

Other works in the series

The Romance Languages
The Celtic Languages
The Slavonic Languages
The Germanic Languages
The Semitic Languages
The Indo-European Languages
The Dravidian Languages
The Uralic Languages

THE TURKIC LANGUAGES

EDITED BY
Lars Johanson
and
Éva Á. Csató



LONDON AND NEW YORK

9 Chaghatay	166
Hendrik Boeschoten and Marc Vandamme	
10 Ottoman Turkish	179
Celia Kerslake	
11 Turkish	203
Éva Á. Csató and Lars Johanson	
12 Turkish Dialects	236
Bernt Brendemoen	
13 The Turkish Language Reform	242
Bernt Brendemoen	
14 Azerbaijanian	248
Claus Schönig	
15 Turkmen	261
Claus Schönig	
16 Turkic Languages of Iran	273
Gerhard Doerfer	
17 Tatar and Bashkir	283
Árpád Berta	
18 West Kipchak Languages	301
Árpád Berta	
19 Kazakh and Karakalpak	318
Mark Kirchner	
20 Noghay	333
Éva Á. Csató and Birsel Karakoç	
21 Kirghiz	344
Mark Kirchner	
22 Uzbek	357
Hendrik Boeschoten	
23 Uyghur	379
Reinhard F. Hahn	
24 Yellow Uyghur and Salar	397
Reinhard F. Hahn	
25 South Siberian Turkic	403
Claus Schönig	

26 Yakut	417
Marek Stachowski and Astrid Menz	
27 Chuvash	434
Larry Clark	
Appendix	453
Index	457

List of Maps and Figures

Map i	The Turkic languages	xvi
Figure 9.1	Relations of Chaghatay with other written varieties	168

List of Tables

i	Consonants (unnumbered)	xix
1.1	Speakers of Turkic languages	13
4.1	Vowels in Proto-Turkic first syllables	70
4.2	Vowels in Proto-Turkic non-first syllables	70
5.1	Oghuz and non-Oghuz lexical differences	120
7.1	The consonant system	139
7.2	Possessive suffixes	142
7.3	Case suffixes	142
7.4	Personal pronouns	143
7.5	Demonstrative pronouns	143
7.6	Simple past	145
7.7	Optative–imperative	146
10.1	Personal pronouns	189
11.1	Short vowels (in Turcological notation)	203
11.2	Consonants (in Turcological notation)	204
11.3	Possessive suffixes	210
11.4	Case suffixes	210
11.5	Paradigms of personal pronouns	211
11.6	Optative	215
14.1	Case suffixes	251
14.2	Possessive suffixes	252
14.3	Present and aorist (<i>al-</i> ‘take’)	254
14.4	Simple past (<i>at-</i> ‘throw’, <i>öl-</i> ‘die’)	255
14.5	Imperative	255
14.6	Optative	255
14.7	Conditional	256
15.1	Case suffixes	264
15.2	Possessive suffixes	264
15.3	Present and aorist (<i>al-</i> ‘take’)	267
15.4	Simple past (<i>at-</i> ‘throw’ and <i>öl-</i> ‘die’)	267
15.5	Optative–imperative (<i>gel-</i> ‘come’)	268
16.1	Personal suffixes	274
16.2	Khorasan Turkic dialect features	275
16.3	Divergence from Turkish	279

17.1 Possessive suffixes	287
17.2 Case suffixes in Tatar	287
17.3 Case suffixes in Bashkir	288
17.4 Declension with possessive suffixes in Tatar	288
17.5 Declension with possessive suffixes in Bashkir	288
17.6 Declension of personal pronouns	289
17.7 Declension of reflexive pronouns in Tatar	289
17.8 Present in Tatar	292
17.9 Present in Bashkir	292
17.10 Simple past	292
17.11 -GAN past	293
17.12 Aorist	293
17.13 Prospective	294
17.14 Optative-imperative	294
17.15 Conditional	295
19.1 Consonants	320
19.2 Possessive suffixes	323
19.3 Case markers	323
19.4 Personal pronouns	323
19.5 Copulas	324
19.6 Personal markers	326
19.7 Optative-imperative	326
20.1 Vowels	333
20.2 Consonants	334
20.3 Case suffixes	336
20.4 Personal pronouns	337
21.1 Short vowels	345
21.2 Consonants	345
21.3 Possessive suffixes	347
21.4 Case suffixes	348
21.5 Personal pronouns	348
21.6 Copula suffixes	349
21.7 Personal markers	350
21.8 Optative-imperative	351
22.1 Consonants	359
22.2 Personal and demonstrative pronouns	362
22.3 Copula forms	364
22.4 Thematic suffixes	366
23.1 Vowel phonemes	380
23.2 Consonant phonemes	381
23.3 Vowel phonemes in suffixes	383
23.4 Possessive markers	389
23.5 Case marking	389
23.6 Personal pronouns	391

23.7 Demonstrative pronouns	391
24.1 Yellow Uyghur vowels	398
24.2 Yellow Uyghur consonants	398
24.3 Salar vowels	400
24.4 Salar consonants	401
25.1 Suffix-initial consonants of the plural and accusative suffixes	406
25.2 Plural possessive suffixes	408
25.3 Declension of first-person singular pronouns	409
25.4 Declension of the 3p.sg. pronoun <i>ol</i>	410
25.5 Cardinal numbers from 1 to 10	411
25.6 Tens of cardinal numbers	411
26.1 Vowels	418
26.2 Consonants	418
26.3 Sound harmony	419
26.4 Consonant and vowel variants in suffixes	420
26.5 Case suffixes	421
27.1 Vowels	435
27.2 Consonants	435
27.3 Possessive suffixes	438
27.4 Case suffixes following plural forms	438
27.5 Case suffixes	439
27.6 Personal pronouns	439
27.7 Demonstrative pronouns	440
27.8 Reflexive pronouns	441
27.9 Numerals	442
27.10 Personal markers of finite verbs	444
Appendix	
1 Uzbek script	454
2 Kazakh script	454
3 Uyghur writing systems	455

List of Contributors

Árpád Berta, University of Szeged, Hungary.
Hendrik Boeschoten, Tilburg University, The Netherlands.
Bernt Brendemoen, University of Oslo, Norway.
Larry Clark, Indiana University, Bloomington, Indiana, USA.
Éva Á. Csató, University of Uppsala, Sweden.
Gerhard Doerfer, University of Göttingen, Germany.
Marcel Erdal, University of Frankfurt, Germany.
Peter B. Golden, Rutgers University, Newark, New Jersey, USA.
Reinhard F. Hahn, University of Seattle, Washington, USA.
Lars Johanson, University of Mainz, Germany.
Birsel Karakoç, University of Mainz, Germany.
Celia Kerslake, University of Oxford, UK.
Mark Kirchner, University of Frankfurt, Germany.
Astrid Menz, University of Mainz, Germany.
András Róna-Tas, University of Szeged, Hungary.
Claus Schönig, Free University Berlin, Germany.
Marek Stachowski, University of Cracow, Poland.
Marc Vandamme, University of Utrecht, The Netherlands.

Preface

A General Introduction to the Turkic Family

The purpose of this volume is to present data, both synchronic and diachronic, to characterise one of the world's major language families. The book aims at achieving a reasonable coverage of the diversity of both ancient and modern Turkic languages, focusing on certain key aspects of their structure and history.

This introduction differs from previous surveys of the Turkic languages by trying to meet the requirements not only of Turcologists, but of a variety of readers. It is intended to be accessible to a wide readership without any previous knowledge of Turcology. The theoretical basis is relatively neutral, and less widely known terms are, as a rule, explained.

First, the book is designed to serve as a general source of reference for all readers with an interest in the Turkic family. It is also intended as a textbook for undergraduate or graduate courses in Turcology and Linguistics. Furthermore, it is hoped that the book will offer information of value to scholars in various disciplines, such as general linguists, typologists and historical linguists. The volume may also be useful for Turcologists searching for information on languages different from the ones they are already familiar with. Finally, the information offered may prove helpful as background knowledge to a study of Oriental history and literature. The editors have tried to ensure that the book provides an up-to-date survey of current knowledge in the entire field covered, and that it is sufficiently clear and expository for all the purposes mentioned here.

The Structure of the Volume

All chapters in this volume are written by specialists in the respective fields. Most chapters are devoted to individual contemporary languages, whereas some deal with a whole branch or sub-branch. The attention accorded to each language has been determined on purely linguistic and historical grounds, independent of external political considerations. The remaining chapters deal with the distribution and history of the Turkic-speaking peoples, the salient structural features of Turkic, extinct written languages, alphabets and transcription problems, the Turkish language reform, and the Turkic languages spoken in a particular area, Iran. Diachronic issues are basically

limited to two chapters, one on the reconstruction of an ancestor of the Turkic languages – with a brief discussion of the controversial question of the Altaic linguistic unity – and one on the history of Turkic.

The descriptive chapters have been organised according to structural linguistic criteria. The editors have tried to ensure that a similar range of core topics is covered in all of them. The order of sections is as similar as possible to make it more easy for the reader to find comparable information on each of the languages.

Indications of further sources in English and other widely read languages are provided under References and Further Reading at the end of each chapter. In general, these references have been kept to a minimum, but a number of titles relevant for most chapters have been listed under Chapter 5, 'The History of Turkic'. Chapter 1, 'The Speakers of Turkic Languages' is partly based on material contributed by the authors of the individual language chapters. For the present geographical location of the Turkic languages, reference should be made to the map on pp.xvi–xvii.

Limitations

The natural length restrictions within a volume of this kind have often faced the authors and the editors with difficult choices as to which aspects should be included. Considerable reductions and omissions of details have been necessary, sometimes at the risk of some overgeneralisation. A detailed coverage of certain interesting languages such as Gagauz, Karaim and Fu-yü has proved impossible within the limits of the volume.

However, the authors of this survey have also been confronted with another difficulty: the incontrovertible fact that the Turkic family as a whole is not yet sufficiently investigated and that much of the necessary detailed linguistic analysis has not yet been carried out. In particular, the sociological aspects on language use are under-represented.

On the other hand, possible deficiencies of these kinds may serve to highlight areas for potential future research, to stimulate new investigations into the rich data offered by the Turkic languages. This family forms a rather homogeneous group, ideal for comparative and typological work and offering ample evidence for the reconstruction of its historical evolution.

Diversity and Harmonisation

The multiple authorship inevitably conditions a certain diversity according to the contributors' differing perspectives and ranges of interest. The individual chapters differ somewhat from each other in the relative balance between the sections and in the selection of the topics dealt with in detail. There are obvious differences of emphasis and presentation. As expected, there are sometimes also discrepancies in the theoretical approaches adopted and disagreement about how to interpret the linguistic data.

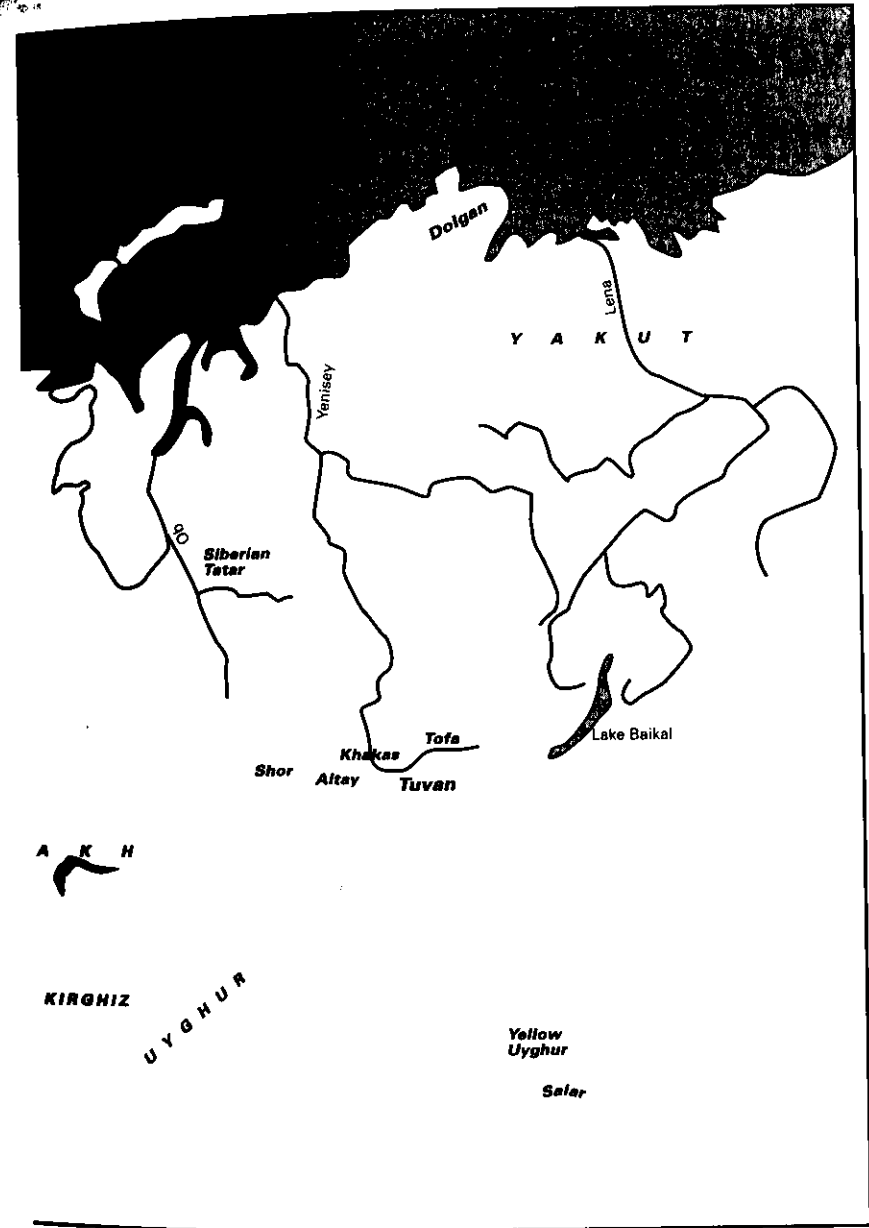
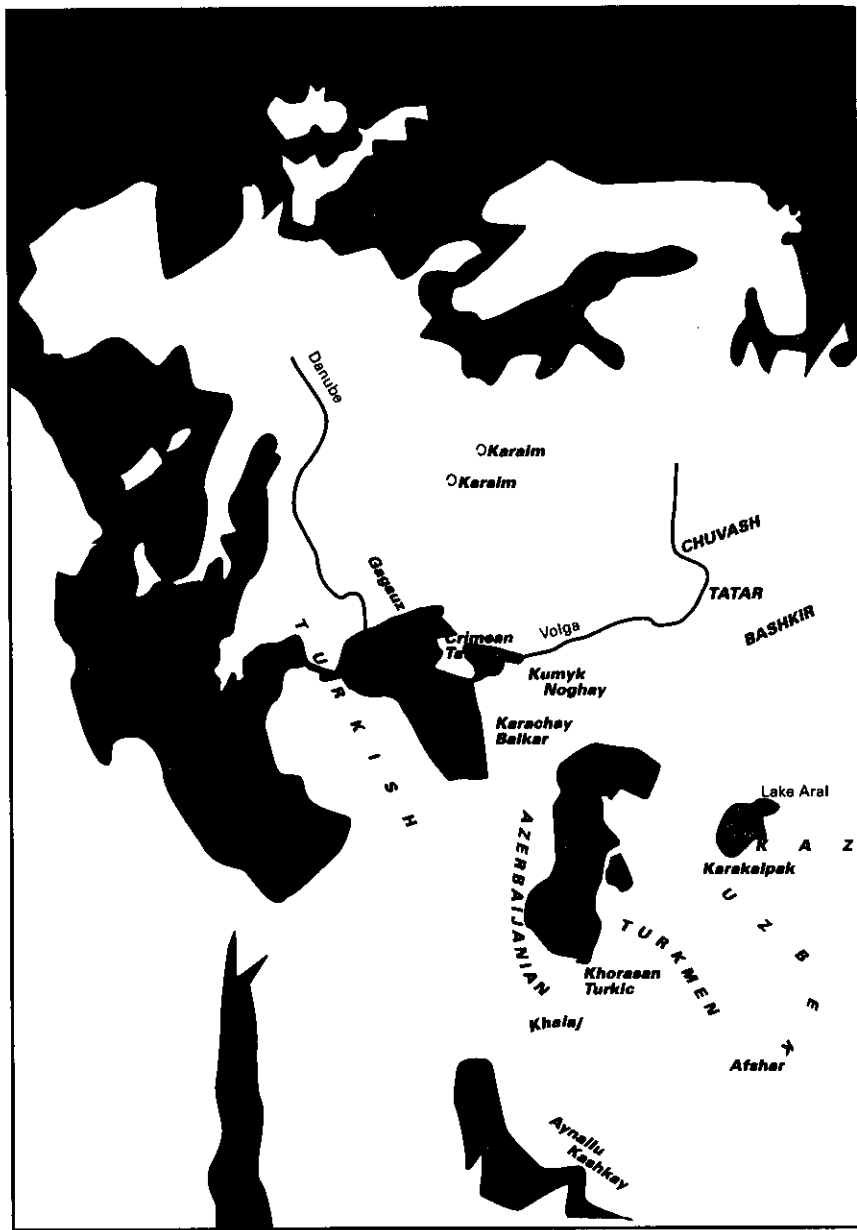
This diversity has not caused any major contradictions between the chapters. The editors have not aimed at a complete resolution of such problems, but have allowed the different perspectives to stand side by side. Many linguistic problems have necessarily been left unsolved. Some questions are of such a nature that they cannot get clear-cut and unanimous answers, e.g. the issue of the genetic relationship of Altaic languages. After all, the book aims at documenting the results of current and previous work in a realistic way, reflecting the factual situation in linguistic Turcology.

In view of the handbook character of the present volume, a certain harmonisation or standardisation has been necessary. The aim has been to employ, as far as possible, a unified terminology without all too idiosyncratic expressions used exclusively in Turcological writings. In the descriptions of verbal categories, the notions and terms introduced by Lars Johanson (see References and Further Reading, Chapter 5) have largely been adopted. This choice has been motivated by the fact that the verbal systems of several Turkic languages have been described within that framework.

Acknowledgements

The editors would like to acknowledge the cooperation of all the individual contributors, thanking them most sincerely for their openness and readiness to accept the teamwork conditions of this enterprise. Our special thanks are due to Bernard Comrie, who originally suggested the publication of the volume and scrutinised our first proposal for it, and Jonathan Price, who decisively encouraged us to go ahead with our plans. The editors are extremely grateful to the editorial staff at Routledge for their highly positive attitude and wonderful patience during the preparation of the book: particular thanks are due to Sarah Hall and Samantha Parkinson. Last, but certainly not least, we want to express our gratitude to Rosemary Morlin for her skilful and empathetic copy-editing.

Lars Johanson and Éva Ágnes Csató



Map i The Turkic languages

Notes on Transcription and Symbols

Transcription

The official orthography of the Republic of Turkey is used for Modern Turkish. For languages spoken outside Turkey, a variety of alphabets and orthographical conventions are applied, which makes transcription necessary. The recently introduced Latin-based scripts for Azerbaijani, Turkmen, etc. are not yet sufficiently established. A transliteration of the Cyrillic scripts, which differ considerably from each other, would distort the phonetic shape and make comparison between the languages difficult. All examples from languages which use the Cyrillic or Arabic script will thus be given in transcription. The system applied, referred to as the 'Turcological notation', is basically identical to the traditional 'Fundamenta' notation as represented in Deny *et al.* 1959, pp. xiv–xv (see References and Further Reading, pp. 122–5).

The sections 'Sound system(s)' primarily deal with the sounds of the language in question and relations among them. In spite of differing approaches to the problems, these sections generally contain remarks on the inventory of phonemes, i.e. which sounds can be used in the language to distinguish words. Though some authors tend more strongly towards a phonemic representation, the transcription will in general not only represent phonemes, but also major allophonic variants, thus conveying an idea of the actual pronunciation. For example, front *g* and *k* sounds will as a rule, irrespective of their phonemic status, be rendered as *g* and *k*, and their back equivalents as *ɣ/g* and *q*. The phonetic difference between a front *l* and a back *ɫ* will not, however, be marked. The transcription is thus relatively broad. Wherever necessary, finer phonetic details are given in the transcription of the International Phonetic Association (IPA).

Vowel Notations

The transcription employed in the present volume differs from the 'Fundamenta' notation mainly with respect to the representation of *e*-sounds. Thus, the normal low type [ɛ] with all its variants is written as *e* instead of the 'Fundamenta' notation *ä*. If a particular language also has a higher *e*

phoneme, a so-called 'closed *e*', it is rendered as *è*. If, in addition to *e*, a lower phoneme is present, it is written *æ*. The low front unrounded Uzbek vowel is transcribed as *â*.

The sign used to denote a front unrounded high vowel is *i*. A corresponding back vowel is written as *ï* (in Turkish orthography *ı*). A front rounded high vowel is written as *ü*, a back rounded high vowel as *u*, a back rounded low vowel as *o*, and a back unrounded low vowel as *a*. A labialised variant of *a* is rendered by *â*, e.g. in Uzbek. Centralised, retracted or lowered variants of vowels are written with an under-dot, e.g. *ï*. Centralisation can also be denoted by an over-dot. Thus, in the transcription of Karaim words, *ü̇*, *ö̇* are to be read as retracted *ü* and *ö* respectively. The weaker element of a diphthong is marked with a raised letter, e.g. *ʼa*, *ʼe*, *ʼo*, *ʼö*. Reduced vowels are written with the diacritic breve, e.g. *ă*, *ĕ*. The sign [◌] is used to signal a reconstructed, probably reduced, vowel entity. Long vowels are indicated by a macron, e.g. *ā*, and semi-long vowels, e.g. in Khalaj, by a dot following the vowel, e.g. *a˙*.

Consonant Notations

The following table includes the most important consonant signs occurring in the volume.

Table i Consonants

	Labial	Prepalatal	(Post)palatal	Velar, uvular, glottal
stop	p, b	t, d	k, g	q, ɡ, ʔ
fricative	f, v, θ, ð	s, z	š [ʃ], ž [ʒ]	ɣ, ɣ̥, h
nasal	m	n	ɲ	ŋ
affricate		c [tʃ], dʒ	č [tʃ̣], ǰ [dʒ̣]	
glide	w		y	
liquid		l, ɫ, r		

In the above table, some IPA correspondences are given in square brackets []. In some chapters, the term velar is used to cover both postpalatal and velar/uvular dorsals.

Palatalisation is marked with an apostrophe, as in *kʼ*. In some cases, the diacritic is placed over the consonant sign; thus, the palatalised *n* and *s* are written *ɲ* and *ʃ*, respectively. Reduced voice is denoted with an under-ring, e.g. *ɣ̥*. Some authors use small capitals for lenis variants with reduced voice, e.g. Chuvash [ɛbir] 'we' = *ɛb̥ir* (p. 435). Aspiration is indicated with a superscript *h*, e.g. *p^h*. The acute accent sign marks high pitch accent, e.g. *á*, whereas the vertical stroke ' indicates stress accent on the following syllable. Deviating notations are explained in the relevant chapters.

Suffix Notations

The sections on morphophonology deal with the ways in which the phonemic shape of an item can change in different morphological forms. Suffixes have different – largely phonologically predictable – realisations depending on phonotactic and harmonic properties of stems. They may thus be cited in standardised forms which summarise all possible realisations, e.g. $-(y)A$ for $-e$, $-a$, $-ye$, $-ya$ in the Turkish dative suffix.

In our suffix notations, capital letters indicate morphophonemes, e.g. vowel alternants due to intersyllabic sound harmony variation and consonant alternants due to contact assimilation. Note that not all authors make consistent use of these shorthand formulas. The following rules account for how the abbreviations are to be read.

Parentheses

Parentheses around an initial vowel of a suffix indicate that this vowel is dropped when the suffix is added to a vowel-final stem. The vowel is thus realised only if the suffix follows a consonant, e.g. Turkish 1p.sg. possessive suffix $-(I)m$ in *araba-m* 'my car' but *kardeş-im* 'my brother'.

Parentheses around the initial consonant of a suffix indicate that this consonant occurs only if the suffix is added to a vowel-final stem, e.g. the Turkish dative suffix $-(y)A$ in *araba-ya* 'to a/the car', but *kardeş-e* 'to a/the brother', the distributive suffix $-(ş)Ar$ in *üç-er* 'three each', *iki-şer* 'two each'.

Parentheses around a final consonant of a suffix indicate that this consonant occurs before case suffixes, e.g. Turkish $-(s)I(n)$ in *evinden* 'from his/her house'.

Parentheses around an initial zero sign \emptyset in a suffix indicate that the final vowel of the stem drops when the suffix is added, e.g. Chuvash present in $-(\emptyset)At$, Turkish in $-(\emptyset)Iyor$.

Parentheses around a colon in a suffix indicate that the addition of the suffix to a stem-final vowel yields length, e.g. Turkmen $-A(:)n$, e.g. *oqān* 'having read' ← *oqa-* [ɔkɔ] 'read', but *gelen* 'having come' ← *gel-* 'come'.

Vowels

The capital letter *V* indicates any vowel that can occur in suffixes, thus in many Turkic languages all vowels except *o* and *ö*.

The capital letter *U* indicates alternation of high rounded vowels, *ü* and *u*.

The sign $^{\circ}$ stands for Old Turkic vowels of controversial phonetic nature (see pp. 107–8), claimed by some scholars to be reduced vowels, e.g. the accusative suffix $-(^{\circ})G$ as in *sü-g* 'army', *at-^{\circ}γ* 'horse'.

The capital letter *A* indicates low vowel alternation, at least the twofold variation of *e* (α , \acute{a}) and *a* (\grave{a}). The vowel is *e* when the preceding syllable is front, and it is *a* when the preceding syllable is back, e.g. the Turkish dative suffix $-(y)A$ as in *araba-ya* 'to a/the car', but *kardeş-e* 'to a/the brother', or the Turkmen present suffix $-(y)ArIn$ as in *Alyārin* 'I take', but *Gülyārin* 'I laugh'. The alternation may also be three- or fourfold, additionally compris-

ing rounded variants such as *o* and *ö*, e.g. the Turkmen plural suffix $-lAr$ as in $-lar$, $-ler$, $-lor$, $-lör$.

The capital letter *I* indicates high vowel alternation, from a twofold alternation of *i* and *ı* up to a fourfold one comprising *i*, *ĩ*, *ü*, *u*, e.g. Turkmen $-dIK$ as in $-dik$, $-dīq$, $-dük$, $-duq$. In some contributions to this volume, this fourfold alternation is indicated with the symbol *X*. The rules for it are, e.g. in Turkish, as follows:

- 1 The suffix vowel is *i* when the preceding syllable is front and its vowel unrounded (*i* or *e*).
- 2 The suffix vowel is *ü* when the preceding syllable is front and its vowel is rounded (*ü* or *ö*).
- 3 The suffix vowel is *ı* (in Turkish written *ı*) when the preceding syllable is back and its vowel unrounded (*ı* or *a*).
- 4 The suffix vowel is *u* when the preceding syllable is front and its vowel rounded (*u* or *o*).

Examples: Turkish third-person possessive suffix $-(s)I$ in *araba-sı* 'his/her car', *kardeş-i* 'his/her brother', *üzüntü-sü* 'his/her worry', *kuş-u* 'his/her bird'.

The capital letter *Ė* indicates an analogous alternation in languages of the Volga type, from a twofold alternation of \check{e} and \check{i} (mostly in Tatar) up to a fourfold alternation of \check{e} , \check{i} , \check{o} , $\check{ö}$ (mostly in Bashkir).

Consonants

One type of consonant alternation in suffixes depends on whether the last sound of the stem is voiced or voiceless. Members of consonant clusters often undergo progressive assimilation with respect to voice.

The capital letter *D* mostly denotes the twofold alternation of voiced *d* and unvoiced *t*, e.g. the Turkish locative suffix $-DA$ as in *yol-da* 'on (the) way', but *süt-te* 'in (the) milk'.

The capital letter *J*, in Turkish written *C*, denotes the twofold alternation of voiced *J* (in Turkish written *c*) and unvoiced \check{c} (in Turkish written \check{c}). Example: the Turkish derivation suffix $-CI$ in *yol-cu* 'traveller', but *süt-çü* 'milkman'.

The capital letter *G* indicates the twofold alternation of front and back consonants of the type *g* vs. χ , e.g. Azerbaijani *y* vs. γ , Turkish *y* vs. \emptyset , both written $\langle\check{g}\rangle$. It also stands for fourfold suffix-initial alternations (front vs. back, voiced vs. unvoiced) of the type *g*, χ , *k*, *q*, e.g. the Chaghatay dative suffix $-ya$, $-ge$, $-qa$, $-ke$, the Kazakh participle suffix $-GAN$. In some cases, *G* only signals a twofold alternation between *g* (voiced) and *k* (unvoiced), e.g. the Turkish deverbal derivative suffix $-GI$ in *sev-gi* 'affection', but *iç-ki* 'drink'.

The capital letter *K* indicates the twofold alternation of front and back dorsal stops of the type *k* and *q*, e.g. Azerbaijani $-mAKdA$ with *k* and *g*, Turkish $-mAKtA$ as in *gelmekte* 'coming' with [k] vs. *almakta* 'taking' with [k], both written $\langle k \rangle$.

Notes on Transcription and Symbols

Transcription

The official orthography of the Republic of Turkey is used for Modern Turkish. For languages spoken outside Turkey, a variety of alphabets and orthographical conventions are applied, which makes transcription necessary. The recently introduced Latin-based scripts for Azerbaijanian, Turkmen, etc. are not yet sufficiently established. A transliteration of the Cyrillic scripts, which differ considerably from each other, would distort the phonetic shape and make comparison between the languages difficult. All examples from languages which use the Cyrillic or Arabic script will thus be given in transcription. The system applied, referred to as the 'Turcological notation', is basically identical to the traditional 'Fundamenta' notation as represented in Deny *et al.* 1959, pp. xiv-xv (see References and Further Reading, pp. 122-5).

The sections 'Sound system(s)' primarily deal with the sounds of the language in question and relations among them. In spite of differing approaches to the problems, these sections generally contain remarks on the inventory of phonemes, i.e. which sounds can be used in the language to distinguish words. Though some authors tend more strongly towards a phonemic representation, the transcription will in general not only represent phonemes, but also major allophonic variants, thus conveying an idea of the actual pronunciation. For example, front *g* and *k* sounds will as a rule, irrespective of their phonemic status, be rendered as *g* and *k*, and their back equivalents as *ɣ*/*ġ* and *q*. The phonetic difference between a front *l* and a back *ʎ* will not, however, be marked. The transcription is thus relatively broad. Wherever necessary, finer phonetic details are given in the transcription of the International Phonetic Association (IPA).

Vowel Notations

The transcription employed in the present volume differs from the 'Fundamenta' notation mainly with respect to the representation of *e*-sounds. Thus, the normal low type [ɛ] with all its variants is written as *e* instead of the 'Fundamenta' notation *ä*. If a particular language also has a higher *e*

phoneme, a so-called 'closed *e*', it is rendered as *è*. If, in addition to *e*, a lower phoneme is present, it is written *æ*. The low front unrounded Uzbek vowel is transcribed as *â*.

The sign used to denote a front unrounded high vowel is *i*. A corresponding back vowel is written as *ï* (in Turkish orthography *ı*). A front rounded high vowel is written as *ü*, a back rounded high vowel as *u*, a back rounded low vowel as *o*, and a back unrounded low vowel as *a*. A labialised variant of *a* is rendered by *â*, e.g. in Uzbek. Centralised, retracted or lowered variants of vowels are written with an under-dot, e.g. *î*. Centralisation can also be denoted by an over-dot. Thus, in the transcription of Karaim words, *û*, *ô* are to be read as retracted *ü* and *ö* respectively. The weaker element of a diphthong is marked with a raised letter, e.g. *ʼa*, *ʼe*, *ʼo*, *ʼö*. Reduced vowels are written with the diacritic breve, e.g. *ă*, *ĕ*. The sign *°* is used to signal a reconstructed, probably reduced, vowel entity. Long vowels are indicated by a macron, e.g. *ā*, and semi-long vowels, e.g. in Khalaj, by a dot following the vowel, e.g. *a˙*.

Consonant Notations

The following table includes the most important consonant signs occurring in the volume.

Table i Consonants

	Labial	Prepalatal	(Post)palatal	Velar, uvular, glottal
stop	p, b	t, d	k, g	q, ġ, ʔ
fricative	f, v, θ, ð	s, z	š [ʃ], ž [ʒ]	ɣ, χ, h
nasal	m	n	ñ	ŋ
affricate		c [tʃ], dz	č [tʃ], ĵ [dʒ]	
glide	w		y	
liquid		l, ł, r		

In the above table, some IPA correspondences are given in square brackets []. In some chapters, the term velar is used to cover both postpalatal and velar/uvular dorsals.

Palatalisation is marked with an apostrophe, as in *kʼ*. In some cases, the diacritic is placed over the consonant sign; thus, the palatalised *n* and *s* are written *ñ* and *ś*, respectively. Reduced voice is denoted with an under-ring, e.g. *ḥ*. Some authors use small capitals for lenis variants with reduced voice, e.g. Chuvash [ebir] 'we' = *eb̃ir* (p. 435). Aspiration is indicated with a superscript *h*, e.g. *pʰ*. The acute accent sign marks high pitch accent, e.g. *á*, whereas the vertical stroke ' indicates stress accent on the following syllable. Deviating notations are explained in the relevant chapters.

Suffix Notations

The sections on morphophonology deal with the ways in which the phonemic shape of an item can change in different morphological forms. Suffixes have different – largely phonologically predictable – realisations depending on phonotactic and harmonic properties of stems. They may thus be cited in standardised forms which summarise all possible realisations, e.g. $-(y)A$ for $-e$, $-a$, $-ye$, $-ya$ in the Turkish dative suffix.

In our suffix notations, capital letters indicate morphophonemes, e.g. vowel alternants due to intersyllabic sound harmony variation and consonant alternants due to contact assimilation. Note that not all authors make consistent use of these shorthand formulas. The following rules account for how the abbreviations are to be read.

Parentheses

Parentheses around an initial vowel of a suffix indicate that this vowel is dropped when the suffix is added to a vowel-final stem. The vowel is thus realised only if the suffix follows a consonant, e.g. Turkish 1p.sg. possessive suffix $-(I)m$ in *araba-m* 'my car' but *kardeş-im* 'my brother'.

Parentheses around the initial consonant of a suffix indicate that this consonant occurs only if the suffix is added to a vowel-final stem, e.g. the Turkish dative suffix $-(y)A$ in *araba-ya* 'to a/the car', but *kardeş-e* 'to a/the brother', the distributive suffix $-(ş)Ar$ in *üç-er* 'three each', *iki-şer* 'two each'.

Parentheses around a final consonant of a suffix indicate that this consonant occurs before case suffixes, e.g. Turkish $-(s)I(n)$ in *evinden* 'from his/her house'.

Parentheses around an initial zero sign \emptyset in a suffix indicate that the final vowel of the stem drops when the suffix is added, e.g. Chuvash present in $-(\emptyset)At$, Turkish in $-(\emptyset)Iyor$.

Parentheses around a colon in a suffix indicate that the addition of the suffix to a stem-final vowel yields length, e.g. Turkmen $-A(:)n$, e.g. *oqān* 'having read' ← *oqa-* [ɔkɔ] 'read', but *gelen* 'having come' ← *gel-* 'come'.

Vowels

The capital letter V indicates any vowel that can occur in suffixes, thus in many Turkic languages all vowels except o and \ddot{o} .

The capital letter U indicates alternation of high rounded vowels, \ddot{u} and u .

The sign $^\circ$ stands for Old Turkic vowels of controversial phonetic nature (see pp. 107–8), claimed by some scholars to be reduced vowels, e.g. the accusative suffix $-(^\circ)G$ as in *sü-g* 'army', *at-^\circ\gamma* 'horse'.

The capital letter A indicates low vowel alternation, at least the twofold variation of e (α , \acute{a}) and a (\grave{a}). The vowel is e when the preceding syllable is front, and it is a when the preceding syllable is back, e.g. the Turkish dative suffix $-(y)A$ as in *araba-ya* 'to a/the car', but *kardeş-e* 'to a/the brother', or the Turkmen present suffix $-(y)\bar{A}rIn$ as in *Alyārin* 'I take', but *Gülyārin* 'I laugh'. The alternation may also be three- or fourfold, additionally compris-

ing rounded variants such as o and \ddot{o} , e.g. the Turkmen plural suffix $-lAr$ as in $-lar$, $-ler$, $-lor$, $-lör$.

The capital letter I indicates high vowel alternation, from a twofold alternation of i and \ddot{i} up to a fourfold one comprising i , \ddot{i} , u , e.g. Turkmen $-dIK$ as in $-dik$, $-diq$, $-dük$, $-duq$. In some contributions to this volume, this fourfold alternation is indicated with the symbol X . The rules for it are, e.g. in Turkish, as follows:

- 1 The suffix vowel is i when the preceding syllable is front and its vowel unrounded (i or e).
- 2 The suffix vowel is \ddot{u} when the preceding syllable is front and its vowel is rounded (\ddot{u} or \ddot{o}).
- 3 The suffix vowel is \ddot{i} (in Turkish written ι) when the preceding syllable is back and its vowel unrounded (\ddot{i} or a).
- 4 The suffix vowel is u when the preceding syllable is front and its vowel rounded (u or o).

Examples: Turkish third-person possessive suffix $-(s)I$ in *araba-sı* 'his/her car', *kardeş-i* 'his/her brother', *üzüntü-sü* 'his/her worry', *kuş-u* 'his/her bird'.

The capital letter \tilde{E} indicates an analogous alternation in languages of the Volga type, from a twofold alternation of \tilde{e} and \tilde{i} (mostly in Tatar) up to a fourfold alternation of \tilde{e} , \tilde{i} , $\tilde{\ddot{o}}$, $\tilde{\ddot{o}}$ (mostly in Bashkir).

Consonants

One type of consonant alternation in suffixes depends on whether the last sound of the stem is voiced or voiceless. Members of consonant clusters often undergo progressive assimilation with respect to voice.

The capital letter D mostly denotes the twofold alternation of voiced d and unvoiced t , e.g. the Turkish locative suffix $-DA$ as in *yol-da* 'on (the) way', but *süt-te* 'in (the) milk'.

The capital letter J , in Turkish written C , denotes the twofold alternation of voiced j (in Turkish written c) and unvoiced \check{c} (in Turkish written \check{c}). Example: the Turkish derivation suffix $-CI$ in *yol-cu* 'traveller', but *süt-çü* 'milkman'.

The capital letter G indicates the twofold alternation of front and back consonants of the type g vs. x , e.g. Azerbaijani y vs. γ , Turkish y vs. \emptyset , both written $\langle\check{g}\rangle$. It also stands for fourfold suffix-initial alternations (front vs. back, voiced vs. unvoiced) of the type g , x , k , q , e.g. the Chaghatay dative suffix $-ya$, $-ge$, $-qa$, $-ke$, the Kazakh participle suffix $-GAN$. In some cases, G only signals a twofold alternation between g (voiced) and k (unvoiced), e.g. the Turkish deverbal derivative suffix $-GI$ in *sev-gi* 'affection', but *iç-ki* 'drink'.

The capital letter K indicates the twofold alternation of front and back dorsal stops of the type k and q , e.g. Azerbaijani $-mAKdA$ with k and g , Turkish $-mAKtA$ as in *gelmekte* 'coming' with [k] vs. *almakta* 'taking' with [k], both written $\langle k \rangle$.

The letter *K* also stands for fourfold suffix-final alternations front vs. back, voiced vs. unvoiced of the type *g, x, k, q*, e.g. Noghay prospective $-(A)yAK$, Ottoman $-mAK$, $-DUK$, $-(y)AJAK$, Azerbaijani $-mAK$ with the alternation *y, x, k, g*, Turkish $-IK$ with the alternation *y, ø, k, q*. The voiced variants occur before initial suffix vowels. The Turkish alternation is represented orthographically by <ğ> and <k> alone, e.g. the derivative suffix $-IK$ in *çiftlik* 'farm', but *çiftliğ-i* 'his/her farm', *aylık* 'monthly salary', but *aylığı* 'his/her monthly salary', the prospective suffix $-(y)AcAK$ in *Yazacağ-ım* 'I will/shall write', but *Yazacak-lar* 'They will/shall write'.

The capital letter *L* indicates alternants such as *l, n, d, t*. A plural suffix indicated as $-LAR$ may thus maximally have the variants $-lar, -ler, -lor, -lör, -nar, -ner, -nor, -nör, -dar, -der, -dor, -dör, -tar, -ter, -tor, -tör$.

The capital letter *R* in Chuvash suffixes stands for *t* and *r* (sometimes also *č*), e.g. locative $-RA = -rA, -tA$, ablative $-RAn = -rAn, -tAn$, 1p.sg. simple past $-RĀm = -rĀm, -tĀm$. In Kumyk, it stands for the alternation between *r* and loss of *r*, e.g. $-ler$ (plural nominative) vs. $-lege$ (dative plural).

Further language-specific alternations are discussed in the individual chapters.

Other Signs

A dash to the right of a word signifies a verbal stem, e.g. Turkish *git-* 'go'. A dash to the left signifies a bound element, a suffix, e.g. Turkish $-IK$. The Turcological practice of using plus signs to indicate nominal stems and denominal suffixes is not followed here. $ø$ is the sign for a zero element.

Simple arrows are used for derivation. Thus, \leftarrow means 'is derived from', and \rightarrow means 'is derived as'. Double arrows are used for copying (e.g. 'borrowing') of foreign elements. Thus, \Leftarrow means 'is copied from', and \Rightarrow means 'is copied as'. The sign $<$ means 'has developed from', and $>$ means 'has developed into'.

Square brackets [] are used for phonetic transcriptions, slashes / / for phonemic representations, and angular brackets $< >$ for graphic representations. Whenever possible, brackets and slashes are dispensed with in order to facilitate reading.

The asterisk * is used in historical contexts for reconstructed (not attested) forms. In descriptive contexts, it is the sign for an unacceptable form.

Name Forms

In view of the considerable and confusing variation in the practices of writing names of Turkic peoples and languages, the forms of those names have been unified, e.g. Kirghiz (not Kyrgyz), Azerbaijani (not Azeri or Azerbaijani), Yakut (not Sakha). The ending $-ic$ is used when referring to a whole language family, e.g. Turkic, Mongolic and Tungusic.

Abbreviations of Grammatical Terms

1p.pl.	first-person plural	INSTR.	instrumental
1p.sg.	first-person singular	LIM.	limitative
2p.pl.	second-person plural	LOC.	locative
2p.sg.	second-person singular	MULT.	multiplicative
3p.pl.	third-person plural	NEC.	necessitative
3p.sg.	third-person singular	NEG.	negation
ABL.	ablative	OPT.	optative
ACC.	accusative	PART.	participle
CAUS.	causative	PASS.	passive
COLL.	collective	PL.	plural
COM.	comitative	POSS.	possessive
COMP.	comparative	POSTP.	postposition
COND.	conditional	POSTT.	post-terminal
CONV.	converb	PRES.	present
COORD.	coordinative	PRESUMP.	presumptive
COP.	copula	PROSP.	prospective
DAT.	dative	PST.	past
DER.	derivational suffix	Q.	interrogative
DIR.	directive	R.	orist
EQU.	equative	REC.	reciprocal
FUT.	future	REFL.	reflexive
GEN.	genitive	RES.	resultative
HAB.	habitual	SG.	singular
IMP.	imperative	SIM.	similitive
INF.	infinitive	VOL.	voluntative

1 The Speakers of Turkic Languages

Hendrik Boeschoten

Introduction

As will become apparent from the following chapters in this book, it is not difficult to show that the languages spoken by the different Turkic peoples are genetically related. The often spectacular differences between the subdivisions of the Indo-European language family are not echoed in the Turkic case. For questions of genetic relatedness, see Chapter 4.

Cultural patterns among the Turkic peoples are submerged in larger wholes. There is no intra-Turkic unifying factor, apart from the shared linguistic background. Nevertheless, much of the formation of the modern languages took place in the context of Islamic civilisation, which at the same time reinforced the common linguistic background of most of the tribes. The Islamic states eventually provided the basis for many of today's nationalities, although these did not emerge until the eighteenth century, ultimately to be shaped by the Atatürkist revolution in the case of Turkey, and by the socialist systems of the Soviet Union and China elsewhere.

Thus, the bulk of speakers of Turkic languages live in the old Islamic heartland, and in terms of numbers as well as economic importance the other peoples can be viewed as peripheral. From the perspective of Turkic studies, however, these other peoples carry their own weight. In modern times there are other poles of identification as well. Turkey has always been partly a European country, and has had a strong western orientation ever since the Republic of Turkey came into being.

The Turkic peoples in the former Soviet Union, on the other hand, formed part of the community of Soviet peoples for generations. In the 1990s, we are witnessing a process of reorientation, with diverging and unsettled outcomes in the different republics, while Islamist ideology is asserting itself to a certain extent even in Turkey. Note that allegiance to ideological and political systems at large is symbolised by choices of different alphabets for the literary languages; today the main choices are between a Latin, a Cyrillic or an Arabic alphabet as the basis for writing systems. Some smaller languages are under the heavy pressure of dominant languages surrounding them, such as some of the languages in South Siberia and Yellow Uyghur in China. At least some

languages such as Karaim, spoken by small groups in Lithuania and the Ukraine and Fu-yü in China are on the brink of extinction.

Literary Languages

Literary languages in general are relatively recent creations. Some written idioms emerged, or branched off as national varieties, during the nineteenth and early twentieth centuries from earlier stages that developed in an Islamic context, e.g. modern Turkish from Ottoman Turkish, and Uzbek, Kazakh and Turkmen from Chaghatay. Other literary languages were created, first by Turcologists and missionaries, and later by agencies in the Soviet Union. Soviet policies on nationalities have led to wholesale manipulation of language issues.

Among other things, the Karakalpak, closely related to the Kazakh, attained their own autonomous republic inside the borders of the Soviet Republic of Uzbekistan – not Kazakhstan – and their own literary language. The Karachay and Balkar peoples, who share the same written language, were separated by placing the Karachay in an autonomous republic together with the Cherkessians, and putting the Balkars in a separate autonomous republic together with another Caucasian people, the Kabardinians. At the same time, the differences between the dialects of these two groups were greatly amplified in two separate literary languages. Conversely, the dialects subsumed under the Uzbek literary language belong historically to three different branches of the Turkic language family.

Dividing Factors

At the same time, the cultural and political history of the regions concerned have developed in such a way that the old tribal entities have still retained some relevance in certain cases. This is notably true for the Oghuz peoples. Not only do the languages of these peoples form a clearly discernible, closely related block, but their cultural and political history has also linked them more closely right up to the modern age.

Two moments in history have been decisive in shaping the modern nations. First, the rise of Shī'ism as a state religion in Iran in the sixteenth century, and the ensuing hostility between the Shī'ite Oghuz tribes and the Ottoman state, have ultimately led to the modern situation in which the Azerbaijanians are wedged between the overwhelmingly Sunnī Turks of Turkey and the Turkmens, to mention the three large Oghuz groups of today.

Later, Soviet isolationism rendered close cross-border contacts between Turkey and the Soviet republics impossible. The other tribal denotation often used is that of Kipchak, a generic term for a variety of mainly Islamic peoples of the steppe belt in west-central Eurasia. The numerically and culturally most important peoples of Central Asia or Turkestan proper, e.g. the Uzbeks and

Uyghurs, have for centuries been so ethnically mixed that they cannot be linked in any linguistically meaningful way to a specific larger tribal entity.

Languages and Ethnic Groups

Proceeding to the various Turkic-speaking peoples, the reader should keep in mind that, until recently, the larger groups lived in multi-ethnic empires, and that in general, contact with other languages continues to be a characteristic of many Turkic languages, spread as they are over the vast area of Eurasia. While we will describe the whereabouts of the different peoples and provide their numbers of speakers, one should remember that there is no automatic match between ethnic groups and languages, and that boundaries may be very ill-defined.

Relationships within Turkic may also be quite complex, but these are usually problems of a linguistic nature, which can be kept apart rather easily from problems of ethnic allegiance. Thus, the numbers listed in Table 1.1 on pp. 13–14 are intended to give a fair indication of first-language speakers in all cases, and are not to be taken as absolute truth. Besides, not all numbers are accurate enough to be included in the table. For Turkey, for instance, there are no statistics about the small refugee groups who speak Turkic languages other than Turkish. Quite apart from the impossibility of verifying fully all sources, we simply do not know enough about the patterns of language shift to be encountered in different places.

Speakers in Turkey

As a case in point, the most important single language is Turkish. In order to distinguish the modern language from earlier versions, i.e. Ottoman Turkish, and from the other Turkic languages, the term 'Republican Turkish' has been employed by some Turcologists, but this usage is unsatisfactory, because there are now several Turkic republics and Turkish is also spoken outside Turkey. Turkish has been written since 1928 with a Latin alphabet that supplanted the Arabic alphabet used for Ottoman.

Turkish is spoken as a first language by approximately 55 million people. The total population of Turkey numbers 65 million, but not all of these are first-language speakers of Turkish.

Kurds, speaking either Kurmanji or Zaza, are the most important linguistic minority, making up about ten million of the total population, but there is no reliable information about patterns of language choice and language acquisition among the Kurds, nor among other linguistic groups or dialects.

Many other smaller ethnic groups, living all over Turkey, have preserved their own languages. These include Indo-European languages such as Greek, Judeo-Spanish, Armenian and Serbo-Croat, Caucasian languages such as Cherkas, Georgian, Laz, and the Semitic languages Arabic and Syriac, but also some Turkic languages spoken by refugees from the Russian Empire and

the Soviet Union, and their descendants. The reader is referred to Andrews (1989) for details on the language situation in Turkey.

The linguistic Turkification of Turkey is a process that has been going on since the times of the first principalities in the Middle Ages, and has been speeded up after World War I as a result of the republican language policy.

Turkish Speakers Outside Turkey

Outside Turkey, Turkish is spoken in different areas of the Balkans. The language is under pressure for political reasons, notably in the Greek province of Western Thrace. In Kosovo, where a large group of Albanians adopted Turkish as their mother tongue in Ottoman times, the future of the language does not look bright at the present. In Macedonia, a country where Turkish used to have a strong position, today only 4 per cent of the population speak it according to the official figures.

The only country in the Balkans where the position of Turkish is still strong is Bulgaria. Official attitudes towards the language during the last decades have varied from very oppressive to tolerant. There must be about a million first-language speakers of Turkish in Bulgaria. The position of the language is strengthened by the fact that a large group of Muslim Roma also speak Turkish as their first language.

The Turks of Cyprus were settled there by the Ottoman government in the sixteenth century. The protection of their minority rights in the Republic of Cyprus proved to be a precarious matter, and the attempts to attain unification with Greece resulted in the 1970s in the breakaway Turkish Republic of Northern Cyprus, under the protection of Turkey.

A Turkish-speaking group from the region in Georgia that borders on Turkey, the so-called Meshkheti, were deported to Uzbekistan in Stalinist times. They featured in the news early on in the break-up of the Soviet Union, when ethnic violence was apparently directed against them.

Finally, the migrant workers who went to western, central and northern Europe in the 1960s and 1970s, their families and their offspring form another large group of speakers of Turkish, a total of more than two million people. Various groups of these migrants turned immigrants live scattered over industrial cities in Germany and the Netherlands, and in smaller numbers in Austria, Belgium, France and the Scandinavian countries. Patterns of maintenance or loss of Turkish are varied, but it looks as if at least a considerable part of the immigrants will retain Turkish as their first language in the immediate future. Sizeable groups of recent Turkish migrants also live in Australia. Although the Turkish varieties in these different localities have quite different characteristics in some respects, they all share the same standard language, i.e. the standard variety of Turkey.

Gagauz

A different people, the Gagauz, speak a language intimately related to Turkish. This small group, numbering some 150,000 speakers in southern Moldova, about 32,000 speakers in the southwestern Ukraine, about 10,000 in the Russian Federation, 1,000 in Kazakhstan and a few speakers in northern Bulgaria, Romania and Greece, adhere to the Orthodox faith, and have their own history.

Azerbaijanians

Another language that is closely related to Turkish is Azerbaijani. As previously mentioned, Azerbaijani and Ottoman Turkish have developed in more or less separate contexts during the last few centuries. A further split separated the two main bodies of the Azerbaijanians, when some were included in the Russian empire, while the rest came to live in Iran.

The now independent Republic of Azerbaijan has been a major oil-producing country ever since the beginning of this century. Modernising, i.e. secularist, tendencies have been much stronger there than in Iranian Azerbaijan. The literary language of Azerbaijan has been standardised in Soviet times. A Latin alphabet will probably replace the Cyrillic one in the near future. The Azerbaijanians living in northern Iran, with Tabriz as their main city, number at least 13 million speakers and make up about one quarter of the population of Iran as a whole. It may be, however, that a large number of them are not actually first-language speakers of Azerbaijani, but of Persian.

Further Groups in the South

Along the southern edge of the Turkic world we find, in the midst of a large sedentary population, groups of semi-nomads who have retained forms of tribal organisation. There is a Turkish- or rather Azerbaijani-speaking part of the population of northern Iraq which is sometimes called 'Turkmen', similar to the Yürük tribes in the Balkans and in Anatolia. Oghuz peoples of Iran include the important and large confederation of the Kashkay (Qaşqā'ī), who until quite recently were semi-nomads in the provinces of Fars and Khuzistan and have an identity quite separate from the Azerbaijanians. Besides the Kashkay, there are the Aynallu, and, in Afghanistan, the Afshar, who also speak an Oghuz language.

In northeastern Iran, we find Khorasan Turks, who are not Turkmen, but form a separate entity, linguistically and in their own conception. There is also a sizeable Turkmen population in this area. The main body of Turkmen live in the republic of Turkmenistan. The Turkmen have retained at least a certain resemblance of tribal organisation. The standard language is mainly based on the dialect of the Tekke tribe. Turkmen herdsmen and villagers are also numerous in the region of Maimana in north-western Afghanistan.

A very different group is the small Khalaj population in central Iran. Khalaj is spoken by 28,000 people who live in isolated villages, lacking clear cohesion.

For a more detailed account of Turkic languages in Iran, see Chapter 16.

The Caucasus Area

In the central and northern Caucasus several small Kipchak nations live amongst their Caucasian neighbours. The Kumyk language is spoken by approximately 130,000 people in the north-eastern area of the Caucasus. The Karachay and Balkar (Malqar, Bolqar) peoples appear in historical sources at approximately the same period in the seventeenth century. The ethnonym *Balqar* has erroneously been associated with the name of the Bulgars in the Kuban region. Tradition has it that the ancestors of the Karachay and Balkar peoples lived on the northern slopes of the Caucasus and the plain beyond them, and moved into the mountains as late as the Mongol period. They were strongly influenced by the Alans until the eleventh century.

The number of Karachays inhabiting the triangle formed by the Taberda and Kuban Rivers as well as the high Caucasus can be estimated at almost 70,000. The Balkars live to the south and southeast of them and may number as many as 40,000. In 1944, both peoples were deported to Siberia and it was only after 1957 that they were permitted to return to their homeland in the Caucasus region. Small groups of Karachays live in Turkey, where they show a strong desire to preserve their language – in contrast to other Turkic groups, who are usually quickly Turkicised.

Noghay Groups

A tribal confederation of Kipchak and Kipchakised Mongol clans called Noghay first became known in the fifteenth century and have played an important role in the history of the steppe ever since. The Noghay identity has remained tribal and has not led to national consciousness. The remnants of the Noghay live in three different areas. The Qara Noghay in Dagestan are under strong Kumyk influence, whereas the Aq Noghay in the Karachay-Cherkessian republic and in Chechnia are under Cherkessian influence. A third Noghay group lives in the region of Stavropol north of the Caucasus. Other Noghay scattered groupings have been absorbed by various Tatar ethnic units in the Black Sea area and in Turkey.

Tatars and Bashkirs

The ethnonym Tatar is problematic in itself. Today, several Turkic groups are called Tatar. Originally it was probably a Mongolian tribal name, but after Chingis Khan the term became increasingly associated with the, at least numerically, increasingly dominant Kipchak Turkic element in the armies of

his successors. This is notably true for the Russian usage of the term; they tended to call all Turkic peoples in their empire Tatar. Kazan was firmly established as the capital of the Tatars during the last century. The Russians cooperated with the local intellectual elite in a programme of modernisation of the Islamic peoples within the empire. In fact, the important nations of the Kazakhs and Kirghiz were only effectively Islamicised by missionary work directed from Kazan during the nineteenth century. New ideas coming from Kazan were highly influential in Central Asia. At the same time, exchange of ideas with the Ottoman Empire and the Crimean Tatars never ceased.

In the Soviet Union, the Tatars from the Volga region attained their own autonomous republic, though not a separate Soviet republic. Today, the Tatars in that region, who make up about half of the population of Tatarstan, are trying to establish greater autonomy. Ethnically, the modern Tatars are a mixture of Turkic, Mongolic and other elements. The Misher Tatars are a distinct group with a strong Finno-Ugric element. In the centre of Tatar territory, we find the main body of the Tatars, often called Kazan Tatars. Theirs is a flourishing literary language. Further east we find scattered groups of Tatars in western Siberia, along the Tara, Tobol, Ishim and Irtysh rivers and in the Baraba steppe, as well as in the region of Tomsk. Another Tatar people, the Crimean Tatars, will be discussed below.

The Bashkirs, who live further to the north on the southern slopes of the Urals, are often presented as a people that is hardly distinguishable from the Tatars, and it is believed that the Russians more or less artificially divided the Tatars and Bashkirs, who attained their own autonomous republic with Ufa, the former Orenburg, as its capital. However, although the Bashkirs have over the centuries been under strong Tatar cultural influence, they have a strong sense of their own ethnic identity. The Bashkirs form a minority in their own republic, about one quarter of the population.

Chuvash

Another Turkic people in the Volga area are the Chuvash, who, like the Tatars, regard themselves as descendants of the Volga Bulgars in the historical and cultural sense. It is clear that Chuvash belongs to the Oghur branch of Turkic (see pp. 81–2), as the Volga Bulgars did, but no direct evidence for diachronic development between the two has been established. As there were several distinct Oghur languages in the Middle Ages, Volga Bulgar could represent one of these and Chuvash another (pp. 24, 85, 434).

The Russian Orthodox Chuvash have quite a different history from their neighbours, the Tatars, but they have nevertheless been in close cultural contact with them, as is clear from linguistic evidence. About half of the Chuvash live in the Chuvash Republic the capital of which is Cheboksary (Šupaškar in Chuvash) in the southwest corner of the great bend in the Volga river. According to the 1989 census, 1.8 million Chuvash lived in the USSR,

nearly all of these in the Russian Federation. About half of the Chuvash live in the Chuvash Republic, where they form 68 per cent of its population. However, slightly more than half live outside the republic, mostly in the southern districts of the Tatar Republic (135,000), the central and western districts of the Bashkir Republic (119,000), and the Samar (118,000) and Ulyanovsk oblasts (117,000).

Karaims

A small Kipchak people are the Karaims, whose most important and still inhabited settlements are in the Lithuanian town of Trakai (Troki in Polish), and in the Ukrainian town of Halič in Galicia. The Crimean peninsula originally was their centre, but the remaining Karaims do not speak their Turkic language any more. The term Karaim refers to both a people and to a religious system. Karaims are believers in the Old Testament but consider themselves to be of Turkic ethnic origin. They have traditionally used the Hebrew alphabet for writing their language, which is also used for ceremonial purposes. Their national identity includes a possible connection with the Khazars and the Khazar Khanate, the ruling house of which converted to Judaism around the year 800. The Karaim language is on the brink of extinction, as only the oldest generation still speaks it. The Lithuanian Karaim community including about 200 members is making efforts to revitalise the language. In 1997, there were about 50 speakers of the Lithuanian dialect and only 6 speakers of the Halič dialect. A few Karaims still having some knowledge of the language live in Poland.

Crimean Tatars

As their name indicates, the Crimean Tatars lived in the Crimea until World War II. The entire nation was deported to Central Asia by Stalin. Today, most of the Crimean Tatars, numbering almost half a million, live in rural areas of Uzbekistan, where many of them seem to be undergoing a process of language loss. Some Crimean Tatars live in the Crimea, and many attempt to return there. Crimean language use, as reported from the time before the exile, shows complex patterns of interaction between varieties which had been influenced in different degrees by Ottoman Turkish.

Uzbeks

If we turn our attention to Central Asia, a large part of Transoxania (beyond the river Oxus, i.e. Amu Darya) is covered by the Republic of Uzbekistan, where most of the Uzbeks live. First-language speakers of Uzbek number a total of about 20 million: 70 per cent of the population of the Republic of Uzbekistan (16 million), and smaller groups in adjacent areas of Northern Afghanistan (1.5 million), Tajikistan (1.5 million), Kirghizstan (600,000),

Kazakhstan (350,000), Turkmenistan (350,000) and Xinjiang (a small group of 15,000).

As contact with other languages has been, and still is, important for a number of features of modern Uzbek, it is important to mention the other major language groups in Uzbekistan: out of the total 23 million of its inhabitants, 5 per cent speak Tajik, a northern Iranian language, mainly in the oases of Bukhara and Samarkand, and in the Ferghana valley to the east, 8 per cent speak Russian, mainly in the capital Tashkent. The Turkic languages represented are from the Kipchak family: 2 per cent Karakalpak, mainly in the Karakalpak autonomous region, 4 per cent Kazakh and 2 per cent Crimean Tatar.

The language situation of Uzbek is complex for a number of reasons. In the first place, the language has inherited the mixed character of its predecessor as a literary language, Chaghatay. This is not just a linguistic matter: the complex ethnic make-up of the Uzbek nation is reflected in a much broader way.

The ethnonym Uzbek derives from the Kipchak Uzbeks who were actually the latest group to arrive in Central Asia, but who have been dominant politically ever since. The original Turkic population, that spoke a different brand of Turkic associated with the Karluk and Chighil tribes, actually sustained closer ties with the Iranian speaking Tajiks, both in the cities and in rural areas, well into this century. Nevertheless, their brand of Turkic has made the largest contribution to the structure of standard Uzbek. A third linguistic group are the so-called Oghuz Uzbeks who live in Khwarezm and in adjacent areas in Karakalpakistan and Turkmenistan. All these varieties function in relationship to one literary language, modern standard Uzbek.

The economy of Uzbekistan in the Soviet era was based on the production of cotton. The republic has inherited severe problems connected with the monoculture of this crop: problems with the management of water – the huge Lake Aral has almost completely dried up – and health problems in rural areas caused by the widespread use of insecticides.

Uyghurs and Other Groups in China

The number of persons who use (modern) Uyghur as their first language is estimated to lie somewhere between 7 and 10 million. The vast majority of these live in the oases of the Chinese-administered Xinjiang Uyghur Autonomous Region, also known as Eastern Turkestan, where their former ethnic and linguistic dominance is now being challenged as a result of large-scale immigration of non-Turkic people from eastern China. In reaction to this, there is among the Uyghur a tendency to emphasise their social and cultural identity, which finds its expression, among other things, in their preference for the Arabic alphabet. Sizeable Uyghur groups are also found in Kazakhstan, Kirghizstan and Uzbekistan. The Uyghurs, historically,

culturally and linguistically, are closely related to the Uzbeks.

A very different group are the Yellow Uyghur or Yughur in the Hexi corridor of China's Gansu province. This small group of approximately 12,000 people, who are traditionally Buddhist in faith, are almost equally divided into speakers of Turkic, Mongolic and Chinese. Turkic Yellow Uyghur is spoken in the western part of this region.

Most of the Salar, approximately 88,000 in the 1990 census, are Muslim and inhabit farming communities along the Yellow River, in the Xunhua Autonomous County and in the Hualong Hui Autonomous County in Gansu. Smaller Salar-speaking enclaves are found in Xinjiang, approximately 3,700.

Kazakhs and Karakalpaks

The northwestern part of Uzbekistan, to the east of the Amu Darya, and to the south of Lake Aral, is formed by Karakalpakistan, formerly an autonomous Soviet republic, nowadays an area with a special status within the Uzbek republic. This region is more than 30 per cent inhabited by Karakalpaks, who are, as has been mentioned above, closely related to the Kazakhs.

The main body of the Kazakhs live in the vast territory of the Republic of Kazakhstan in the steppe belt between the Caspian Sea and the Altay area, with a higher degree of density in the southern parts. The Kazakhs make up roughly half of the population of the republic. In the northern and northeastern parts of the country we find Russian and Ukrainian settlers as well as so-called Volga Germans and other groups who were exiled there in Stalin's time. The Kazakhs were Islamicised relatively late, something that was mainly achieved by Tatar missionary work during the last century. Of the large Turkic peoples, the Kazakh, together with the Turkmen, have retained longest the traditional socio-economic way of life as nomadic herdsmen.

Considerable Kazakh minorities can be found in the northern regions of the Chinese province of Xinjiang (more than one million), in the Russian Federation and in Uzbekistan. Smaller minorities live in western Mongolia and Turkmenistan. Since the independence of Kazakhstan in 1991, a trend can be noticed of Kazakhs leaving their diaspora countries, especially Mongolia, to settle in Kazakhstan, where they take the places left behind by Russians and Germans leaving the republic. Although the demographic situation in Kazakhstan seems to imply otherwise, there are reports about a rapid process of shift to Russian among the Kazakhs. This means that in spite of the changes in the population structure in favour of the Kazakhs, the process of regression of the Kazakh language has not yet come to an end.

Kirghiz

Most of the approximately 3 million Kirghiz live in the independent Republic of Kirghizstan (about 2.5 million), Uzbekistan and Tajikistan. More than 100,000 Kirghiz live in the Autonomous District of Qizilsu in the far west of

the Chinese Xinjiang Uyghur Autonomous Region. A small group called Fu-yü "Kirghiz" who were deported to Manchuria in the seventeenth century managed to retain their Turkic language well into this century. Their language seems to originate from South Siberia, and their ethnic ties with the Kirghiz are unclear (see p. 83). Another small but nevertheless famous Kirghiz community from the Pamir mountains in Northern Afghanistan fled to Pakistan and eventually settled in the Van area of Turkey.

Peoples of South Siberia

A number of small Turkic-speaking groups live in the Altay Mountains and neighbouring parts of South Siberia.

The Altay Turks, formerly called Oyrot, live in the Altay Mountains. Their southern tribes are pastoral nomads and show close connections with the Kazakhs and Kirghiz. The northern ones are basically forest hunters and have, like most other tribes in South Siberia, a different ethnic composition.

In the Novokuznetsk region and the southern region of Krasnoyarsk district, we find Turkic groups which until recently have absorbed Yeniseyic, Samoyedic and other elements. The ethnonym Khakas, also the name of one of the two literary languages in the area, was adopted by the local intelligentsia after 1917. The other language, Shor, is presently one of the endangered Turkic tongues.

Small groups of the so-called Chulym Turks have lived long on the Chulym river, but their language and culture now seem to be extinct.

The Tuvans mainly live on the slopes of the Sayan Mountains, but also in the Altay Mountains and northern Xinjiang. At least those living in the Autonomous Republic of Tannu-Tuva are of Buddhist belief with a strong Shamanist substratum. Most of them are sedentary cattle breeders. Their closest relatives are the Tofa, previously called Karagas, living on the northeastern slopes of the Sayan Mountains, and the small group of the Toja north of them.

Yakuts and Dolgans

The Yakuts (self-denomination: Sakha) is the Turkic nation that lives farthest to the east; they comprise about 365,000 persons, mainly in Yakutia, east Siberia. The capital is Yakutsk on the Lena. Some groups of Yakuts live in other territories, e.g. in South Taymyr, Sakhalin, on the Amur, etc.

The earliest history of the Yakut people is little known. It seems clear that they did not leave their southern homeland – after which they lost contact with the other Turkic tribes – before the thirteenth century. Most scholars consider the fifteenth or sixteenth century as the period of the Yakut migration northwards.

It is beyond question that the modern Yakut people consists of diverse ethnic elements. In the sixteenth century, the ancestors of the Yakuts came to

their new homeland, where they met other ethnic groups. The most important of these for the further development of the Yakut people and language were Tungusic tribes. The other main substrate group was probably formed by Yeniseyic tribes.

Two facts made the Yakuts superior to all ethnic groups they encountered in Siberia: cattle and horse breeding and production of iron. Even the Russian conquest in the seventeenth century did not change the ethnic hierarchy on the Lena. It was only in the 1920s that the Russians began to dominate in the region. The small group of the Dolgans, whose language is close to Yakut, comprises about 5,000 speakers.

Unity and Diversity

This short overview should have convinced the reader that the Turkic peoples are spread out over a vast territory, in greatly varying surroundings. They follow different religions and are in very different stages of economic development, degree of urbanisation, etc. The peoples in their vicinity that dominate them culturally, or are dominated by them, vary enormously from place to place.

This being the case, surprisingly naive ideas about the unity of the Turkic peoples are still entertained by many. In respectable Turkish newspapers one can often read articles about the '165 million people all over Eurasia who speak Turkish'. In a sense, these are echoes of the ideology of Pan-Turkism, an imitation of and reaction to the ideology of Pan-Slavism. Indeed, in the Russian Empire and later in the Soviet Union the 'Centre' was always very concerned about the threat posed by the mass of Muslim subjects living in the southern belt of the empire, and in the context of Russia/the Soviet Union, 'Muslim' naturally was practically synonymous with (Muslim) Turkic. The prospect of unifying the Turkic speakers in one empire has had a superficial attraction for some Turkish and Tatar intellectuals, but has never been popular among the peoples concerned.

It seems natural that old regional and super-regional patterns of cultural and economic contact will establish themselves with frontiers that are so much more open than before. Ever since the establishment of the independent Turkic-speaking republics, Turkey has been actively trying to establish cultural and economic links with them. The European parts and Kazakhstan naturally look more to the west for cooperation, and the eastern parts have close contacts with the Asian industrial nations. Republics such as Turkmenistan, Uzbekistan and Kirghizstan are more isolated.

One field in which the Turkic peoples can and should be studied as a whole quite obviously is linguistics, and the present work is an attempt to strengthen further this line of research. There are, however, other elements of a common cultural heritage shared by the Turkic peoples, the layers of which can notably be found in oral literature, and certain traditions shared by peoples geo-

graphically far apart. The newly regained freedom of so many Turkic peoples has brought them to the attention of a broader public in the west. It is hoped that the present book, even if it concerns a more or less specialist field, might offer a contribution to help remove the still prevailing ignorance concerning the Turkic-speaking world.

Table 1.1 Speakers of Turkic languages

Country	Total number of inhabitants	Turkic language/people	Estimated number of speakers		
Afghanistan	17 million	Uzbek	1.4 million		
		Turkmen	380,000		
		Kazakh	2,000		
		Karakalpak	2,000		
		Kirghiz	500		
		Afshar	45,000		
Armenia	3.7 million	Azerbaijani	40,000		
Australia	18 million	Turkish	40,000		
Azerbaijan	7.5 million	Azerbaijani	6 million		
Bulgaria	8.5 million	Turkish	approx. 1 million		
		Tatar	11,000		
China	1,175 million	Gagauz	approx. 5,000		
		Uyghur	more than 7 million		
		Kazakh	more than 1 million		
		Kirghiz	140,000		
		Salar	approx. 74,000		
		Uzbek	15,000		
		Yellow Uyghur	5,000		
		Tuvan	400		
		Cyprus	700,000	Turkish	140,000
Greece	10.3 million	Turkish	30,000		
Georgia	5.6 million	Azerbaijani	300,000		
Iraq	20 million	'Iraq Turkmens'	approx. 400,000		
Iran	62 million	Azerbaijani	13 million		
		Kashkay	570,000		
		Khorasan Turkic	400,000		
		Turkmen	500,000		
		Khalaj	28,000		
		Kazakhstan	17 million	Kazakh	7.3 million
		Uzbek	350,000		
		Tatar	340,000		
		Uyghur	245,000		
		Chuvash	23,000		
Kirghizstan	4.6 million	Gagauz	1,000		
		Kirghiz	2.4 million		
		Uzbek	600,000		
Lithuania	3.7 million	Karaim	50		
Macedonia	2.2 million	Turkish	80,000		
Moldova	4.5 million	Gagauz	150,000		

Table 1.1 Continued

Country	Total number of inhabitants	Turkic language/people	Estimated number of speakers
Mongolia	2.5 million	Kazakh Uyghur Tuvianian	100,000 1,000 6,000
Poland	39 million	Karaim	approx. 20
Romania	23 million	Tatar Turkish Gagauz	24,000 approx. 23,000 not known
Russian Federation Altay region and South Siberia	148 million	Oyrot, Teleut Khakas Shor Tuva Yakut Dolgan	52,000 58,000 10,000 200,000 400,000 5,000
Sakha		Chuvash	900,000
Chuvashia		Tatar	2 million
Tatarstan		Chuvash	135,000
Bashkiria		Bashkir Tatar	1 million 1 million
Caucasus		Chuvash Kumyk Noghay Karachay Balkar Gagauz	119,000 300,000 77,000 70,000 40,000 10,000
Tajikistan	5.8 million	Uzbek	1.4 million
Turkey	65 million	Turkish	more than 50 million
Turkmenistan	4 million	Turkmen Uzbek Kazakh	3 million 350,000 80,000
Ukraine	51.7 million	Karaim Gagauz	6 32,000
Crimea		Crimean Tatar	200,000
Uzbekistan	22 million	Uzbek Karakalpak Crimean Tatar Kazakh Chuvash	16 million 450,000 300,000 900,000 9,000
Western Europe		Turkish	more than 2 million
Yugoslavia	10.8 million	Turkish	approx. 20,000

References and Further Reading

- Akner, S. (1986) *The Islamic Peoples of the Soviet Union*, London: Routledge & Kegan Paul.
- Andrews, P. (1989) 'Ethnic Groups in the Republic of Turkey', in B. Reihle (ed.) *Beihefte zum Tübinger Atlas des Vorderen Orients*, 60, Wiesbaden: Reichert.
- Bainbridge, M. (1993) *The Turkic Peoples of the World*, London, New York: Routledge & Kegan Paul.
- Comrie, B. (1981) *The Languages of the Soviet Union*, Cambridge: Cambridge University Press.
- Forsyth, J. (1992) *A History of the Peoples of Siberia*, Cambridge: Cambridge University Press.
- Golden, P. (1992) *An Introduction to the History of the Turkic Peoples*, *Turcologica* 9, Wiesbaden: Harrassowitz.
- Johanson, L. (1988) 'Grenzen der Turcia. Verbindendes und Trennendes in der Entwicklung der Türkvölker', in *Turcica et Orientalia*, Stockholm: Swedish Research Institute in Istanbul, 51-61.
- Menges, K. H. (1995) *The Turkic Languages and Peoples. An Introduction to Turkic Studies*, Veröffentlichungen der Societas Uralo-Altaica 42, Wiesbaden: Harrassowitz.

2 *The Turkic Peoples: A Historical Sketch*

Peter B. Golden

Origins and Early Notices

The modern Turkic peoples descend from an ancient grouping of tribes who through conquest, interaction and assimilation extended its language and elements of its culture across Eurasia. Their origins and earliest history can be reconstructed only with difficulty. Turkic belongs to the Altaic language 'family'. The nature of this relationship (see p. 77), whether genetic or the result of long-standing interaction, is much debated. Recent research posits an Altaic 'community' – whatever its origins – c.4000–3000 BC, from which Ancient Turkic emerged c.3000–500 BC. On the basis of ancient borrowings from Uralic and Indo-European, it would appear that the Turkic-speaking grouping was the westernmost of the Altaic family. The location of this earliest Turkic *urheimat* (original habitat) is unclear. Since Indo-Europeans appear to have pioneered the development of equestrian-based pastoral nomadism in the fourth to third millennia BC, the ancestors of the earliest Turkic peoples must have inhabited regions where this technology could be transferred to them, i.e. the forest–steppe zone that ringed the Eurasian plains from the north. This original habitat has been identified with southern Siberia, from the Yenisey to the Pacific (in particular the Altay region or Trans-Baikalia), or with the trans-Caspian zone. In the course of the first millennium BC, the bearers of Proto-Turkic spread over hitherto predominantly Indo-European (Iranian) central and inner Mongolia. It was in the steppe that this equestrian, pastoral nomadic culture matured, most probably in further contact with Indo-European (Iranian) pastoral nomads.

Nomadism was one of the determining forces in Turkic history. Evolving out of animal husbandry specialisations adapted to ecological niches that were marginal to settled society, Eurasian nomadism required interaction with the sedentary world. Given their mobility, easily translated into military advantage, the nomads raided or traded with agrarian-urban society depending on which of these strategies brought the greater reward and least risk. Eurasian nomadic political structures were equally fluid, moving from loosely held tribal unions to conquest states, as opportunities and outside stimuli presented themselves. Egalitarian in principle, nomads did not willingly submit to

strong central authority and a state structure. Charismatic warlords and dynamic conquerors were occasionally able to impose these institutions on them. More often, however, nomads, having conquered already existing states, were adapted to them (not without a struggle) and in time largely sedentarised. Nomads did not necessarily seek to conquer sedentary society, rather they sought regular access to its goods and products.

The ancestors of the Turkic-speaking peoples were, undoubtedly, part of the Hsiung-nu/Asian Hunnic union that began to trouble the borders of China in the third century BC. The ethno-linguistic affiliations of the Hsiung-nu *per se* – Iranian, Palaeosiberian, Altaic – remain unclear. The Hsiung-nu polity, like many other Inner Asian confederations formed through tribal superstratification, was polyethnic and polyglot. When it began to break up after suffering a number of serious defeats at the hands of Han China and its steppe allies (first and second centuries AD), elements drifted towards western Eurasia, tacking on new ethnic groupings as they moved. These became the Hyôn/Xiyon of the Iranian borderlands and the Huns who, after crossing the Volga c.AD 370, proved to be disquieting neighbours of the Roman Empire.

The Hsiung-nu were organised as an 'imperial confederacy', which presented itself as a centralised state in its dealings with the outer world while remaining more loosely structured internally. Their supreme leader was the Shan-yü (= *tarxan* or *jabyu/yabyu*), a heavenly mandated ruler, whose power and authority in the political, judicial and spiritual spheres was much like that of the kaghans and khans of later Turko-Mongolian states. The overall structure of this state with its ruling charismatic clan, division into left and right wings, decimal military structure, Tengri (sky-god) cult and imperial ideology of heavenly mandated rule, established patterns which thereafter would be followed by virtually all of the Eurasian Turkic states and advanced tribal unions.

Hsiung-nu military activities touched off major migrations – such as that of the Indo-European Yüeh-chih c.176 BC – of inner Asian peoples to central and western Eurasia, and undoubtedly brought about dislocations and movements among the Turkic peoples as well. Hsiung-nu dominance was followed by that of the Proto-Mongolic Hsien-pi confederations, one of which, the Tabyač – the T'o-pa or Northern Wei dynasty, AD 386–534 – gained control over much of northern China. Although Turkic elements were present in some of these statelets, they were not politically predominant. When the Türk people are first clearly identified in our sources, in the sixth century AD, it is within the context of relations between Tabyač China and another Hsien-pi state, the Jou-jan/Avar Kaghanate in Mongolia. It is interesting to note, however, that the first clearly attested Turkic peoples appear before this in the western Eurasian steppes.

The European Huns, whose relationship, ethnically and linguistically, to the Asian Hsiung-nu remains problematic, arrived at the Volga c.AD 350, driving before them Turkic and other groupings into the Ponto-Caspian

steppes. Iranian, Germanic and other ethnic elements were added to their union. After a terrifying raid, in AD 395, into the Sāsānid and Roman empires via the Caucasus, they settled into the familiar pattern of raiding and mercenary soldiering for the two empires, exploiting their mutual enmity and domestic weaknesses. This, again, was a pattern that would be typical of the Turkic nomads who, from this point onwards, replaced the Iranian tribes (Scythian, Sarmatian, Alanic) as the predominant political element in the western Eurasian steppes. It was under Attila that Hunnic raiding, based in Pannonia, began to pose a serious problem for Rome. Defeated in Gaul in AD 451 and repulsed in Italy the next year, Attila died in AD 453 and his 'state' immediately crumbled in the wake of rebellions by subject tribes. The Huns now faded into the background, becoming part of the various Turkic tribal unions that were entering the steppe. The Oghuric Turkic peoples (**Šarayur*, *Oγur*, *Onoγur*), who had formed distinct units in the territory of today's Kazakhstan and speaking a form of Turkic that already displayed a number of important divergences from Common Turkic (e.g. the *r-z* and *l-š* shifts; Chapter 4), appeared in the Ponto-Caspian steppes c.463. They had been driven thither by the Sabirs and ultimately by pressures emanating from the Avar/Jou-jan empire. These Oghuric tribes appear to have belonged to a larger confederacy called in the Chinese sources the Ting-ling, later the T'ieh-lê, which included many of the Turkic peoples that subsequently formed part of the Türk Kaghanate. In the following decades, the western Eurasian steppes appear to have been populated by a bewildering array of tribes (Kutrigurs, Utigurs and others). Some established alliances or ententes with the Byzantine Empire while others engaged in predatory raids into imperial territories or did both. Constantinople responded with diplomacy, attempting to convert to Orthodox Christianity those 'Huns' that were closest to imperial territories – there were Byzantine outposts in the Crimea – and seeking to foster inter-tribal rivalries elsewhere.

The Bulghar tribal union, first noted as an ally of Byzantium in AD 480, was a fusion of Turkic Oghur (especially Onoghuric) and Hunnic elements organised under a ruling house that claimed descent from the Attilids. By AD 515, the Sabirs, a Turkic people of undetermined type, were in the Volga steppe zone and had established relations with both the Byzantine and Sāsānid Empires and occasionally raided Transcaucasia and Anatolia. The Ponto-Caspian nomads were overrun in the late 550s by the Avars, derived, it seems, from elements of the fugitive Jou-jan or Hephthalites – both states were based on older War (Proto-Mongolic Hsien-pi) and Hunnic groupings – , who came westward after their overthrow by the Türks in 552 and 557 respectively. The Türks soon arrived, driving the European Avars and some of their subject tribes into Pannonia. It is at this point that the various strands of Turkic history clearly come together in the form of the first Türk Kaghanate.

The Türk Kaghanate

In AD 552, Bumīn of the A-shih-na, leader of the Türk confederation, toppled the Jou-jan Kaghanate in Mongolia and proclaimed himself *Ilig Qayan*, that is 'polity-/realm-possessing (i.e. ruling) Emperor', in Soghdian *bḡ βwmyñ ḡḡ'n* 'Lord Bumīn Kaghan'. The origins and history of this tribe or tribal union before this signal event are little known. Attempts to identify them with graphically similar ethnonyms in ancient Greek and Semitic sources are highly speculative. The ethnonym, which has yet to be etymologised successfully, appears in a variety of sources: Chinese *T'u-chüe* = **Türküt*, perhaps taken from Soghdian *twrk* (*Türk*), *twrkt*, *twrkw* etc. (*Türkit*, *Türküt*), Tibetan *dru-gu*, Khotanese *ttrukä*, *tturka*, *tturki* (*Türk*), Greek *Τούρκος* (*Tourkos*), Arabic *T'rk*, plural *Atrāk*. Róna-Tas, building on older theories, has suggested that *Türk*, which often appears in a hendiadys *erk türk* ('power/strength' + 'period of prime of one's power') was an *epitheton ornans* of the ruling clan which meant 'strength and authority reached by somebody when he is at biological maturity' and thence 'the mighty, those who have the authority, the strength...'. The fact that a number of the personal names of the early Türk kaghans appear not to be Turkic (e.g. Bumīn, Ištemi, Muqan, Taspar, **Nivar/Neber/Nébar*), may point to the non-Turkic origins of the A-shih-na. This latter name, for example, Beckwith has compared with *Ἀρσίλας* (*Arsilas*) (**Aršila*) of Menander, which he associates with the Tokharian title *Āršilānci*. Intimate ties with the Indo-European Tokharians of Eastern Turkistan are attested by a wide variety of borrowings from very early periods. Tokharian titles and ideological formulations (e.g. Old Turkic *teḡri teg teḡride bolmiš* 'heaven-like, heaven-conceived' referring to the kaghan) also had a clear influence on Türk usage.

The ethnogenetic legends reported in the Chinese sources associate the Türks with the Hsiung-nu and hint at migrations that brought them to southern Siberia/northern Mongolia in the mid-fifth century AD. They came from the Kansu region to where they had moved, c.265, as a consequence of buffetings in the Hsiung-nu union. In the southern Altay, they became the blacksmiths of the Jou-jan. With the establishment of their kaghanate, Turkic now became the predominant linguistic element in Mongolia and the steppelands around and in what is now Turkistan and extending into the Pontic zone, supplanting hitherto dominant Indo-European (c.1000 BC–AD 500).

Having destroyed their erstwhile overlords, the Türks, from their holy ground and capital on the Orkhon, set about establishing their hegemony over the Silk Route. Under the Yabghu Kaghan (*yabghu qayan*) Ištemi (552–575/576), the brother of Bumīn, central and western Eurasia were conquered, extending Türk dominion to the borders of the Sāsānid sphere of influence in Transcaucasia and the Byzantine territories in the Crimea. Diplomatic and commercial relations with these imperial neighbours were established. Meanwhile, Bumīn's son Muqan (Chinese Mu-han, reigned 553–572) 'subjugated all the states beyond the border of China'. These

conquests brought the Soghdians, whose commercial diasporan colonies could be found along the Silk Route, into the Türk state where they played an important role as both merchants and functionaries in the administrative apparatus. This joining of Turkic warrior and Iranian bureaucrat became a common feature of many subsequent Turkic states. The nomad-based empires, controlling and exploiting important trade routes, had great need of specialists from the sedentary world. The Kaghanate was organised as a type of dual kingship, common to many Turkic polities, with the supreme rulership residing in the eastern half. The Kaghan was viewed as a sacral figure who ruled by heavenly mandate and by the possession of heavenly good fortune (*qut*). In theory, his blood could not be shed.

The Western Kaghanate, led by the Yabghu Kaghan, would, in time, seek to assert its independence. The Türk state – and many subsequent Turkic polities – adhered to certain notions of the collective sovereignty of the ruling clan over the empire. In principle, any male of the ruling, charismatic clan capable of enforcing his claim to power could rule. Although clear lines of succession were enunciated as the ideal, ‘bloody tanistry’ was frequently the result. This rough school of politics often produced very able rulers. But, it just as often led to instability the manifestations of which were felt well into the Ottoman era.

The pro-Buddhist Taspar (572–581) was the last ruler of a united empire. Caught up in continuing internecine strife within the extended ruling clan, the Eastern Kaghanate (630) fell to the T’ang dynasty (618–906) of China. The Western Kaghanate, which had fissured into two competing power blocs, the Tu-lu and Nu-shih-pi of the Chinese sources, termed the *On Oq* ‘Ten Arrows’ confederation, was brought under Chinese control by AD 659. By this time, the Türk-ruled Khazar Kaghanate and Bulghar confederation in western Eurasia had struck out on their own. Conflict between the two led to the defeat of the Bulghar union and the migration of elements of their confederation to Danubian Europe and the Balkans. Here, in 679, the Balkan Bulghar state took shape. The Khazars, thus, were the successor state of the western Türks in the Ponto–Caspian steppes.

The eastern Kaghanate was revived through the heroic efforts of a scion of the A-shih-na, Qutluy (throne-name Ilteriş, reigned 682–691). These and the deeds of his immediate successors were recorded in the Orkhon inscriptions (Chapter 4). His brother Qapayan (691–716) is said to have ‘made the poor rich and the few many’. The second Türk Kaghanate, however, was constantly at war with its recalcitrant subjects. In 741, a revolt of subject peoples toppled the dynasty and briefly brought the Basmil to power supported by the Uyghurs (*uyğur*) and Karluks (*qarluq*). In 744, the Uyghurs seized supreme power in the eastern Kaghanate – driving their erstwhile allies, the Karluks, westward into central Eurasia – and ushered in a new age of closer, less inimical ties with China. The Uyghurs also fell heir to the culture of the Türks. This included the runiform alphabet used in the Orkhon inscriptions as well

as a cursive script. Both were derived from the Soghdian alphabet system – although this is less clear with the runiform script – , ultimately going back to Eastern Aramaic/Syriac script systems (Chapter 6). The Uyghur alphabet was later adopted by the Mongols and through them by the Manchus. The spread of writing systems was intimately intertwined with the movement of religions. Here again, the Soghdians and to a lesser extent the Tokharians played a major role. Thus, in addition to indigenous shamanism, the *Tegri* (heaven-/sky-god) and other cults, Mazdaist, Buddhist and foreign religious influences may be discerned.

The Western Kaghanate, although plagued by internecine strife, provided a spirited challenge to the Arabo-Islamic advance, which had conquered much of Iranian Central Asia by the early eighth century, and the Tibetans who were expanding into Eastern Turkistan. In 751, near the Talas river in Kazakhstan, a T’ang army which had been imposing its will on the *Türgeš*, the leading grouping of the On Oq, was defeated by a Muslim force, aided by the defection of the Karluks. This clearly delineated the Iranian Central Asian territories and their steppe borderlands as within the Islamic orbit. The Karluks benefited as well, gaining possession, by 766, of the western Türk lands and laying one of the foundation stones of the Karakhanid state.

Successor States of the Türks

The Uyghur kaghans, from their centre on the Selenga river in Mongolia, vied with Tibet over Eastern Turkistan and aided by their Soghdian advisers, skilfully exploited China’s domestic weakness caused by the revolt of An Lu-shan in 755. Periodically called in to suppress rebels, the nomads became one of the props of the T’ang. In return for this they were richly rewarded and granted most favourable trading privileges. In 762, the Uyghur ruler, Böğü Kaghan, under the influence of his Soghdian courtiers, converted to Manichaeism. Here, too, we see another, subsequently familiar pattern in Irano–Turkic politico–cultural relations: Iranian intermediaries in the proselytising of a Near-Eastern religion. The Uyghur Kaghanate became a powerful and highly cultured state. In 840, weakened by domestic strife, it was overrun by the Kirghiz (*qirγiz*), a Turkic or Turkicised people of still uncertain ethnic antecedents whose primary habitat was in the Yenisey region. The Uyghurs dispersed to the Chinese borderlands, ultimately creating several statelets in East Turkistan (hitherto Tocharian and Iranian) and Kansu.

For the Kirghiz, who remained largely in their Siberian home, Mongolia, the traditional centre of nomadic empires in inner Asia, did not appear to hold the same attractions. With the Uyghurs gone, the way was open for the movement in of new, largely Mongolic-speaking tribes hitherto located in eastern Mongolia and adjoining regions. Some of them had already been part of the Türk and Uyghur states. In 924, one of these tribal confederations, the Qitan – already a power in the Sino–Manchurian borderlands as the Liao

dynasty (907–1125) – marched into the old Uyghur capital, reclaiming it in the name of traditional nomadic conquerors and even offering to allow the Uyghurs to return to their former abodes (as subjects). The latter, however, already sedentarising in East Turkistan, declined. The Qitan movements appear to have led to the Mongolisation of Mongolia.

In western Eurasia, the Khazar Kaghanate, a vast empire comprising Common Turkic and Oghuric, as well as Iranian, Finno-Ugric and other elements, and ultimately extending from the middle Volga to the north Caucasus and from Kiev to the Khwarezmian deserts, emerged as the successor state of the Türks c.650–680. Like the Turks of central Eurasia, they were engaged in protracted warfare with the advancing Arabs. This, combined with the dictates of geopolitics and older Türk traditions, led to a Byzantine–Khazar entente. Although the Arabs compelled the Khazar Kaghan, in 737, to convert, briefly, to Islam, they were not able to extend their power beyond Bāb al-Abwāb/Darband in the north Caucasus. In the late eighth to early ninth century, the ruling house converted to Judaism. This little influenced the Türk-derived Khazar state structure and political culture. Here, however, the Kaghan became a completely sacralised figure and the everyday affairs of state were handled by a sub-Kaghan (*qayan-beg*, *šad* or *yilig*). Khazaria grew wealthy from the east–west Islamic trade. A decline in revenue, growing sedentarisation and the pressure of new opponents brought about its collapse in 965 at the hands of the Rus' allied with elements of the Western Oghuz.

Their immediate successors in the western Eurasian steppes were the Pechenegs. Often in conflict with the equally aggressive Rus', they were defeated by the latter in 1036 and largely pushed into Danubian Europe and the Byzantine Balkans. The Torks/Western Oghuz who followed them suffered a similar fate in 1060. More spirited opposition came with the advent of the Kipchaks (*qipčaq*) whose origins are to be sought in the Kimek confederation and a series of movements of peoples in inner and northern central Eurasia. The Kuman–Kipchak confederation dominated the steppes from the Danube to Khwarezm and western Siberia, interacting with Rus', Byzantium, Hungary, Georgia, the Islamic lands – the Kipchak steppe having become a major source of ghulams (*yulām*), caliphal military slaves – both in the Middle East and the Khwārazmšāh state. Facing sedentary states that were either too weak or divided to engage in prolonged wars of conquest with them, the Kipchaks, like the Pechenegs, never formed a state but were able to survive and retain their nomadic economic system by integrating themselves into the larger state-system through adroit diplomacy and by providing professional soldiers.

The early foes of the Khazars, the Bulghar union, divided in the latter half of the seventh century. One grouping, after 679, under Asparukh settled in the Balkans, imposing themselves on an already existing Slavic tribal union. Here, they came to be deadly foes of the Byzantine Empire. Heirs to both the Attilid and Türk traditions, the Bulghar rulers styled themselves Kaghans and

continued to use the twelve-year animal cycle calendar of inner Asia. In 864, under Byzantine pressure, the Balkan Bulgars converted to Orthodox Christianity. Already Slavifying by this time, Bulgaria became one of the pioneer centres of Orthodox Christian Slavic culture.

The Volga Bulgars derived from the same tribal confederation defeated by the Khazars in the seventh century. They moved to the middle Volga region by the mid-eighth or early ninth century. Here, they played a prominent role in the fur trade and later rivalled the Khazar capital, Atil/Itil, their overlord, in trade with the Islamic lands. Islam came, in some measure as a counterbalance against Khazaria, in the early tenth century resulting in a developed Islamic urban culture.

The Turkic Peoples and Islam

After the Qitan 'conquest' of Mongolia, the focus of Turkic history turned to the Islamic steppe borderlands of central Eurasia. Here, a new Irano-Islamic culture was taking shape under Iranian service dynasties of the 'Abbāsīd Caliphs: the Tāhīrīds (821–873), governors of Khorasan (Xurāsān), and the Sāmānīds (819–1005), governors of Transoxania (Mā warā'n-nahr). Soghdian was being replaced by Persian (Fārsī), influenced by Islamic culture and written in the Arabic script. Islamic Neo-Persian literature was pioneered in Central Asia. Persian speech was associated with Islamic culture; the Iranian population came to be known as *Tājīk* (< Arab tribal name *Tayyi'* + *-čik*). Islam here, although nominally Sunnī, took on an Iranian cast, occasionally with a radical religious and social agenda. This was the Islam, in its folk form, that was brought to the Turkic peoples. It ultimately led to a Turko-Persian cultural symbiosis which had a profound impact on the Islamic cultures that developed not only in Central Asia but in western – the Seljuk and Ottoman realms – and southern Asia as well. The Sāmānīds, in particular, became a major source of ghulams, who, by the mid-ninth century, began to occupy increasingly powerful military and administrative positions. Some caliphs were their virtual prisoners.

It is at this stage that the Turkic populations of central Eurasia came more fully to the attention of the Islamic sources. In addition to the tribes of the Western Türk Kaghanate, we find mention of the Karluks, possessors of the western Türk centres around İssīq Kōl, the Ili, Ču and Talas rivers, who arrived here c.745, the Oghuz, who are noted in the Syr Darya region by the late 770s, the Čigil, Tuxsi, two subconfederations of the Karluks, the Yaghma, also associated with the Karluks, camping between Lake Balkhash and the Ala Kūl, all in the immediate environs of the Muslim forts of the region. Further away were the Kimeks ranging across Kazakhstan and western Siberia, the Uyghurs in Eastern Turkistan and a number of other Turkic and Turkicising peoples. Maḥmūd al-Kāšyarī, writing in 1077, notes tribes that spoke 'pure Turkic', others whose speech was corrupted by contact with

Persian and Soghdian speakers in the cities and yet others who spoke languages other than Turkic or Iranian – perhaps Mongolic, Palaeo-Siberian or Tokharian. Islam spread to the Turkic tribes, already relatively sophisticated in religious matters, not so much by the sword as by the activities of merchants and Šūfīs, many of whom were Turks. Nomads living closer to the urban regions with which they traded were converted and in turn preached Islam to their more distant kinsmen. The early converts among the Oghuz – and some Karluks – were called ‘Türkmen’ for reasons that are still obscure. The Karluks and Oghuz were ruled by yabghus – although after 840 the Karluks may have claimed the kaghanal dignity – , the Kimeks by a kaghan and others by lesser ranks. All retained some vestiges of Türk political organisation.

The Karakhanids, Ghaznavids and Seljuks

The tenth century proved to be particularly fruitful for the propagation of Islam among the Turkic peoples. In the early 920s, the Volga Bulgars (see pp. 85, 434) moved decisively towards Islam. Mass conversions of ‘Infidel Turks’ are noted in 960 and 1043. More typical than these grandiose movements were smaller-scale conversions unrecorded by the historians. The Turks were not passive in this process. The Karakhanid dynasty, of possible A-shih-na origins – but unclear affiliations, Karluk, Yaghma or perhaps Uyghur – under Satuq Buġra (d. c.955), adopted Islam and actively promoted it in the Turkic world. The term ‘Karakhanid’ is itself an artificial name developed by scholars. The self-designation was rather the ‘kaghanal (ruling house)’, or ‘khanal kings of the Turks’ (*al-xāqāniyya, al-mulūk al-xāniyya al-atrāk*). By 999, the Karakhanids had driven the Sāmānids from much of Transoxania and were in possession of Bukhara. Creators of the first Islamo-Turkic polity, they imposed a still imperfectly understood Turkic tetrarchic ruling structure – two kaghans and two sub-kaghans, divided along east–west lines – on the pre-existing Islamo-Iranian (Sāmānid) state. Their empire extended from Western Türkistan to sizeable portions of Eastern Turkistan where Kāshghar was one of their centres. Karakhanid rule brought greater numbers of Turkic tribesmen into Iranian Central Asia than ever before. At the same time, a Turkic ghulam state founded by slave-soldiers of the fading Sāmānids in 962 was established at Ghazni in Afghanistan. Aided by an Iranian bureaucracy, it ruled over an Iranian and north Indian peasant and urban populace and acquired great wealth through lucrative and destructive raids into Hindu India. These ‘Ghaznavids’, prototypes, in some respects, of the later Mamlūk regime in the Near East, duly invested in their offices with caliphal approval, became one of the greatest powers in the Islamic world. Thus, two very different Islamo-Turkic states had come into being. They were soon to be joined by a third with an even more dynamic nature: the Seljuks.

The name ‘Seljuk’ is a political rather than ethnic name. It derives from Selċük, born Toġaq Temir Yalıġ, a war-lord (*sü-başı*), from the Qınıġ tribal grouping of the Oghuz. Seljuk, in the rough and tumble of internal Oghuz politics, fled to Jand, c.985, after falling out with his overlord. Here, having converted to Islam, he ended his days as a fighter for the faith. His sons – all of whom bore Old Testament names indicating some earlier Judaic or Christian influences – became caught up in the Karakhanid and Ghaznavid rivalry. By the mid-1030s, much of the family, now under his grandsons Toġrul and Ćayrı, had been dispersed by their foes and migrated to Khorasan. When the Ghaznavid Sultān Mas‘ūd (1031–1041) sought to end their depredations in this wealthy part of his empire, he suffered an unexpected, disastrous defeat at Dandānqān (23 May 1040). The Ghaznavid hold over eastern Iran collapsed and the Seljuk-led Oghuz tribesmen became masters of the area. Seljuk authority advanced into western Transoxania and westwards towards Transcaucasia and the Byzantine border. Relations were soon established with a weak caliphate seeking a military protector that could rid them of the fading Shī‘ite Buyids who had dominated Baghdad since 945. In 1055, the Seljuks entered the ‘Abbāsīd capital. Toġrul was proclaimed sultan of the Sunnī Islamic world. He and Ćayrı – who ruled in the east, in keeping with old Turkic notions of bipartite political organisation – realising that Islam was the only ideology supporting their claims to universal rule that would be acceptable to both their nomadic followers – very much a minority – and newly conquered sedentary populations, sought now to strengthen their hold on the rich Islamic lands of the Near East. They were not always able to control the anarchic dynamism of their tribal followers. The more troublesome of these were directed to the Byzantine frontier to engage in the already centuries-old traditions of *jihād* there, while the Seljuks concentrated their primary efforts on securing the Islamic heartland.

In 1071, however, Alp Arslan, son of Ćayrı (1063–1072), who held supreme authority, in an unsought conflict inflicted a stunning defeat on the Byzantines at Manzikert. Anatolia now lay open to the Oghuz tribesmen who, independently of the dynasty, began to migrate thither. A branch of the Seljuk family established itself as the Sultanate of Rūm (= Rome = the East Roman/Byzantine empire). In this we may see the origins of Turkey. Seljuk power under Alp Arslan and his son and successor Malikšāh (1072–1092), aided by their brilliant vezir, Niẓām al-Mulġ (d. 1092) was also extended to the Karakhanid realms. The Great Seljuk state was, like the Karakhanid, a Turkic monarcho-military structure grafted onto an already existing Islamo-Iranian state with traditions extending deep into the pre-Islamic past. The dynasty, however, with its notions of collective sovereignty, faced numerous, bloody throne struggles. The nomads, never fully reconciled to a system of absolute kingship and the pressure to sedentarise, also weakened the foundations of the state. Sanġar (1118–1157), who lost Central Asia to the Kara Khitai (1141), was the last effective Great Seljuk ruler. The dynasty succumbed to the

Khwārazmshāh state in 1194. The latter, ruled by the Anušteginids, a ghulam dynasty of Oghuz origin sent by Malikshāh to rule there, replaced the Seljuks, in the course of the twelfth and early thirteenth centuries, as the major power in Western Turkistan and Iran. In Rūm, however, the Seljuks survived and in time thrived, reaching the zenith of their power under Kay Qubād (1219–1237), on the eve of the Mongol invasion of the Middle East.

The Mongol Invasions and Aftermath

The Mongol conquests launched by Chingis (Činggis) Khan (1227) and completed by his sons and grandsons had, by the middle of the thirteenth century brought virtually all of the Turkic world – as well as states such as China, Russia and Iran – from inner Asia and Siberia to western Eurasia and the Near East under their control. The Seljuks of Rūm, conquered in 1237, became tributaries and following an abortive revolt in 1277 were fully absorbed into the Il-Khanid realm, the Chingisid subordinate *ulus* ('appanage') centred in Iran, which had destroyed the 'Abbāsīd Caliphate in 1258. In the Near East, the Mongols were only stopped by their own internal problems and the resistance of the largely Kipchak Mamlūks.

The demographic fallout of these conquests in the Turkic world was enormous. Large numbers of central Asian Oghuz tribesmen, as well as many other Turkic groupings, entered the Middle East, swelling the ranks of those that had come here in the Seljuk era. In Iranian Central Asia the consequences were just as dramatic. The number of inner Asian nomads – overwhelmingly Turkic as were most of the 'Mongol' armies – entering the region surpassed the total number of those that had, over the centuries, preceded them, making certain the Turkicisation of central Eurasia, Anatolia and northern Iran. In the process, many of the old Turkic confederations, e.g. the Kipchaks, were broken up and dispersed giving rise to new units, armies based on now dispersed tribes and clans. These became the nuclei of the modern Turkic peoples. Some of these groupings took the name of the founding *xan* as their political name. In eastern Europe–western Eurasia (Ulus of Joči), Central Asia (Ulus of Čayataj) and in the Near East (the Il-Khanid state), the Turkic element prevailed linguistically, absorbing the Mongol minority. In the course of the first half of the fourteenth century, the Turko-Persian Islamic culture of these regions prevailed as well.

In the Near East, the last effective Il-Khanid ruler, Abu Sa'īd, died in 1335. By that time, however, dynamic Turkic statelets (*beyliks*) had formed on the frontier of the Turko-Islamic and Byzantine worlds in Asia Minor. One of these was led by Osman (1324) whose followers were known as Ottomans (*Osmanlı*). Advantageously located on the Byzantine border for the waging of *ğazā* (military raids in the name of Islam), Ottoman successes attracted *ğāzīs* from other regions. In 1352, the Ottomans penetrated Europe – the isthmus of Gallipoli – and, by 1400, much of the Balkans and Asia Minor had

come under their domination. It was in the latter region that they came into conflict with Tamerlane (Turkish Aqsaq Temür, Persian Tīmūr-i Lang), the great Central Asian conqueror who, starting from his base in the Chaghatay Khanate, had sought to re-establish the Chingisid world empire. In a battle fought at Ankara (1402), the Ottoman Bayezid I (1389–1403) was defeated and carried off to die in captivity. Tamerlane, whose repeated campaigns had brought great destruction to Central Asia, India, the Near East, and fatally weakened the Ulus of Joči (1405), the Golden Horde. He died in 1405, while preparing to invade China. The Ottomans recovered and under Mehmed the Conqueror (d. 1481) took Constantinople (1453), ushering in a new imperial age. The Ottoman Sultan was Khan, heir to the Oghuz Central Asian tradition, Basileus/Caesar, heir to the Roman–Byzantine Empire and Islamic champion. Mehmed's control over eastern Anatolia was secured by his victory over the Oghuz/Türkmen tribal confederation of the Akkoyunlu (*aq qoyunlu*, 1473). His grandson Selim I (1512–1520) defeated the Shī'ite Šafavids of Iran, whose military base rested on the heterodox, Oghuz soldiery of Azerbaijan and parts of Anatolia (1514) and then went on to conquer the Mamlūk state (1516–1517) and thereby gain mastery over the Arabo-Islamic heartland, including the holy cities in Arabia.

The Formation of the Modern Turkic Peoples

While the Ottomans were transforming themselves into a world power, spanning western Asia, eastern Europe and north Africa and reaching the zenith of their might under Süleyman the Magnificent (1520–1566), the Chingisid realms of Eurasia were in decline. The Ulus of Joči fragmented into the Great Horde (collapsed in 1481), the Khanates of the Crimea (1443), Kazan (by 1445), Kasimov, a puppet state of Muscovy (1452) and Astrakhan (by 1466), the Khanate of Sibir (early fifteenth century), the Noghay Horde and the Uzbek (Özbek) tribal union. By 1475, the Crimean Khans became vassals of the Ottomans. The Volga Khanates were conquered by Muscovy, Kazan in 1552, Astrakhan in 1556 and the Siberian Khanate falling by 1598. The Noghays, prominent in the throne-struggles of their Chingisid neighbours, but usually lacking central leadership, began to break up by the mid-sixteenth century, elements eventually contributing to the composition of the Bashkirs, Kazakhs, Uzbeks and Crimean Tatars. The Uzbek union, a conglomeration of Kipchak-speaking Golden Horde soldiery, came under the dynamic leadership of Abu'l-Khair Khan (d. 1468) who attempted to create a stronger central political structure. This ultimately led to the revolt of other Chingisid princes whose followers became the Kazakhs (*qazaq* 'rebel'). Weakened by this and by the western Mongol Oirats, who were now a powerful presence, the Uzbek horde seemed to be dissolving. One of Abu'l-Khair's grandsons, Muḥammad Šaybānī Khan, c.1500, brought his horde into Transoxania and conquered what is now called Uzbekistan, ending Timurid

rule there and putting into place the final ethnic components of the Uzbek people of today. One of the Timurids forced out of the region was Bābur (d. 1530), the founder of the Muḡal dynasty in India. Sunnī Šaybānīd Uzbek Transoxania, however, was blocked off from the Sunnī Muslim world of the Middle East, in particular their Ottoman allies, by Shī'ite Iran. Some have argued that this led to intellectual stagnation and the militarisation of Transoxanian society. Others suggest that shifts in international trade patterns were the cause. Whatever the merits of the different arguments, the Iranian impact was felt almost immediately. Indeed, Muḡammad Šaybānī Khan himself fell in battle with Shah Ismā'īl in 1510. This struggle also determined the modern borders of Iran and Uzbekistan. The Šaybānīd realm eventually divided into three distinct polities: Bukhara-Samarkand, Khwārezm/Khiva, which contained a strong Türkmen/Oghuz element, and later the Ferghana valley state of Khokand/Qoqan, with a sizeable Kirghiz population. Bukhara-Samarkand became a Russian protectorate in 1866 and was brought into the Soviet Union in 1920. A similar fate befell the Khivan Khanate in 1873 and Qoqan in 1876.

The nemesis of the Uzbeks, the Kazakh horde, comprising many of the same tribal elements that formed the Uzbek union, remained in possession of the Jočid–Eastern Kipchak lands. In the course of the sixteenth century, political fissures surfaced among the Kazakhs as well. By the time of Ḥaqq Nazar (1538–1580), the Great, Little and Middle Hordes had taken shape. Periodically involved in Uzbek politics, the Kazakhs also faced a dangerous foe in the Oirat union which ultimately destroyed the Kazakh supreme khanate. Under continuing Oirat pressure, the Kazakhs accepted Russian overlordship between 1731 and 1742. Occasionally allied to the Kazakhs were the Kirghiz who had undergone a number of transformations and now shared many ethnic elements with the former. Under Oirat pressure they were increasingly pushed into the regions they presently occupy, different groupings within them coming, in time, under the overlordship of the Qoqan Khanate, the Xwājas of Eastern Turkistan, the Oirats and later the Manchus. The Russian Empire absorbed them by 1876. The Turkī-speaking populations of the East Turkistanian oases, who had been living under semi-clerical rule, the Xwājas, since the sixteenth century – by which time the ethnonym *uyğur* was no longer in use – were brought into the Manchu Empire in 1757. Local Turkī intellectuals revived and adopted the name Uyğhur in 1921.

The Ottoman realm, after 1566, entered into a period of decline, punctuated by occasional revivals. The decline may be attributed to dynastic, military and economic causes. Reform and modernisation movements in the nineteenth century – which helped to shape similar trends among Turkic Muslims under Tsarist rule – proved unequal to the task of preserving the failing empire. It collapsed, along with the other great, land-based empires, in World War I. The modern, democratic, secular Turkish state that came into being in Anatolia and the remaining European regions of the Ottoman Empire,

under the leadership of Mustafa Kemal Atatürk (d. 1938), despite some difficulties, has largely pursued his programme of modernisation and now offers itself as a model to the newly independent Turkic Muslim states of the former Soviet Union, Azerbaijan, Turkmenistan, Uzbekistan, Kazakhstan, Kirghizstan.

Further Reading

- Barfield, T. J. (1989) *The Perilous Frontier. Nomadic Empires and China*, Oxford: Basil Blackwell.
- Barthold, W. (1932–4) *Zwölf Vorlesungen über die Geschichte der Türken Mittelasien*. Reprinted Hildesheim, 1962.
- (1968) *Turkestan down to the Mongol Invasion* (3rd edition), E. J. W. Gibb Memorial Series NS5, London.
- Beckwith, C. (1987) *The Tibetan Empire in Central Asia*, Princeton: Princeton University Press.
- Bosworth, C. E. (1963) *The Ghaznavids: their Empire in Afghanistan and Eastern Iran*, Edinburgh: Edinburgh University Press.
- Canfield, R. L. (1991) *Turko-Persia in Historical Perspective*, Cambridge: Cambridge University Press.
- Czeglédy, K. (1983) 'From east to west: the age of nomadic migrations in Eurasia', *Archivum Eurasiae Medii Aevi* 3: 5–125. Revised edition of *Nomád népek vándorlása napkelettől napnyugatig*, Budapest, 1969.
- Frye, R. (1975) *The Golden Age of Persia. The Arabs in the East*, London: Weidenfeld & Nicolson.
- Golden, P. B. (1980) *Khazar Studies*, 2 volumes, Bibliotheca Orientalis Hungarica 25: 1–2, Budapest: Akadémiai Kiadó.
- (1992) *An Introduction to the History of the Turkic Peoples*, Turcologica 9, Wiesbaden: Harrassowitz.
- Hambly, G. (ed.) (1969) *Central Asia*, New York: Delacorte Press.
- Inalcık, H. (1973) *The Ottoman Empire. The Classical Age 1300–1600*, London: Weidenfeld and Nicolson.
- al-Kāšyarī, Mahmūd (1982–5) *Compendium of the Turkic Dialects (Dīwān Luyāt at-Turk)*, ed. and trans. R. Dankoff in collaboration with J. Kelly, Cambridge, Mass.: Harvard University Printing Office.
- Khazanov, A. M. (1984) *Nomads and the Outside World*, Cambridge: Cambridge University Press.
- Lewis, B. (1966) *The Emergence of Modern Turkey*, Oxford: University Press.
- Liu, Mau-tsai (1958) *Die chinesischen Nachrichten zur Geschichte der Ost-Turken (T'u-küe)*, 2 volumes, Wiesbaden: Harrassowitz.
- Róna-Tas, A. (1991) *An Introduction to Turkology*, Studia Uralo-Altaica 33, Szeged.
- Spuler, B. (1965) *Die Goldene Horde. Die Mongolen in Russland 1223–1502* (2nd revised edition), Wiesbaden: Harrassowitz.

3 The Structure of Turkic

Lars Johanson

Introduction

Throughout their history and in spite of their huge area of distribution, Turkic languages share essential structural features. Many of them are common to Eurasian languages of the Altaic and Uralic types. While often dealt with in typologically oriented linguistic work, most aspects of Turkic structure still call for more unbiased and differentiated description. The following survey will give some examples of characteristic common features and of more language-specific phenomena.

Sound Systems

Vowels and Consonants

Many Turkic languages, e.g. Turkish, exhibit eight vowel phonemes, *a, i, o, u, e, i, ö, ü*, which can be classified with respect to the features front vs. back, unrounded vs. rounded, and high vs. low:

	Front		Back	
	unrounded	rounded	unrounded	rounded
high	i	ü	ɨ	u
low	e	ö	a	o

Some Turkic languages display fewer or less clear-cut distinctions than indicated in this scheme. Thus, modern Uyghur lacks a clear differentiation *i : i̇*, while Iranised Uzbek dialects show centralising tendencies which affect all the distinctions *e : a, ö : o, ü : u, i : i̇*.

On the other hand, several Turkic languages display more distinctions than shown in the scheme. Some, such as Azerbaijani, exhibit an opposition between an open *e* and a more closed *è*. Many languages have a phonemic contrast between long and short vowels. These length distinctions may be primary, as in Yakut, Turkmen and Khalaj. Long vowels may also be present in loanwords from Arabic, Persian etc., or developed secondarily through

consonant contractions. Long vowels are often diphthongs. Reduced vowels, on the other hand, are typical of the Volga region: Chuvash, Tatar, Bashkir. In these languages, due to the shifts dealt with in Chapter 5 (p. 92), the realisation of vowel distinctions deviates considerably from the scheme given above, though the basic relations in principle remain valid.

The consonant systems are more different from each other. The phonetic realisations with respect to the distinction front vs. back varies a good deal. Gagauz and Karaim, which are strongly influenced by Slavic, show palatalised front consonants. The Sayan Turkic languages Tuvan and Tofa exhibit a glottal element functioning as a fortis (strong consonant) signal, e.g. *aʔ* 'horse'. Atypical sounds include the fricatives *f, v, ž, θ*, and the affricates *ts, dz*. Long consonants normally only emerge at morpheme boundaries, but may also be found, for instance, in numerals and affective words, e.g. Uzbek *ikki* 'two', Turkish *anne* 'mother'.

Syllable Structure

A Turkic syllable typically consists of a vowel with one preceding and/or subsequent consonant, e.g. *qum-da* 'in the sand'. Vowel hiatus and initial consonant clusters are avoided. Final clusters with one nasal, liquid or sibilant occur commonly, e.g. *Türk* 'Turk', *üst* 'upper side'. When two morphs join, maximally three consonants may cluster together, e.g. *dostlar* 'friends'.

Word-initial *n, m, ŋ, l, r* are avoided, the only seemingly native exception being the interrogative *ne* 'what' (cf. p. 106). Loanwords beginning with nasals and liquids are often provided with prothetic vowels, e.g. Kazakh *orıs* 'Russian'.

The most general sound harmony phenomenon is an intrasyllabic front vs. back harmony that affects whole syllables with their vowel and possible consonants. Each syllable is classified as front or back. The frontness or backness is signalled by both vowels and consonant segments. For example, a back syllable may be realised as *qut* 'slave', whereas its front counterpart is realised as *kül* 'ashes'. There are, however, numerous exceptions to this phonetic situation. Not all consonants have clearly distinguished front and back variants. The reduced vowels of some languages do not signal frontness and backness in a clear-cut way. In loanwords, phonetically front vowels may go with back consonants and vice versa, e.g. Turkish *kâr* [kɑ:r] 'profit' (< Persian). Nevertheless, the syllable as a whole is classified as phonologically front or back.

Normally, the front consonants *k* and *g* occur in front syllables, whereas the back consonants *q, ğ* and *ɣ* occur in back syllables. In Yakut, however, *k* also occurs with high back vowels, e.g. *kīs* 'girl', whereas *χ*, pronounced as an affricate [k^x], only occurs with the low vowels *a* and *o*, e.g. *χās* 'goose'.

Morphophonological Variation in Primary Stems

As regards morphophonological variation in primary stems, the second syllable of certain stems has an unstable vowel that does not appear in front

of suffix-initial vowels, e.g. Tatar *kürək* 'beauty', *kürkë* 'its beauty', Turkish *oğul* 'son', *oğlu* 'her, his son'. In primary stems with an unstable vowel, consonant assimilation may also create cases of variation such as Tuvan *egin* 'shoulder' → *ekti* (< **ekni* < **egni*) 'his/her shoulder', *ās* (< **ayiz*) 'mouth' → *aqsi* 'his/her mouth' (< **aqzi* < **ayzi*), with syllable-final devoicing triggering progressive assimilation.

Certain Turkic languages show morphophonological variation of lenis (weak) obstruents: partial or full devoicing in syllable-final position, e.g. Turkish *ad* [ɑd] 'name', *git-* 'go', *gök* 'heaven', *kanat* 'wing', Azerbaijani *dodağ* 'lip', Uzbek *âzād* [ɑ:zɑ:dɑ] 'free', *ortâq* 'friend', Kazakh *esik* 'door', *taraq* 'comb', *qap* 'sack', *etik* 'boots', but voicing in front of a vowel, e.g. Turkish *kanadı* 'its wing', *giden* 'going', Azerbaijani *dodayım* 'my lip', Kazakh *esigi* 'its door', *tarayım* 'my comb', *qabıg* 'your sack', *etigi* 'his/her boots'.

Intervocalic lenes are often further weakened, that is fricativised or deleted, e.g. Turkish *göğü* [œy] 'its heaven', Uyghur *ayiyi* 'its foot', Tuvan *balı* (< *balıyi*) 'its fish', Khakas *azām* (< *azayım*) 'my foot'. Morphemes with final lenis thus have two allomorphs. Polysyllabic stems end in lenes, while monosyllabic stems may also end in fortes (strong consonants), e.g. Turkish *yük* 'burden', *yüki* 'its burden'. This kind of lenis vs. fortis variation is not manifested in all Turkic languages, e.g. Khalaj *topuq* 'ankle', *topuqum* 'my ankle', compare Turkish *topuk*, *topuğum*.

Primary stem variation may also emerge through contraction, e.g. Uyghur *qı-p* 'doing' ← *qıl-* + *-ip* 'doing', Tuvan *ā-p* 'taking' ← *al-* + *-ip*, or through regressive simplification of a consonant cluster, e.g. Uzbek *Tāškēngā* 'to Tashkent' ← *Tāškēnt* + *-gā*.

The reason may also be regressive devoicing, e.g. Kazakh *ḡassam* 'if I write' ← *ḡaz-* + *-sam*. This kind of assimilation is frequent in Siberia, e.g. Yakut *akka* 'to the horse' ← *at* 'horse' + *-GA*. Siberian languages also exhibit other cases of primary stem variation due to diachronic processes, e.g. Yakut *as-* (before consonants) ~ *any-* (before vowels) 'stick, pierce' (see p. 107). A specific kind of morphophonological variation in primary stems is caused by Modern Uyghur so-called umlauting, a regressive assimilation of low vowels, e.g. *baš* 'head' → *bēšim* 'my head' (see pp. 93, 383).

Morphophonological Variation in Suffixes

The intersyllabic phonotactic relations are characterised by a number of assimilations of morphs within word forms.

Intersyllabic Sound Harmony

The rules of intersyllabic sound harmony vary across languages. Suffixes are often non-harmonic at an early stage of development. Their development into harmonic suffixes may also be blocked by foreign influence, e.g. by Iranian influence in dialects of Uzbekistan, Afghanistan and Iran.

Front vs. back harmony, mostly called 'palatal harmony', is a systematic

neutralisation of the phonological distinction front vs. back in suffix syllables under the influence of a dominant immediately preceding syllable. This means that the quality of the last syllable of a stem determines the quality of a following suffix with respect to front vs. back. Primary stems allow a free choice of front and back syllables, e.g. Tatar *at* 'horse', *it* 'meat', *ēt* 'dog', *ut* 'fire', *üt* 'pass!', *öt* 'win!', *öt* 'sing!'. In the case of suffixes subject to intersyllabic front vs. back harmony, however, a front or a back variant must be selected to match the last stem syllable. Though this harmony manifests itself most clearly in the choice of vowels, it affects the whole syllable. Thus, dative suffixes are of the type *-GA* with variants such as *-ke*, *-ge*, *-qa*, *-ya*. If the harmony rules are applied consistently, back and front syllables exclude each other within word forms, e.g. Turkish *ev-ler-im-e* 'to my houses', *at-lar-ım-a* 'to my horses'.

It would be wrong to claim that the harmony is caused by the preceding vowel. It is also observed after stems with weak or vanishing 'minimal vowels', e.g. in Tatar dialects, or after the neutral *i* in Uyghur. In loanwords, the final syllable may have a back vowel, but still be front, i.e. take on front suffixes, and vice versa, e.g. Turkish *rol-ü* 'its role', *harb-i* 'its war'. Apart from non-harmonic suffixes, there also are many cases of phonetically less clear vowel realisations, notably in strongly Iranised dialects. But not even here do we find any general breakdown of the syllabic front vs. back harmony.

Some languages only display this kind of harmony, e.g. the Tatar third-person simple past forms *çıqtı* 'went out', *kettë* 'went', *ötti* 'won', *öttë* 'singed'. Others also apply a rounded vs. unrounded harmony, the so-called 'labial harmony', which is more of a real vowel assimilation. It implies neutralisation of the distinction rounded vs. unrounded in suffix syllables: the quality of the vowel of the last stem syllable determines the suffix vowel. One additional property of the preceding syllable is thus reflected in the suffix.

In many languages, this harmony only affects high suffix vowels, in Tatar and Bashkir centralised vowels (p. 92). This creates suffixes with vowels displaying a fourfold harmony such as the Turkish first-person singular possessive suffix *-(X)m* or the Bashkir third-person simple past suffix *-DX*, e.g. Turkish *kızım* 'my girl', *atım* 'my horse', *elim* 'my hand', *ipim* 'my rope', *pulum* 'my stamp', *yolum* 'my way', *gülüm* 'my rose', *gölüm* 'my lake', Bashkir *sıqtı* 'went out', *kettë* 'went', *öttö* 'won', *öttö* 'singed'.

Some languages such as Yakut and Kirghiz go further, applying labial harmony to low-vowel suffixes as well ('labial attraction'). The latter thus exhibit four variants, e.g. Yakut *-LAR* in *ayalar* 'fathers', *oyolor* 'children', *kihiler* 'persons', *börölör* 'wolves'. However, no Turkic language applies this harmony consistently. For example, the roundedness of the Yakut stem vowels *u*, *ū*, *ü*, *ū*, *o*, and *ö* is not reflected in suffixes, e.g. *ular* 'waters', and Kirghiz labial harmony does not apply after *u*, e.g. *quşqa* 'to a/the bird', *sūya* 'to water'. The scope of labial harmony must be specified for each language.

The distinction low vs. high always has semantic implications and is not subject to harmony, e.g. Turkmen *ādama* 'to a/the man' (dative), *ādami* 'the man (accusative)'.

Consonant Assimilations

Consonant assimilations create further suffix allomorphs. A very common phenomenon is progressive devoicing after voiceless consonants, $d > t$, $j > č$, $g > k$, etc., e.g. Turkish *gel-di* 'came' vs. *git-ti* 'went', Uzbek *üy-gä* 'to the house' vs. *eşik-kä* 'to the door'. The devoicing is relatively weak in some languages, and not always indicated orthographically. In the present volume, it is ignored in the notations for a few languages, e.g. Azerbaijani *-mAKdA*.

Certain languages display progressive assimilation of suffix-initial *l* to *n*, *d*, *ð*, *t*, etc., e.g. in the plural suffix *-LAR*: Kazakh *at-tar* 'horses', *köl-der* 'lakes', Bashkir *taw-ðar* 'mountains', Tuvan *χol-dar* 'hands', *nom-nar* 'books'. Some Chuvash suffixes have allomorphs beginning with *q* = written *t* (after *l*, *r*, *n*) and with *r* (otherwise), e.g. *värman-qa* 'in the forest', *tu-ra* 'on the mountain' (see Table 27.5, p. 439).

Many suffixes have allomorphs with an initial consonant after stem-final vowels, and with an initial vowel after stem-final consonants. With the type *-(V)C*, the vowel is dropped when the stem ends in a vowel, e.g. Turkish *ev-im* 'my house', *baba-m* 'my father'. With the type *-(C)V*, the consonant is dropped when the stem ends in a consonant, e.g. Turkish *başla-yan* 'beginning', *ol-an* 'being', *iki-şer* 'two each', *üç-er* 'three each'. There are diachronic reasons for this variation, and it would be wrong to claim that the segments indicated in brackets are 'connective' sounds inserted epenthetically to prevent hiatus or to break up consonant clusters.

A special kind of suffix variation is due to the so-called 'pronominal *n*' occurring in many languages between third-person possessive suffixes and case suffixes, e.g. Turkish *-(s)I(n)* in forms such as *baba-sın-a* 'to his/her father'.

Prosodic Phenomena

The main factor in word-level accent is the capacity to carry high pitch. Underived items are accentable, e.g. *at* 'horse', or unaccentable, e.g. *dA* 'and, too'. Most Turkic languages have pitch accent, that is increase of the tone height, on the last syllable of native lexical items.

Suffixes are classified into accentable ones, e.g. Turkish *-DIm* in *Uyudım* 'I slept', and non-accentable ones, e.g. *-(y)dIm* in *Uyurdım* 'I would sleep'. Pitch accent occurs on the last accentable syllable in word forms, e.g. Turkish *köylerdé* 'in the villages'.

As a rule, personal suffixes of the pronominal type, copula markers, negation suffixes (except the negative *-mAz* or *-mAs* aorist) etc. are unaccented, e.g. Uyghur *Sen kimsen?* 'Who are you?', *Yazmıdı* '(S)he did not write'. This is also true of enclitic particles such as *dA* 'and, too'.

There is also an interacting changeable dynamic stress accent, charac-

terised by more energy of articulation. It tends to fall on the first syllable, and seems to be the original factor of intersyllabic progressive sound harmony and of rhyme patterns in Old Turkic poetry. Being sensitive to phonetic factors such as weight, it often falls on heavy syllables, that is closed syllables or syllables with a long vowel, e.g. Turkish *evde* ['evdÉ] 'at home'.

Pitch and stress accent may coincide in a non-final syllable, followed by a corresponding fall in the next syllable, which yields a higher degree of prominence. Lexical items displaying this feature are mostly of foreign origin, recent borrowings or place and personal names. The accent falls on the nearest heavy penult or antepenult, e.g. Turkish *lokanta* [lɔk'ánta] 'restaurant', *pencere* [péndzere] 'window', or, if none of the syllables is heavy, on the nearest light syllable. Such patterns may, however, be replaced by word-final accent if the words in question are nativised.

Non-accentable suffixes often produce a coincidence of pitch and stress accent on the immediately preceding syllable, e.g. Turkish *Güzel-sin* [y'zél-sın] 'You are beautiful'. This phenomenon allows minimal contrasts between words such as Uzbek *yazmá* ['józ-mæ] 'do not write' and [józ-mæ] 'writing', *Atıg* [ótıg] 'Throw!' and [ótıg] 'your horse'. Conjunctions and adverbially used elements tend to be accented on the first syllable, e.g. Tatar *emma* [émma] 'but'. Turkish *yalnız* [j'ótı-nız] 'only'; compare [j'ótı-nız] 'alone' (adjective). Items of compound origin and reduplications behave similarly, e.g. Tatar *niçe* [nítʃe] 'how many', *appaq* [óp:ak] 'very white', Turkish *nasıl* [násıl] 'how', *şimdi* [şím-di] 'now'. This accent is often used for affective or emphatic effects, e.g. Turkish vocative *kardeş* [kárdeʃ] 'brother!' ← ['kardéʃ].

Both accent types are subordinated to higher pitch and stress patterns at phrase, clause and sentence levels. The components of word accent are also distributed differently in the individual Turkic languages. Central Asian languages often tend to give more prominence to the initial syllable; languages of the Volga-Kama region to the last syllable. The differences between pitch and stress accent are usually ignored in studies of Turkic accent systems. In general, reliable data on Turkic prosodic phenomena are rather scarce. Intonation patterns are particularly poorly investigated.

Morphology

Word Structure

Synthesis and Juxtaposition

The structure of the Turkic word is agglutinative, that is characterised by a highly synthetic structure with numerous bound morphemes, and a juxtaposing technique with clear-cut morpheme boundaries and predictable allomorphic variation. An example is the Turkish word *parasızlıklarından*

'because of their poverty', consisting of *para* 'money', *-sız* privative suffix '-less', *-lik* abstractness suffix '-ness', *-ları(n)* 3p.pl. possessive suffix, *-Dan* ablative suffix.

There are rich possibilities of expanding stems by numerous bound morphemes, which serve word formation and the expression of grammatical notions. For many grammatical notions, the synthetic expression is the only available method, e.g. Turkish passive *yapıl-* 'be done', causative *sevdір-* 'cause to love', genitive *krālm* 'the king's, of the king'. A high degree of combinability allows long chains of morphs. Since hundreds of forms may be derived from each single primary stem, it is difficult to present complete paradigms.

The bound morphemes mostly have a highly generalised content and thus a high applicability. For example, Turkish *-lik* '-ness' and *-ci* 'professional' are much more productive than corresponding English devices, e.g. *gazeteci* 'journalist' ← *gazete* 'newspaper', *avcı* 'hunter' ← *av* 'hunting', *sucu* 'water seller' ← *su* 'water'.

Regularity

All these factors contribute to a considerable morphological regularity. The morphemes have few and phonologically predictable allomorphs, added rather mechanically to the stem according to the rules of assimilation mentioned above. The agglutinative technique yields transparency: regular, easily segmentable structures. The content is readily matched with its segmental expression, e.g. Turkish *iş-ler-in* 'affair + plural + genitive', without any fusion of significant as, for instance, in the Latin counterpart *rerum*.

This regularity should, however, not mislead to unjustified simplifications concerning the meaning of complex forms. Certain combinations of morphs have grammatical functions not derivable from the functions of the components. Accent may also distinguish seemingly identical forms.

The few exceptions to phonological predictability include the choice of causative allomorphs and the choice between high and low vowels in the 'aorist' suffix *-(V)r*, e.g. Turkish *bil-dir-* 'let know', *anla-t-* 'let understand', *bil-ir* 'knows', *dön-er* 'turns'. Some irregularities in primary stems have been pointed out above. Nevertheless, Turkic languages basically lack declensional and conjugational classes, irregular verbs, suppletive forms, etc.

The method is consistently affixing, one morph following after the other. The primary stem is the leftmost morph (not segmentable synchronically). The affixes are thus suffixes, mostly monosyllabic ones. The primary stem can always be used as a free form, e.g. Turkish *at* 'horse' (nominal stem), *At* 'Throw!' (verbal stem). It remains intact, without infixes, additive or replacive elements. The few exceptions relate to the declension of pronouns, e.g. Turkish *ben* 'I', *ban-a* 'to me'. Neither are there usually real prefixes. Elements preceding primary stems are mostly nominal elements that can also

occur as free forms, e.g. *ön* 'front' in the Turkish loan translation *öngör-* 'foresee, provide for' ← French *prévoir*. Some languages under strong foreign influence have a few copied prefixes, e.g. Uzbek *nā-toyri* 'incorrect'.

The originals of loanwords often represent morphological processes alien to Turkic, prefixation, ablaut, metatheses, etc. One Arabic root may be represented by various forms, e.g. *hkm* in Turkish *hüküm* 'judgement', *hikmet* 'wisdom', *hākim* 'judge', *hakīm* 'wise', *mahkūm* 'sentenced', *mahkeme* 'law court', *muhākeme* 'lawsuit', *muhkem* 'solid', etc. Such processes are unproductive in the Turkic languages that incorporate the loanwords in question. Copies of Arabic plurals with internal inflection are sometimes used as singulars and provided with Turkic plural suffixes, e.g. Turkish *tüccār-lar* 'merchants' ← singular *tācir*.

Use of Suffixes

It is a typical feature of Turkic to use morphological devices economically and avoid redundancy. There are few cases of agreement. Third-person singular forms are often unmarked and the singular is used after cardinals, and certain suffixes such as number, case, possessive and copula markers may be shared by several syntactically parallel segments and only attached to the last of them, e.g. Turkish *Görmüş ve duymuş-lardı* 'They had seen and heard it'.

The order of suffixes is subject to rigid rules. Suffixes form distributional classes according to their ability to occupy relative positions within the word, that is their relative distance to the primary stem. Suffixes modifying the primary stem directly are closest to it, which means that derivational suffixes precede inflectional ones. Each added suffix tends to modify the whole preceding stem, e.g. Kirghiz *üylörömdö* ('house + plural + my + in') 'in my houses'.

Suffixes and Enclitic Particles

Suffixes must be distinguished from enclitic particles, which are free unaccentable units. They include postpositions, relators of other kinds, modal items, etc., e.g. Uzbek *burun* 'before', *edi* 'was', Turkish *ki* 'that'. Enclitics are similar to suffixes in that they are often subject to assimilatory processes such as sound harmony, e.g. Turkish *da* 'and, too'. Note that some free enclitic markers have suffixes of the unaccentable type as variants, e.g. Turkish *idi* ~ *-(y)DI* 'was', *ile* ~ *-(y)LA* 'with', Ottoman *içün* ~ *-çün* 'for'.

Stem Formation

Verbal and nominal stems are sharply distinguished, homonymous stems such as English *face* being extremely few, e.g. *āj-* 'be hungry' and *āj* 'hunger, hungry'. All stems, whether primary or secondary, can be used as free forms.

From verbal and nominal stems, expanded verbal or nominal stems are formed. Since nominal stems take on denominal suffixes, and verbal stems take on deverbal suffixes, there are four possibilities of derivation:

- 1 denominal nominal stems, e.g. Turkish *yoldaş* 'fellow traveller' ← *yol* 'way'
- 2 deverbal nominal stems, e.g. Turkish *yatak* 'bed' ← *yat-* 'lie'
- 3 denominal verbal stems, e.g. Turkish *yolla-* 'send' ← *yol* 'way'
- 4 deverbal verbal stems, e.g. Turkish *yolla-n-* 'be sent' ← *yolla-* 'send'.

The last two devices should be compared with the possibilities of analytical derivation of verbal stems. Denominal verb formation also includes lexicalised verbal phrases containing auxiliary verbs (pro-verbs) such as *et-*, *eyle-*, *qil-*, *yap-* 'do' + incorporated nominal elements which do not function as free objects, e.g. Turkish *imza et-* 'sign' (compare *imza-la-*), *affet-* 'forgive'.

Similarly, a converb and a form of a second verb may form verbal phrases with strong semantic fusion, e.g. Uzbek *ālip kēl-* 'bring' ('taking come'), *yiylāb yubār-* ('crying send') 'start crying'. For postverb constructions, see p. 42. There are also combinations of thematic stems (see p. 42) with auxiliary verbs such as *er-* 'be', *bol-* 'become'.

Word Classes

The main word classes of Turkic languages are nominals and verbals. This division is not identical with the classification into nominal and verbal stems, since verbals may also be nominal stems. Nominals comprise nouns, adjectives, pronouns, and numerals. The remaining parts of speech will be referred to as indeclinables.

Nominals

Nouns

Wherever an indefinite article is used, it is formally identical with the numeral 'one', e.g. *bir* in Turkish *iyi bir at* 'a good horse'. Many languages use this device rather infrequently, e.g. Kirghiz *žaqšī at* 'a/the good horse'. Turkic lacks definite articles, though demonstrative pronouns may sometimes seem to be used in a similar way. Grammatical gender is also absent and thus cannot constitute an agreement factor. Even morphologically marked feminine lexical items are lacking, except a few derived with copied foreign suffixes, e.g. Turkish *kraliçe* 'queen', Karaim *Karayka* 'Karaim woman'.

Nouns may contain plural, possessive and case suffixes. Their order and combinability is basically common to all Turkic languages, though in Chuvash possessive suffixes precede the plural suffix. Examples of inflectional paradigms are given in the individual chapters of the present volume.

The plural suffix is generally *-lAr* or *-LAR*, e.g. Tatar *qolaqlar* 'ears', the Chuvash counterpart being *-sem*. Unmarked forms, e.g. *qolaq* 'ear', are referred to as singular forms, though they also have generic and collective uses (see p. 51).

Possessive suffixes typically express possession and correspond in function to English possessive pronouns. They often exhibit forms such as *-(l)m*, *-(l)g*, *-(s)l(n)*, *-(l)mlz*, *-(l)ηlz*, *-LARl(n)*, signalling person and number of the possessor. The first- and second-person plural forms contain a plural element *-lz*.

There are usually five core cases expressed by accentable suffixes: a genitive in *-NIg* or *-(n)Ig*, a dative in *-GA* or *-(y)A*, an accusative in *-NI* or *-(y)I*, a locative in *-DA*, and an ablative in *-DAn*. Headless genitives are created with suffixes of the type *-KI*, e.g. Kirghiz *Bul at ayamdıqi* 'This horse is my brother's'. The nominative is suffixless, identical with the bare stem. There are many deviations from this scheme, e.g. coincidence of accusative and dative in Chuvash, of accusative and genitive in Uzbek dialects, Kumyk and Karachay-Balkar, lack of the genitive in Yakut and partitive use of *-DA* in the same language. Many case suffixes correspond in function to English prepositions. The core cases do not signal very specific relations, but have rather wide functional areas. The genitive, the accusative, and partly the dative fulfil abstract-relational functions.

Certain descriptions also reckon with a more peripheral set of cases, equative ('like'), directive ('towards'), terminative ('until'), comitative ('together with'), instrumental ('by means of'), prodesive ('for'), comparative ('than'), etc. The markers of this group are mostly unaccentable. Some of them are rather like postpositions, since they govern primary cases. Many similar suffixes, including several Old Turkic ones commonly regarded as case markers, are unproductive and are only present in adverbial relicts.

Adjectives

Adjectives are not clearly distinguished from nouns in morphological respect. However, some suffixes primarily form adjectives, e.g. Turkish *-II*, which is only attached to nouns. Adjective phrases are formed from adverbial locative phrases with suffixes of the type *-GI* or *-KI*, e.g. Uzbek *yāzdāgi* 'taking place in summer' ← *yāzdā* 'in summer', Chuvash *kunti* 'local' (with contraction) ← *kunta* 'here'. Under foreign influence, some languages may mark adjectives with special copied suffixes, e.g. Ottoman *-ī*, copied from Arabic, Turkish *-sal*, a neologistic suffix, Altay *-niy*, *-skiy*, Khakas *-nay*, *-skay*, copied from Russian in invariable forms.

The adjective in comparative constructions (see pp. 56–7) is in some languages followed by a comparative suffix, which may also simply signal a high degree of a property, e.g. Uzbek *kōprāq* 'more, very much'. Yakut has a so-called comparative case in *-TĀyAr* (p. 421).

Intensive forms may be formed from adjectives and adverbs with a preposed reduplication of the first syllable. Between the reduplication and the stem, an *m*, *p*, *r* or *s* is inserted, replacing a possible syllable-final consonant, e.g. Old Uyghur *ap ariy* 'quite clean', Uzbek *yām-yāšil* 'quite green', Kirghiz

köpkök 'entirely blue', Chuvash *çup-çura* 'jet-black', Turkish *büsbütün* 'altogether'.

Pronouns

Personal and demonstrative pronouns form a morphologically distinct nominal subclass. They often exhibit oblique stems that differ from their nominative stems, whereas most other pronouns are inflected like nouns. Gender distinctions are lacking, e.g. Bashkir *ul* 'it, she, he'.

As for personal pronouns, plural forms, with corresponding predicative forms, are often used for polite address. Reflexive pronouns can sometimes be used as more polite third-person personal pronouns, e.g. Turkish *kendisi* '(s)he [in person]'. As a rule, however, Turkic languages do not display very elaborate honorific systems.

Demonstratives mostly distinguish several deictic types for which features such as choice, distance and visibility seem to be relevant. Most Turkic languages have at least three-way systems, e.g. Bashkir *bil*, *ošo*, *şul*, without exact equivalents in the English system 'near' vs. 'further away'. There are corresponding series of demonstrative nominals and adverbs, e.g. Turkish *burası* 'this place', *buradan* 'from here'. Corresponding interrogatives are Turkish *neresi* 'what place', *nerede* 'where', Uzbek *qayêrgâ* 'where to', etc. Examples of pronominal verbs are Kazakh *büyt-* 'do this way', Yakut *çayâ-*, Tuvan *qanča-*, Ottoman *neyle-* 'do what', Kirghiz *Qantiş kerek?* 'What [is necessary] to do?'

Possessive pronouns are formally genitives of personal and demonstrative pronouns. They are often, but not always, used for emphasis, e.g. Turkmen *meniñ adim* 'my name', Uyghur *mëniñ kitiplirim* 'my books'. Especially in spoken varieties, a free possessive pronoun may also be followed by a noun without a possessive suffix, e.g. Turkish *bizim köy*, Uyghur *bizniñ yêza* 'our village', Uzbek *bizniñ kitâp* 'our book', Kirghiz *menin at(im)*, Chuvash *man(ân) laşa* 'my horse'. Independent forms are created with suffixes of the type *-ki*, e.g. Turkish *bizimki*, Uzbek *bizniki*, Kazakh *bizdiki*, Chuvash *pirënni* 'ours'. There are also corresponding interrogative, reflexive and other pronouns, e.g. Uyghur *kimiñki* 'whose', *niminiñki* 'belonging to what', *özemniki* 'belonging to myself'.

Reflexive pronouns such as *kendi*, *öz*, *bot*, Chuvash *ça* are used attributively in the sense of 'own', e.g. Uyghur *öz kitimim* 'my own book', Kirghiz *öz qolum* 'my own hand', Tuvan *bodumnuñ ažilim* 'my own work', and in the sense of '(my)self' etc., with possessive suffixes, e.g. Uyghur *men(iñ) özem*, Yakut *min beyem* 'I myself', Tuvan *bodumça* 'to myself', Chuvash *esir çävâr* 'you yourselves', Karachay *kesim* 'myself'. There are also reciprocal pronouns, mostly of the type *bir(i)biri* 'each other'.

Numerals

Modern Turkic languages normally have lexical cardinal numerals for the units one to nine, for the tens ten to ninety, for hundred, for thousand, etc. The

tens sixty to ninety are of the multiplicative types *altmış*, *yettiş*, *seksen*, *doqsan*, i.e. they contain the digits six to nine (*altı*, *yetti*, *sekiz*, *doquz*). Some languages use multiplicative juxtapositions formed with *ön* 'ten' for lower tens as well, e.g. Tuvan *üzen* 'thirty', Kirghiz dialects *çiton* 'seventy', Fu-yü *durdin* 'forty'.

Ordinals are often formed with suffixes of the type *-(1)nçil*, e.g. Kirghiz *ekinçi* 'second', Uzbek *nëçänçi*, Crimean Tatar *qaçinji* 'which in order'. Chuvash uses the suffix *-mëş*. Collective numerals are formed with the type *-ĀGU*, e.g. Old Uyghur *üçëgü* 'three together', Kirghiz *altō* 'six together'. There are also the types Uyghur *-(ü)len*, e.g. *ikkilen* 'two together', Tuvan *-(A)lĀn*, e.g. *beželēn* 'five together', Khakas *-(O)lAñ*, e.g. *altolañ* 'six together', and the special Chuvash possessive element *-ëšë*, e.g. *ikkëšë* 'two of them, a pair'. Distributives are often formed with the type *-šAr* or *-(š)Ar*, e.g. Turkish *beşer*, Chuvash *pilëkşer* 'five each', Crimean Tatar *yarimşar* 'a half each'.

Personal Markers on Nominal Predicates

The first- and second-person personal markers on nominal predicates, 'subject representatives', indicating the person and number of the first actant (see pp. 52–3), are unaccentable copula elements developed from personal pronouns, e.g. Turkish *Evdëyim* 'I am at home', Bashkir *Min yađiwsimin* 'I am a writer', Kirghiz *Men qiryizmin* 'I am a Kirghiz'. In the third person, there is mostly no personal marker as a copula, e.g. Kazakh *Dosim žaqşı adam* 'My friend is a good man', Khakas *Ol tojšiči* 'He is a worker'. In some languages, unaccentable suffixes of the type *-Dir* 'is' < *turur* 'stands' may be used in the third person.

Negative copulas are formed with particles such as Kazakh, Kirghiz, Uyghur *emes*, Uzbek *ëmäs*, Chuvash *mar*, Turkish *değil*, Azerbaijani *deyil*, Turkmen *däil*, Tatar, Bashkir *tügël* '(is) not', e.g. Uyghur *U muellim emes* 'He is not a teacher', Kazakh *Men muşalim emespin* 'I am not a teacher', Turkish *Fena değil* 'It is not bad'.

The past tense copula is mostly the past form *ëdi* 'was' of the old verb *ër-be*; e.g. Turkish *Fena değildi* (< *değil idi*) 'It was not bad'. The Chuvash counterpart is *-ččë*. There are also conditional copula particles of the type *ëse* 'if ... is' and indirective ones such as *imiş* and *ëken* (p. 45). The old role of *ër-* has largely been taken over by copula verbs such as *(b)ol-* 'become, be'.

Verbals

Verbal Morphology

The verbal morphology is complex, comprising productive markers of actionality, voice, possibility, negation, aspect, mood, tense, person, interrogation, etc., normally in the order given here. Long derived stems can thus be produced, e.g. Turkish *Kov-ala-n-ma-mış-tı-k* ('persecute + iterative +

passive + negation + post-terminal + anterior + 1p.pl.') 'We had not been persecuted'.

As regards *actionality* (German *Aktionsart*), any verb(al phrase) has a natural actional content with respect to phase structure. *Transformatives* imply an inherent crucial limit, a natural turning point, with the attainment of which a transformation takes place, e.g. *öl-* 'die', *yat-* 'lie down, lie'. They comprise two subclasses: *fini-transformatives*, in which the end of the action is the crucial limit, e.g. *öl-* 'die', and *initio-transformatives*, in which the beginning of the action is the crucial limit, e.g. *yat-* 'lie down, lie'. *Non-transformatives* do not imply any inherent crucial point, e.g. *yaz-* 'write'.

Actionality suffixes, modifying the action expressed by the verb stem, include markers of intensity, frequentativity, etc., e.g. Crimean Tatar *kes-kele-* 'cut continuously', Tuvan *biži-gile-* 'write repeatedly'. There are also desideratives and similatives such as Old Uyghur *-(V)GsA* and *-(V)msIn*, e.g. *körügse-* 'want to see', *kelimsin-* 'pretend to come'.

Simple suffixes of this type are weakly represented in modern languages. Analytical methods of derivation are more strongly developed. As already noted, a converb of a lexical verb and a second auxiliary verb may form a verbal phrase with strong semantic fusion. Such phrases have a common actancy pattern, and insertion of elements between the two verb forms is heavily restricted. The second verb – mostly with lost lexical meaning and generalised grammatical meaning – may contribute to describing the action in a more accurate manner, e.g. Uzbek *ušlap al-* ('grasping take') 'seize', Uyghur *élip bar-* ('taking go') 'lead', Chuvash *ilse pır-* ('taking go') 'bring', Turkish *olup bit-* ('becoming end') 'happen', Karachay *aytib qoy-* ('saying put') 'blurt out', or express actionality, i.e. specify a phase of the content of the preceding verb or indicate whether the action is lasting, repeated, momentary, attempted, etc., e.g. Turkish *yazıp dur-* ('writing stand') 'write permanently', *güliver-* ('laughing give') 'burst out laughing', Uzbek *yiylab yubâr-* ('crying send') 'start crying', Uyghur *oqup kör-* ('reading see') 'try to read'. Such actional auxiliaries, often erroneously called 'aspect verbs', correspond to Indo-European preverbs and may therefore also be referred to as postverbs. There is sometimes fusion of the two verbs, e.g. Uyghur *yéziwal-* < *yézip al-* ('writing take'), Tuvan *biživit-* < *bižip it-* ('writing send') 'write down'.

Various thematic stems can also combine with copulative verbs such as *er-* 'be' and *bol-* 'become', e.g. Old Turkic *-(V)r bol-* ('become doing'), signalling the transition to an intraterminal state (see pp. 43–4), Turkish *-miş ol-* ('become having done'), signalling the transition into a post-terminal state (see p. 44).

Voice is expressed by passive, reflexive-middle, causative and cooperative–reciprocal suffixes, which modify the meaning of the preceding verbal stem and affect its actancy pattern by changing the syntactic roles of actants. The most common passive suffix is *-(V)l*, e.g. *örtül-* 'be covered'. Suffixes of the type *-(V)n* often express the middle voice, e.g. Old Uyghur,

Karakhanid *alin-* 'take for oneself'. Causative suffixes include *-(V)r*, *-GUr*, *-KUr*, *-(V)t*, *-DUr*, *-(V)z* etc., e.g. *ölür-* 'kill', *yêdür-* 'give to eat'. Cooperative–reciprocal suffixes are generally of the type *-(V)š*, e.g. *körüš-* 'see each other, meet'. Note that these suffixes are used to express plurality in Kirghiz verb paradigms, e.g. *Qalıštı* 'They remained', singular *Qaldı*, *Jazışat* 'They write', singular *Jazat*.

Possibility markers are postverbal combinations of converbs with auxiliary verbs such as *bil-* ('know') and *al-* 'take', e.g. Kirghiz *Bere alat* '(S)he can give'. Most of them have developed into suffixes, e.g. Turkish *Verebilir* '(S)he can give'. The verbal negation suffix is *-MA* etc., e.g. Tuvan *Kelbediñ* 'You did not come'.

Verbal predicates, whether finite or non-finite, are marked with thematic suffixes expressing aspect, mood and tense. Certain verb forms may occur both as finite and as non-finite items. It is important to note that they do not have identical meanings in these different syntactic functions. This has often been ignored by grammarians.

Finite Forms

Finite Thematic Forms

Finite items constitute independent sentences and express various aspectual and modal perspectives relative to given temporal orientation points, notably the moment of speaking. Conjugated verb forms minimally consist of a verbal stem and a thematic suffix that signals such a perspective. Though most of them are morphologically nominal stems, e.g. Turkish *gelir* 'coming', *gelmiş* 'having come', *gelecek* 'foreseen to come', they readily fulfil the syntactic role of constituting sentences. As mentioned, however, their finite functions differ from their non-finite ones.

The number of simple and compound aspect–mood–tense forms is relatively high, and their designations are not standardised in the grammatical literature. Modern languages exhibit numerous past tenses, mostly more than one present tense, but seldom genuine future items.

Turkic languages make use of a number of aspect or viewpoint markers, which offer different ways of envisaging events with respect to their limits, that is their beginning and their end: intraterminals, post-terminals and simple terminals. Most languages exhibit rather elaborate aspecto-temporal systems.

Intraterminals, Post-terminals, Terminals With intraterminal items such as *presents* and *imperfects*, the event is envisaged within its limits, that is after its beginning and before its end. Some are more focal, putting a narrower focus on what is currently going on at the orientation point, sometimes in the sense of English progressives, e.g. Uzbek *Kêlâyâtir* '(S)he is just coming', Noghay *Barayatir* '(S)he is just going', Kazakh *Žazıp otır*, Uyghur *Yéziwattidu* '(S)he is writing', Turkish *okumaktayım*, Kirghiz *oqudamın* 'I am

reading'. Less focal items are used for events seen as ongoing within a broader period of time, for protracted, habitual or general events, e.g. Bashkir *ěšley* '(S)he works', Noghay *Baradı* '(S)he goes', Tatar *Yaza*, Uyghur *Yazidu* '(S)he writes', Uzbek *Bilámán* 'I know', Kazakh *Ol ilýiy šay išedi* '(S)he always drinks tea', Qus *ušadı* 'A/the bird flies'.

There are corresponding past items, more or less focal imperfects such as Turkmen *Iyyērdim* 'I was just eating', Azerbaijani *Alirdi* (< *alir idi*) '(S)he was taking, took', Kumyk *Bara edim* 'I was going, went', Chuvash *Širattám* 'I was writing'. Several languages also have special habitual past forms, e.g. Kazakh *Baratın* '(S)he used to go', Kirghiz *Oqūcumun* 'I used to write', Khakas *Xiyırjaqmın* 'I used to read'.

With post-terminal items such as *perfects*, the event is envisaged after its relevant limit, i.e. typically after it has been carried out. The relevant limit varies according to the actional content (see p. 42). More focal items, with a narrow focus on the orientation point, are stative or resultative, e.g. Turkish *Ölmüş bulunuyor* ('is in the state of having died') '(S)he has (just) died'. Less focal items are similar to English perfects, signalling the current relevance of a past event, e.g. Uzbek *Yāzyān* '(S)he has written', Kumyk *Baryan* '(S)he has gone', Uyghur *Bu kitapni men oquyan* 'I have once read this book'. There are also corresponding pluperfects, indicating a post-terminal aspect in the past, e.g. Azerbaijani *Yazmışdığ* (< *yazmış + idik*) 'We had written', Kumyk *Baryan edim* 'I had gone'.

Special negative items include categorical pasts such as Uzbek *Yāzyānim yoq* ('there has not been any writing of mine') 'I have not written at all', Kazakh *Körgenim žoq*, Bashkir *Kürgenēm yoq* 'I have not seen it', Turkmen *Bilemōq* 'I do not know at all', and items denoting that an event has not yet taken place, e.g. Kirghiz *Kelelek* '(S)he has not come yet' (from *-A elek*), Yakut *Bara ilik* '(S)he has not gone yet'.

All Turkic languages have simple terminal items which present the event directly and as a whole, implying the attainment of its relevant limit: a *simple past* ('preterite') of the type *-DI*, e.g. Uzbek *Yāzdim*, Chuvash *Širtám* 'I wrote'.

Modal Forms Turkic also uses verbal suffixes to convey certain modal meanings with respect to the speaker's attitude.

Imperatives exhibit different forms functioning at various levels of politeness. The thematic marker of the second-person singular is \emptyset , e.g. *Al* 'Take!'. Optatives express voluntative modality and often occur in purpose clauses. Optatives have close connections with imperatives and conditionals, sometimes occurring in similar functions. There are also necessitative or obligative items of the types *-mAK*, *-(y)AsI*, *-mAll* etc., e.g. Kirghiz *Jönömökpüz* 'We must set out', Tatar *Barasibiz* 'We must go', Azerbaijani *Gelmeliyem* 'I ought to come', and intentional items such as Uzbek *-mākči*, e.g. *Mēni körmākči* '(S)he will, intends to see me', Uyghur *Men yazmaqčimen* 'I am going to write'.

The so-called 'aorist' in *-(V)r* is mostly modal, expressing disposition, inclination, prospectivity etc., e.g. Tatar *Kiler* '(S)he may/will come', Chuvash *Širáp* 'I will write'. There are also more clear-cut prospectives or future items, e.g. Tatar *Kileček* '(S)he will come'. Most Turkic languages have special presumptive verb forms, e.g. Turkish *Uyuyordur* (intraterminal + *Dir* < *turur* 'stands') '(S)he is presumably sleeping', Turkmen *Ol oqoyānir* '(S)he is probably reading', Uyghur *Yazyandu* (post-terminal + *du* < *turur* 'stands') '(S)he has probably written'.

Indirective Forms Turkic languages also possess indirective categories, certain kinds of evidential items used to qualify the experience of the event spoken about. Indirective statements concern the conclusion regarding an event and thus do not present the event itself in a direct way. The source of information may be hearsay, inference from results, or direct experience ('as is obvious', 'as it turns out', etc.). The expression of this epistemic modification varies across languages. Post-terminals such as *-mlš* and *-(V)ptlr* tend to get indirective interpretations as 'subjective pasts', e.g. Turkish *Ali gelmiş* 'Ali has [reportedly, apparently, obviously] come', Uzbek *Yāziptılar* 'They appear to have written', Uyghur *Yeziptu* '(S)he appears to have written', Kazakh *Baripti*, Altay *Bariptur*, Yakut *Barbit* '(S)he appears to have gone'.

Besides these deverbal past tense devices, there are also tense-indifferent indirective copula particles of the types *imiš*, *ēken*, which combine with nominal stems, e.g. Turkish *imiš/-(y)mlš*, e.g. *Ali geliyormuş* 'Ali is/was [reportedly, apparently, obviously] coming', Uyghur *imiš*, *ēken*, Turkmen *-mlš*, *ēken*, Uzbek *ēmiş*, *ēkän*. The particle *ēken* tends to convey the meaning 'as is/was obvious' or 'as it turns/turned out'.

Personal Markers

The most periphrastic inflectional items are personal markers, 'subject representatives', indicating the person and number of the first actant (see pp. 52–3). The dominant type of first- and second-person markers is the one used after nominal predicates (see p. 41): unaccentable markers of pronominal origin, e.g. Turkish *Geliyor-sun* 'You come, are coming', Kirghiz *Kele-biz* 'We come'. The third person is unmarked or sometimes marked by an unaccentable suffix of the type *-Dir* < *turur* 'stands'. Certain thematic stems, notably the simple past, take on accentable suffixes of the possessive type, e.g. Uzbek *Kēldi-m* 'I came'. Note that some authors in the present volume do not analyse the simple past as *DI* + personal markers *-(I)m*, etc., but rather as *D* + personal markers *-Im*, etc.

The enclitic copula particles, developed from forms of the obsolete verb *er-* 'be', have already been mentioned. Past markers of the type *edi* 'was' locate the thematic perspectives temporally, e.g. Turkish *Gelmişti* '(S)he had come'. The roles of indirective copula particles such as *ēmiş* and *ēken* have just been mentioned.

Non-finite Forms

Non-finite Thematic Forms

Turkic is rich in non-finite predicative forms based on action nouns, participles and converbs. The corresponding suffixes function as thematic markers and serve to non-finitise verbal stems. Some of them take on personal markers.

Action Nouns Action nouns refer to actions and are used to construct complement clauses, e.g. Turkish *-DİK*, *-(y)AcAK*, *-(y)İy*, *-mA*, Uzbek *-Gân*, *-Gânlik*, *-(â)yâtgân*, *-(â)yâtgânligi*, *-âdigân/-ydigân*, *-(i)š*, Turkmen *-A(:)n*, *-jek*, Chuvash *-ni*. They are, as a rule, used with possessive suffixes as 'subject representatives', e.g. Turkish *aldığım* ('my taking') 'that I take/took'. After nominal stems, corresponding copulative markers are used, e.g. forms of copula verbs such as *(b)ol-* 'become, be', copula particles of the type *eken*, e.g. Uzbek *ékân*, *ékânlik*, or other older derivatives of *er-* 'be', e.g. Ottoman *idük*, Turkmen *-DİK*. These markers also carry possessive suffixes, e.g. Turkish *hasta olduğum* ('my being ill') 'that I am/was ill'.

Participles Participles refer to entities participating in actions, and can be used as attributes or without a head. They are often identical in form with action nouns. Thus, the Turkish form *duyduğum* 'my hearing' means, in isolation, both 'the fact that I hear(d)' and 'what I hear(d)'.

Many languages have special intraterminal ('present') participles, presenting the event as current, e.g. Uzbek *-(â)yâtgân*, Kazakh *-A žatqan*, Chuvash *-(A)kAn*, Khakas *-pčAtKAn*, Tuvan *-(V)p turar*, Turkmen *-yĀn*, e.g. Kirghiz *oqup jatqan student* 'a/the student who is/was reading'. Some of these intraterminal participles are less focal, e.g. Uzbek *-âdigân/-ydigân*. There are also participles with post-terminal or terminal meaning (perfect participles), e.g. Azerbaijani *-mİš*, *-DİK* *-(y)An*, Turkmen *-A(:)n*, Tatar *-GAn*, Uzbek *-Gân*, Chuvash *-nĀ*.

There are also prospective (future) participles such as Bashkir *-(y)AsAK*, Kumyk *-(A)žAK*, Tuvan *-(V)r*, Chuvash *-(A)s*, e.g. Kumyk *oXužaq* 'who will read', and necessative participles such as Tatar *-AsĖ/-ysĖ*, Chuvash *-mAllA*, e.g. Tatar *yazasĭ*, Chuvash *širmalla* '(necessary) to write'. One type of participle denotes events that have not yet taken place (participium nondum facti, i.e. participle of the not yet done), e.g. Tuvan *kelgelek* 'not having come yet', Altay *kirgelek* 'not having entered yet', Kirghiz *körö elek* 'not having seen yet' (compare pp. 413, 427).

Active participles, e.g. *-GAn*, may often, notably in older languages, refer to entities different from the first actant of the verbs. This may yield 'impersonal' functions and seemingly 'passive' readings without passive suffixes, which is the basis for 'impersonal' interpretations of relative clauses (see p. 62), e.g. Khakas *say-an inek* ('milk-PART COW') 'a/the cow [that somebody has] milked'.

In such cases, possessive markers may function in a way reminiscent of ergative markers: *kör-gen-im* ('see-PART-my') 'what I have seen'.

Converbs Converbs are adverbial forms of the verb signalling various semantic relations to the content of the superordinate clause. A few converb markers are simple, morphologically unanalysable, e.g. *-(y)V*, *-(V)B*, *-GAll*, *-GAč*, *-sA*. More elaborate forms are based on verbal nouns and mostly marked with adverbial cases such as locative, dative, ablative, instrumental, equative or with postpositions. Some converbs lack a clear one-to-one relationship of affirmative and negative forms.

Intraterminal converbs are formed with *-(y)V*, *-(y)ArAK*, *-(y)Vr* + locative, etc., e.g. Bashkir *ala* 'taking', Tuvan *kele* 'coming', Turkish *giderek* '(by) going'. Converbs in *-(y)V* tend to occur in pairs, e.g. Yakut *oXto oXto* 'running'. There are also post-terminal converbs in *-GAč*, *-mİš* + locative etc., e.g. Bashkir *alyaθ*, Tuvan *alyaš* 'having taken'. All Turkic languages have a terminal converb marker of the type *-(V)B*, Chuvash *-sA*, Yakut *-(A)n*, e.g. Turkish *-(y)İp* in *gidip* 'going (and ...)' (see p. 64). Conditionals may sometimes occur in finite functions.

Personal Suffixes

Action nouns and participles generally take on personal suffixes of the possessive type, e.g. Turkish *gel-diğ-im* ('my coming') 'that I come/came'. Most simple converbs do not carry personal suffixes, but in Yakut they generally conjugate for person and number, e.g. *bar-am-min* 'me going/having gone'. More elaborate converb endings often contain personal markers, e.g. Old Turkic *olor-duq-°m-a* 'as I sat down'.

The simple conditional mostly takes on accentable personal suffixes of the possessive type, e.g. Turkish *gel-se-m* 'if I should come' (hypothetical), which also combine with the past copula particle, e.g. Turkish *gel-se-ydim* 'if I came' (counterfactual). There are also combinations of nominal verb forms with conditional copula markers of the type *èse* 'if ... is', e.g. Turkish *gelirsem* 'If I come'.

Indeclinable Word Classes

The indeclinable word classes include adverbs, postpositions, copula particles, interjections and conjunctions or similar relators signalling connections between the parts of a sentence.

Adverbs

Adverbs do not constitute morphologically well-defined categories in modern Turkic languages, e.g. Gagauz *bün* 'today', Uyghur *bek* 'very', *emdi* 'now', *burun* 'formerly', *tünügün* 'yesterday', Bashkir *biyil* 'this year', Kirghiz *beri* 'hither'. Many of them are fossilised case forms such as old directives or instrumentals, old participles and converbs, or forms of unknown origin, e.g. Uzbek *sonrà* 'afterwards', Bashkir *yeyën* 'in summer', Old Turkic *edgüti*

'well', *birle* 'together'. Adverbs may also be formed with productive equative suffixes such as *-JA* or *-DAY*, e.g. Turkish *gizlice* 'secretly', Uzbek *bundây* 'this way'. Case forms of pronouns are often used as conjunctive adverbs, 'adjunctors' (see pp. 48–9), e.g. Kazakh *sondıqtan* 'thus'. Certain Turkic languages make frequent use of converb forms as adverbs, e.g. Uyghur *harmay* 'unremittingly', *yağşılap* 'in a friendly way'.

Postpositions

Turkic has rich systems of adpositions, grammatical relators which differentiate the relational concepts expressed by the cases. They are free word forms and, according to the left-branching syntax, postpositions. Some are homonymous with adverbs, e.g. *birle* 'with' or *sogra* 'after'. Some go back to converbs, e.g. Uzbek *körâ* 'according to', *âlip* 'from', Uyghur *qarap* 'to'. Many of them govern cases, e.g. the ablative, as in Uyghur *bizdin burun* 'before us', or the dative, as in Gagauz *bana deyni* 'for me'. Some govern the genitive of personal and demonstrative pronouns, e.g. Kirghiz *menin menen*, Uzbek *mên(ig) blân* 'with me', Uyghur *sen/sênig üçün* 'for you'.

One kind of postposition clearly goes back to nouns, notably space nouns (nomina loci), provided with possessive and case suffixes, e.g. Uzbek *âldimdâ* 'before me', *uniğ yânigâ* 'to him/her/it'. The possessive suffix may refer to a preceding noun phrase in the genitive or the nominative, e.g. Turkish *ev(in) önünde* 'in front of the house', *ev(in) arkasından* 'from behind the house', Bashkir *öhtel öhtönde* 'on the table', Uyghur *şehar yénida* 'near the town', Chuvash *têp şine* 'on the ground'.

Conjunctions

Turkic languages have few conjunctions. Even coordinative conjunctions meaning 'and', 'or', 'but', 'for' etc. are often copied from Persian, Arabic, Russian, etc., e.g. Uyghur *peqat* 'but', Shor *no* 'but'. Old Turkic displays a few free conjunctors such as *azu* 'or'; compare Tuvan *azî*. Many languages use enclitic particles such as Uzbek *Dâ* in *Kêldi-dâ, ketti* '(S)he came and left'. Elements denoting 'with', instrumental cases or postpositions such as *birle* and *ile*, may be used instead of coordinative devices, e.g. Uyghur *ata bilen ana* 'father and mother', Turkish *karga ile tilki* 'the crow and the fox', Chuvash *naukapa praktika* 'science and practice'. With increasing Europeanisation, pure juxtaposition without syndetic elements as 'and', 'or' etc. becomes less frequent.

Turkic clauses are embedded by means of suffixed subordinative elements, non-finite predicative markers, referred to here as subjunctors (pp. 59–60). The use of free subordinative items, conjunctions and relative pronouns is thus untypical. Where such items do occur, they are homonymous with interrogatives, e.g. *qaçan* 'when', *kim* 'who', or copied from other languages. Fossilised converbs of verbs meaning 'say', e.g. *dep, deyi, diye, tese, tēyen, dēgân, dēgen*, serve as unquoting particles placed after quoted direct speech or content of thoughts, e.g. Uzbek *kêlâsan dēb* ('saying: you will come') 'with

the idea that you will/would come', *kêtsin dēb* ('saying: may he go!') 'in order for him to go'.

Many items referred to as conjunctions in the literature are in reality conjunctive adverbs, referred to here as adjunctors, e.g. Old Uyghur, Karakhanid *anin* 'therefore', Kumyk *šo sayalı* 'for that reason', Kazakh *sondıqtan* 'thus', Chuvash *ançax* 'however', Uyghur *şuglaşqa* 'therefore, hence'.

Syntax

Nominal Phrases

Since Turkic syntax is basically head-final or left-branching, i.e. dependents precede their heads according to the so-called 'rectum-regens rule', the head of a nominal phrase succeeds modifiers such as adjectival, genitival and participial attributes. A common order pattern is demonstrative pronoun + cardinal number + adjectival attribute + head, e.g. Turkish *bu üç mavi kuş* 'these three blue birds'. In languages that use the numeral *bir* as an indefinite article, the latter tends to stand next to the head, e.g. Turkish *büyük bir ev*, Uzbek *kattâ bir üy* 'a big house'. There is no agreement in number or case between dependents and heads.

One type of nominal phrase is the combination adjectival attribute + noun, e.g. Turkish *büyük evler* 'big houses'. A second type, in which the attribute has predicative force, i.e. can take complements, is the combination participial attribute + noun, e.g. Turkish *bekleyen adam* 'the man who is/was waiting', *kaldığım ev* 'the house I stay/stayed in' (see relative clauses pp. 61–3).

The attributive use of adverbials is limited, e.g. Crimean Tatar *açlıqtan ölüm* 'death by starvation', Turkish *taştan duvar* 'wall of stone', *Londra'da (oturan) bir Türk* 'a Turk (living) in London', *halka (yapılan) hizmet* 'service (rendered) to the people'. Some adverbial expressions can be used attributively with the addition of the *-ki*, e.g. Turkish *sokaktaki adam* 'the man in the street', *bugünkü görüşme* 'today's meeting'.

Genitive Constructions

In the combination genitival attribute + noun, the first element, expressing the possessor, carries a genitive suffix, whereas the head, indicating the entity possessed, is provided with a possessive suffix. Examples with third-person possessive suffixes are Kirghiz *atın başı*, Chuvash *laşan puşê* 'a/the horse's head'. Uzbek *âdâmnig üyi* ('man's house-his') 'the man's house', Turkish *öğretmenin kitabı* 'the teacher's book', *evin kapısı* 'the door of the house', Tatar *kitapnıñ tışı* 'the cover of the book', Bashkir *qalanıñ baqsahı* 'the park of the town'. In Yakut, which lacks a genitive suffix, the attribute is in the nominative, e.g. *kihî jiete* 'the man's house'. If the head is an action noun, e.g. Turkish *Ali'nin beklemesi* 'Ali's waiting', Kirghiz *kündüñ çiyişi* 'the rise of

the sun', it has predicative force, i.e. can take complements. Genitive constructions with headless adjectives may be used to express superlativity, e.g. Uyghur *atniŋ yaǵsisi* 'the best of (the) horses'.

Compounds

The dominant type of nominal compound adheres to the possessive pattern noun + noun + third-person possessive suffix, e.g. Turkish *el çanta-sı* 'handbag', *ev kapı-sı* 'front-door', Turkmen *yöl hereket-i*, Kirghiz *jol qiyimil-i* 'road traffic', Uyghur *uyghur helq-i* 'the Uyghur people', Kirghiz *qiryiz til-i* 'the Kirghiz language'. A possessive suffix expressing personal possession replaces the third-person suffix, e.g. Turkish *el çanta-m* 'my handbag'. Unlike in the genitive construction, no element can be inserted between the nouns. Numerous cases of absence of the possessive suffix, notably in nominative forms of the compound, are observed in older and modern languages, e.g. Karachay-Balkar *alma terek* 'apple-tree' (compare cases such as *alma teregin-den* 'from an/the apple-tree'). There is also a similar neologistic Turkish compound type, e.g. *budunbilim* ('people science') 'ethnology'.

Identity attribution means that two nouns referring to the same entity are juxtaposed asyndetically as qualifying attribute + head, e.g. Turkish *kadın öğretmen* 'woman teacher', *haydut polis* 'brigand policeman', *dostum Ali* 'my friend Ali'. The attribute often refers to materials, e.g. *taş köprü* 'stone bridge', Chuvash *çul şurt* 'stone house', Turkmen *altın bayat* 'golden watch', Turkish *altın elma* 'golden apple'. The attribute may also be a participle with predicative force, that is the basis of a relative clause, e.g. Turkish *konuşan adam* 'the man who speaks/spoke', *sevdiğim kadın* 'the woman I love(d)'.

The asyndetic type noun + noun is also used in coordinative compounds, so-called twin words or binomes. In this case, two parallel nouns with similar meanings form a synonym compound, hendiadys, e.g. Old Turkic *iş küç* ('work, strength') 'efforts', or a hyponym compound to express a higher concept, e.g. Uzbek *bariş-keliş* ('going, coming') 'mutual relations', Karachay *alış-beriş* ('taking, giving') 'trade', Uzbek *âta-âna* ('father, mother'), Yakut *iye-aya* ('mother, father') 'parents', Tuvan *imirä-sëk* ('mosquito, fly') 'flying insects'. This type often includes alliteration and rhyme formations, e.g. Kazakh *kiyim-keşek* 'clothing', Turkish *karıkoca* ('wife, husband') 'married couple', *dedikodu* ('said, laid') 'gossip'. Sometimes only one of the elements has a lexical meaning, e.g. Karakhanid Turkic *yâş yôş* 'vegetables, greens', Turkish *çoluk çocuk* 'family, wife and children'.

Turkic languages often display echolaly constructions expressing 'et cetera', 'and the like': a given word is followed by a reduplicating 'echo word' with an initial labial consonant, *m-*, *b-* or *p-*, which replaces a possible original consonant, e.g. Uzbek *nân pân* 'bread and similar baker's ware'.

Adjective Phrases

Compound adjectives such as Karachay *qaraqaş* 'with black eyebrows' are rather frequent. Due to the head-final structure, adverbials precede their head

in the adjective phrase, e.g. Uzbek *jüdâ issiq* 'very hot', Karaim *astri kati* 'very hard'. Adjectives may also be used as heads of nominal phrases, e.g. Uzbek *yâşlâr* 'young ones'. In older languages and some modern Siberian languages, adjectives may serve as abstracts denoting qualities, e.g. Yakut *bây* 'rich(ness)', without a suffix of the type *-IJK*, which forms abstract nouns in most Turkic languages, e.g. Karachay *bayliq* 'wealth'.

Use of Possessive and Plural Forms

In addition to its normal possessive function, a third-person singular possessive suffix may refer to a known entity or, anaphorically, to something preceding it in the discourse, e.g. Kirghiz *bâri* 'all (of it, them)', Turkish *başkası* 'the other (one)'. This limited function does not make it a definite article in the proper sense. There are often duplications of the possessive suffix such as in Turkish *bir-i-si* 'one of them', Uzbek *köp-i-si* 'most of it'. Third-person singular possessive suffixes can also have adverbialising functions, e.g. Uzbek *kêçâsi* 'at night'.

Plural suffixes mostly signal individual plurality, e.g. Turkish *elmalar* '[single] apples', Chuvash *pürtsem* '[a number of individual] houses'. The singular has a broad, partly number-indifferent range of use, which also includes collective or generic reference, e.g. Turkish *Elma aldım* 'I bought apple(s)'. Plural suffixes may also be used in an honorific sense, to express respect. Plurality expressed by cardinal numerals and other quantifiers mostly excludes agreement in the form of plural marking, e.g. Turkish *iki at* ('two horse') 'two horses'. Numerators (counting words) are sometimes placed after the numeral, e.g. Turkmen *θâni* (literally 'count'), Uzbek *dâna*, Kirghiz *dâna* 'piece', Uzbek *bâş* 'head' (for animals), Tatar *töp* 'root' (for plants). Uzbek has numerators of the type *bittâ*, *iktâ*, *uštâ*, where *-tâ* means 'piece(s)'.

The Syntax of Numerals

The syntax of numerals is rather regular. Complex numerals are formed by combining the primary cardinals for digits, tens, hundred, thousand, etc. Hundreds, thousands etc. are expressed multiplicatively, e.g. Turkish *iki yüz* 'two hundred', Tatar *biş yöz* 'five hundred', *ikë yöz mëŋ* 'two hundred thousand'. Intermediate numbers are expressed additively, e.g. Turkish *on iki* ('ten two') 'twelve'. Higher numerals precede lower ones, the highest decimal place being the leftmost one, e.g. Tatar *unbër* 'eleven', *mëŋ ëlli ikë* '1052', Kazakh *jïyirma jeti miŋ beş jüz toqson bir* '27,591', Turkish *bin dokuz yüz doksan dokuz* ('thousand nine hundred ninety nine') '1999'. Approximative numbers are often expressed by means of juxtaposition, e.g. Turkish *altı yedi* 'six or seven'.

While most Turkic languages count in tens and units, Karachay-Balkar displays, as in earlier Kumyk, a vigesimal system based on the number twenty as a counting unit, e.g. Karachay *jïyirma bile on* ('20 + 10') 'thirty', *eki jïyirma bile on* ('2 × 20 + 10') 'fifty', *beş jïyirma* ('5 × 20') 'one hundred',

on *jüyirma* ('10 × 20') 'two hundred'; compare French *quatre-vingt* ('4 × 20') 'eighty'. Khalaj has interesting variants such as *akki ottuz* ('2 × 30') 'sixty', *akki qirq* ('2 × 40') 'eighty', *üç hottuz u yirmi* ('3 × 30 + 20') 'one hundred and ten', etc.

Yellow Uyghur still partly preserves an older subsystem of higher rank counting, combining the digit numeral with the numeral of the higher ten, e.g. *per yigirma* ('one twenty') 'eleven', *per otuz* ('one thirty') 'twenty-one' (see pp. 74, 144).

Predications

With respect to relational typology, Turkic languages adhere to a so-called nominative–accusative pattern. There is a first actant which may manifest itself as the subject of a predication. The nominative marks the first actant of intransitive verbs, independently of semantic roles, agents and non-agents being coded in the same way. With transitive verbs, the nominative marks the first actant = mostly agent, and the accusative marks the second actant = mostly patient. There are no clear tendencies to code, as in so-called 'active languages', agentive complements of intransitives differently from non-agentive ones, or to mark, as in 'ergative languages', the agent of transitives with a special case. Note that some post-terminal participles and verbal adjectives may refer to non-agentive first actants of intransitive verbs, whereas they refer to non-agentive second actants of transitive verbs, e.g. Turkish *yanık* '(having) burned' ← intransitive *yan-* 'burn', *kesik* '(having been) cut' ← transitive *kes-* 'cut'.

A Turkic predication minimally consists of a predicate. An overt subject is optional, and personal markers are often missing in the third person, e.g. Turkish *Yer* '[He, she, it] eats'. A verbal predicate consists of a predicate core, provided with a thematic marker and mostly with a personal marker, e.g. Turkish [*Sen*] *gel-iyor-sun* 'You come'. A nominal predicate contains a nominal or adverbial item as the predicate core, e.g. Turkish *Öğretmen* '[(S)he is a] teacher', *Burada* '[It is] here'.

Actancy Patterns

The predicate can be expanded according to specific actancy patterns, which determine the overt syntactic relations between predicate core and complements. Most of these relations are signalled by case suffixes and postpositions.

First Actants

A central first actant, typically (but not always) agent, that is source of the action, is necessary for subject realisation. It may be realised overtly as a subject and/or as a personal marker as 'subject representative', e.g. *Ali asker* 'Ali is a soldier', *Uyudu-m* 'I have slept'. Personal markers are used whether an overt subject is present or not. Some thematic stems, however, do not take on personal suffixes, e.g. Turkmen *Men geljek* 'I will come', Uyghur *Men yazyan* 'I have written'. In Salar and Yellow Uyghur, finite verb forms

normally dispense with personal markers altogether, Salar *Men kiler* 'I will come', Yellow Uyghur *Sen parar* 'You will go'. Number agreement between inanimate subject referents and the predicate core is virtually never marked, e.g. Uzbek *Ėşiklar yâpiq* 'The doors are closed'. With animate subject referents, there are varying language-specific rules, but there is a tendency to avoid two plural markers very close to each other, e.g. Turkish *Öğrenciler geldi(*ler)* 'Students/the students have come'.

As was noted in connection with the participles, lack of a first actant referent yields 'impersonal' predications. This is often possible without passive-marking, especially at older language stages. There are remnants of diathetically less elaborated systems, where actant relations are less explicitly marked and where subject omission may suggest that no specific first actant is meant, e.g. Karakhanid *Alin arslan tutar* '[One] can catch a lion by guile'. This diathetical indifference is still common in certain types of relative clauses (see p. 62). Even some finite items in modern languages may be used this way, e.g. the necessitative *-mAll* in Turkish sentences such as *Ne yapmalı?* 'What should one do?'

Second Actants

Second actants typically represent affected or effected goals of the action. In simple active verbal clauses, second actants of transitives are realised as direct objects, typically as 'patients', in the form of accusative or nominative complements according to certain rules of topicality/specificity, e.g. Uzbek *kitâpni oqi-*, Kirghiz *kitepti oqu-*, Chuvash *kënekene vula-* 'read the book'; compare the corresponding syntagms Uzbek *kitâp oqi-*, Kirghiz *kitep oqu-*, Chuvash *këneke vula-* 'read (one or more) books'.

Second actants of intransitives are realised as complements by means of adverbial cases or postpositions, e.g. the dative as with Kirghiz *jaq-*, Karakalpak *una-*, Noghay *yara-* 'please', Noghay *yoliq-* 'meet', Karakalpak *isen-*, Noghay *inan-* 'believe', Kirghiz *oqšo-*, Karakalpak *usa-*, Yakut *mārinnā* 'resemble', the ablative as with Kirghiz *tayman-*, Azerbaijani *gorç-*, Karakalpak *qorq-*, Yakut *kuttan-* 'fear', the locative as with Turkish *israr et-* 'insist on', or postpositions as with Turkish *ile yetin-* 'be contented with'.

Third Actants

A third actant may occur as a complement in the dative, as 'indirect object', typically the 'recipient', the entity that receives a 'patient', e.g. Turkish *bir şeyi bir kimseye ver-* 'give something to somebody', Uzbek *Mëggä kitâpni bërin* 'Give me the book!', Kirghiz *Asan maya kitepti berdi* 'Hasan gave me the book'. It can also be in other cases, e.g. in the ablative such as in Turkish *bir kimseyi birşeyden kurtar-* 'rescue somebody from something'.

Other Relations

Free adverbial constituents, not required by any actancy pattern, function as qualifiers at different levels and express location, direction, origin, means,

beneficent, manner, time, companionship, reason, etc. In addition to their abstract-relational uses, Turkic datives have local, directive, allative, and terminative functions. In some Siberian languages, the dative is also used in functions otherwise typical of the locative, e.g. Tuvan *Men Qizilya čurttap turyan men* 'I have lived in Kyzyl'. The locatives have very wide meanings of place in time or space, 'in, on, at', etc. The ablatives denote, in a concrete or figurative sense, source, origin, starting-point ('from', 'out of'), way, channel ('along', 'through'), reason, means, standard of comparison, etc. Equatives express similarity, but also have mensurative and prosecutive functions, expressing extension in time and space. As noted above, they sometimes serve as general adverb suffixes, e.g. Crimean Tatar *sayıja* 'numerically', Turkish *güzelce* 'beautifully', *Fransızca* 'French' (< 'in the French way').

Other relations are specified by various means, notably postpositions, i.e. relators that form a constituent with a given dependent and connect it with the predicate core, e.g. Turkish *bir kimse ile konuş-* 'talk with somebody'.

Manner adverbials may be expressed by equative suffixes, converbs or other means, e.g. Turkish *gizlice* 'secretly', *gizli olarak* ('being secret') 'secretly', *gizli bir şekilde* 'in a secret way'. Many time adverbials are unmarked, e.g. Turkish *bir gün* 'one day', *bir saat* 'for an hour'.

A secondary predicative relationship in the sense of 'being', 'as', 'in the function of' may be established between a nominal constituent and a second copredicative nominal. Subjective predicatives ('subject adjuncts') are often provided with essive markers meaning 'being', e.g. Turkish *Öğretmen olarak çalışıyor*, Kirghiz *Muylım bolup işteyt* '(S)he works as a teacher'. With certain transitive verbs, a corresponding secondary predicative relationship may obtain between the direct object and an objective predicative ('object adjunct'), even without an essive marker, e.g. Turkish *Ali'yi başkan seçtiler* 'They elected Ali (to be) president'.

Diathetic Patterns

Diathetic relations are encoded by means of passive, causative, reflexive, medial and cooperative-reciprocal voice markers, which systematically modify the basic actancy patterns and the roles of participants.

Cooperative-reciprocals and Reflexives

Cooperative-reciprocal markers such as *-(V)ş* express cooperation or competition of participants.

So-called reflexive markers such as *-(V)n* indicate that the action does not transcend the domain of the first actant referent, but remains immanent, not related to any external entity ('immanence'). The first actant may be the goal of the action ('reflexive'), or the source of an action without a specified goal ('deobjective', 'anti-transitive', etc.). The first actant can also be the beneficiary of the action ('middle voice'), in which case the verb may govern direct objects: 'do something for oneself'. For example, the Tuvan reflexive form *bižittin-* ← *biži-* 'write' can be interpreted as 'write' (deobjective), 'write

for oneself', and 'be written'. Unambiguously reflexive meanings are mostly expressed by reflexive pronouns, e.g. Turkish *kendini öldür-* 'kill oneself'. Furthermore, the suffix *-(V)n* is ambiguous in many Turkic languages, since it may also serve as an allomorph of passive markers, e.g. Turkish *taran-* 'be combed, comb oneself'. Old Turkic *-l'n* and *-(°)d* seem to signal the middle voice, whereas *-(°)K* derives intransitives from transitives in a more general way. Many reflexive verbs have special lexical meanings, e.g. Turkish *sev-* 'love', *sevin-* 'rejoice'.

Passives

Common to passive and causative patterns is that they signal 'transcendence' in the sense that the range of the action transcends the domain of the first actant, which is either the goal or the source of the action (see p. 56). In the first case, with passives, the transcendence is exogenic (originating from outside). In the second case, with causatives, it is endogenic (originating from within).

Passive patterns typically have one actant less than the corresponding active – 'initial' or 'non-diathetic' – pattern. The first actant of the corresponding active pattern, the 'initial subject', is demoted. It is not necessarily expressed, not even with passives derived from intransitives, e.g. Turkish *Burada güzel yaşıyor* 'One lives well here' (cf. p. 228). With passives derived from transitives, a non-first actant of the corresponding active pattern is promoted to first-actantship, typically as 'patient', e.g. Turkish *Ali resim çekti* 'Ali took a picture/pictures' → *Resim çekildi* 'A picture/pictures was/were taken'.

Passivisation is thus often used as a device not only for backgrounding but also for concealing the agent of an action. This function is similar to that of 'impersonal' active constructions not referring to specific first actants (see p. 53). An agent may be optionally indicated by adjuncts based on postpositions such as Old Turkic *üze* or items copied from other languages such as Turkish *tarafından* and Uzbek *tamānidān*. Some languages, e.g. Chuvash, make rather restrictive use of passives.

In languages such as Chaghatay, Uzbek and Uyghur, passive verbs may also occur with accusative-marked direct objects, e.g. Chaghatay *Ol ramazānni Xoǰandta ötkerildi* 'That Ramadan was spent in Khojand', Uzbek *Čāyni içildi* 'The tea was drunk', Uyghur *Aşni yeyilgen* 'The food was eaten'. Here, first actant suspension by means of the passive is combined with direct object topicalisation by means of the accusative (see p. 58).

Passive markers may also be used for pure intransitivising, e.g. Turkish *açıl-* 'open (intransitive)', or to express reflexive meaning, e.g. Turkish *katıl-* 'attach oneself, join'.

Causatives

Causative patterns typically provide one actant more than the corresponding initial non-diathetic patterns. Causative markers signal that the first actant is

the source of the action as its causer (initiator, permitter, etc.).

The first actant of a corresponding non-diathetic pattern ('initial subject') is demoted and expresses the causee in the function of a direct or indirect object. Causative suffixes added to intransitives yield transitives, e.g. *İbrahim, Ali'yi öldürdü* 'İbrahim killed Ali', 'İbrahim caused Ali to die'. Added to transitives, they yield causative patterns in which the direct object of the corresponding non-causative active pattern may occur, e.g. Turkish *İbrahim resmi çekirdi* 'İbrahim had the picture taken'. A causee agent, corresponding to the first actant of the non-diathetic pattern, can be optionally expressed, like the agent in passive patterns, though in this case mostly a dative complement is used, e.g. Turkish *İbrahim resmi Ali'ye çekirdi* 'İbrahim had the picture taken by Ali'.

In some languages, causative suffixes added to transitives may also imply that the first actant is the patient, the entity affected by the action, e.g. Tuvan *ölürt-* 'let kill' or 'be killed' ← *ölür-* 'kill'. This is the meaning of 'transcendence' mentioned above. The agent, the entity performing the action and corresponding to the first actant of the non-diathetic pattern, is in the dative, e.g. Tuvan *Xoy börüye çidirtken* 'A/the sheep was eaten by the wolf'. The first actant referent is animate, and there is an implication of disadvantage for it, though the responsibility for this is not necessarily attributed to it. Such 'reversive' forms and patterns are widespread in Eurasian languages, and present in Old Turkic already. Without being identical to any of them, they show affinity to permissive causative constructions ('let do') and also to passives ('be done'), since an initial non-subject constituent is promoted to subject position.

There are also complex causative-passive, causative-causative, and causative-causative-passive forms such as Uyghur *körsütül-* 'be shown', Turkish *öldürül-* 'be killed', *öldürt-* 'cause to kill', *öldürtül-* 'be caused to kill'.

However, causative suffixes cannot be followed by so-called passive suffixes unless these have antitransitive or reflexive readings, e.g. Turkish *kıvrıldı-* 'cause to curl'. There is little need to causativise passives such as *açıl-* 'be opened', since the primary stems, e.g. *aç-* 'open', function as their transitive counterparts.

Possessive Constructions

Possessive constructions, corresponding to English 'have' constructions, are of the type possessor + genitive suffix + possessed entity + possessive suffix + *bar* 'existent' or *yoq* 'non-existent', e.g. Uzbek *Pulim bâr* ('my money existent') 'I have money', Uyghur *Dadisi yoq* '(S)he has no father'; compare, however, the Khalaj type *hat vâ^arum* 'I have a horse' with *vâ^ar* + possessive suffix. There are also locative constructions of the type Kirghiz *Anda kitep bar* '(S)he has a book', Turkish *Ali'de para var* ('at Ali money existent') 'Ali has money'. In non-finite clauses, the words *bar* and *yoq* are mostly replaced by a verb of the type *bol-* 'become, be' and its negation *bolma-*, e.g. Turkish *parası olmayan* '(the one)

not having money', *parası olmadığı* ('his/her not having money') 'that (s)he does/did not have money', *parası olmayarak* 'without having money'.

Comparative, Equative and Similitive Constructions

The comparative degree of gradable adjectives and adverbs is normally expressed by adding an ablative suffix to the segment denoting the standard of comparison, e.g. Turkish *buzdan soğuk* 'colder than ice', Uzbek *mëndän yaxşı* 'better than I', Chuvash *turan pısäk* 'bigger than the mountain'. The adjective may be preceded by a word for 'more', or, as noted above, followed by a comparative suffix e.g. Turkish *daha iyi*, Tatar *yaxşıraq* 'better', Chuvash *pısäkraq* 'bigger', Uyghur *uzunraq* 'longer'. Superlatives are mostly formed with adverbs such as *eñ*, Chuvash *çi*, in front of the adjective, e.g. Uzbek *eñ yaxşı* 'best'.

In grammatical constructions expressing equality or similarity, e.g. 'A is as big as B' and 'A sings like B', a postpositional phrase or an equative suffix usually marks the standard of comparison ('as', 'like'), e.g. Azerbaijani *geder*, Uyghur *qeder* 'so much', Azerbaijani *kimi* 'like', Turkmen *yäli*, Tatar *këbëk*, Yakut *kurduk*. Thus, 'white as snow' may be rendered as Karakalpak *qarday aq*, Uzbek *qardây äppâq*, Tatar *qar şikëllë aq*, Yakut *xâr kurduk mayan*, etc. The standard marker may be accompanied by a reinforcing emphatic element such as Turkish *tıpkı* 'exactly, just'.

Order of Constituents

Since the Turkic languages are, as noted above, basically head-final, the unmarked order of constituents is subject + object + predicate core. The predicate core is normally clause-final, preceded by complements and free adverbials, the subject being the first complement. Adverbials referring to the entire clause often appear initially. Postpositions per definition take the final position in postpositional phrases. Note that the order of determination within the clause is the reverse of the successive modification through suffixes within the word. The order rules are also valid for non-finite clauses, and even tend to be more rigidly observed there than in finite clauses.

Finite Clauses

Predications are realised as clauses, basic text-building units occurring in pluripredicative constructions within the limits of sentences. They may be finitised as main clauses and non-finitised as dependent clauses.

In the former case, they function as independent sentences, optimally marked for aspect, mood, tense, personal reference and illocution. Their first actants are realised as subjects in the nominative.

In yes/no questions, an interrogation marker is placed after the whole predication or the constituent asked for, e.g. Turkish *Ali bugün geliyor mu?* 'Does Ali come today (or does he not)?', *Ali mi bugün geliyor?* 'Is it Ali who comes today (or who)?', *Ali bugün mü geliyor?* 'Is it today Ali comes (or when)?' The placement of the marker in relation to personal markers varies

across languages, and also some language-internally to some degree, e.g. Uyghur *Sen muellim-mu-sen?* ~ *Sen muellim-sen-mu?* 'Are you a teacher?'; compare the only Turkish option *Sen öğretmen misin?*

The verbal negation can vary its scope without changing its place, whereas more mobile nominal negation particles such as *emes* and *değil* (Chuvash *mar*) 'not' may be used specifically for constituent negation, e.g. Uyghur *Bêşim emes, közüüm ayıydu* 'What aches is not my head, but my eyes'. Note that certain verb forms of a stronger participial nature do not use *-mA*, but such particles, e.g. Tatar *Kileçek tügël* '(S)he is not coming', Turkmen *Men gitçek dæl* 'I will certainly not go'.

Chuvash has a special imperative negation: the postposed particle *mar* in the first persons and the preposed element *an* in the other persons, e.g. *An kil* 'Do not come!'

Discourse-Pragmatic Functions of Word Order

Finite clauses often display systematic deviations from the unmarked constituent order. This versatility is motivated by modifications of the functional sentence perspective, by the assignment of communicative, discourse-pragmatic functions.

Postpredicative Elements

The permitted deviations often include loosening of the final position of the predicate core. The postpredicative position may be occupied by extraposed elements from the clause, defocused constituents, often representing topics already activated, afterthoughts, etc. The elements are often detached from the preceding predicate core by an intonational juncture, e.g. Turkish *Gitti, postacı* 'He has already gone, the postman'. The postpredicative position seems to be a natural place for some types of sentence adverbials, for example those expressing comments on the proposition such as Turkish *herhalde* 'certainly, probably'.

The postpredicative position is not the position for new information, for interrogative pronouns and adverbs or for unmarked direct objects with specific reference. Subject pronouns found in this position cannot have topic function. Note that postpredicative elements are not confined to less carefully planned speech, but have also been part of written varieties throughout the history of Turkic.

Topic

There are mostly no fixed syntactic configurations to the left of the predicate core. Since the syntactic relations between constituents are clear from case-marking, word order modifications may be used for prominence relations, notably for topic and focus assignment.

As expected, the initial position is the preferred place of the sentence topic, the constituent which often contains known information and about which the rest of the predication gives new information. In rhetorically neutral sentences

it is identical with the subject, e.g. Turkish *Ali resmi çekti* 'Ali took the picture'. There are also special topicalisation markers following the topicalised nominal phrase such as Turkish *ise*, Turkmen *bolθa* ('if it is') 'as for', e.g. Turkish *Ali ise resmi çekti* 'As for Ali, he took the picture'.

Other constituents may be topicalised by taking the initial position, e.g. *Resmi Ali çekti* 'As for the picture, Ali took it'. A direct object can be topicalised by passive constructions, e.g. *Resim Ali tarafından çekildi* 'The picture was taken by Ali', but this option is less often chosen because of the devices offered by the constituent order.

Focus

The position immediately in front of the predicate core is used for focused constituents, offering new or relatively important information. This is the natural position for unmarked direct objects and for interrogative pronouns asking for new information, e.g. Tatar *Sin këmnë kürdëñ?* 'Whom did you see?' A non-object in this position is emphasised, so that the sentence cited above, *Resmi Ali çekti*, may also be rendered as 'It was Ali who took the picture'.

This device is more common than only marking focus by stress *in situ*, as in English ('Ali took the picture'), where non-subjects normally do not take the initial position in active clauses. Another method is to use a cleft sentence with a participle as subject in initial position and the focused element as a nominal predicate, e.g. Turkish *Resmi çeken Ali'ydi* ('The one taking the picture was Ali') 'It was Ali who took the picture'.

Position and Specificity

In most Turkic languages, the position in front of the predicate core is open to non-topical nominative direct objects, e.g. *Ali (bir) resim çekti* 'Ali took a picture', and thus allows a formal contrast with respect to specificity. An accusative object in this position seems to be marked as 'specific', e.g. Turkish *Ali, resmi çekti* 'Ali took the picture', *Ali, bir resmi çekti* 'Ali took a certain picture', whereas a nominative object is not. When the direct object is separated from the governing verb, accusative marking is mostly necessary, e.g. Uyghur *Ĵawapni muellim berdi* 'It was the teacher who gave the answer'. The topicalisation itself may also often suggest 'specific' readings.

Participant Reference

Most Turkic languages are relatively unexplicit with respect to participant reference. Under certain conditions, actants may remain overtly unexpressed. A constituent representing an old text topic may often be omitted. Thus, anaphoric pronouns can also largely be dispensed with if the referent is assessed as identifiable from co-text and/or situational context, e.g. Azerbaidjanian *Gelir* '[(S)he/it] comes'. This is also true of causers and causees in causative patterns, e.g. Turkish *Çalıştırıyor* '(S)he lets [someone] work'. Explicit expression is, however, normally required in a predication that starts

a discourse. A subject pronoun such as *o* 'he, she, it' cannot introduce a new text topic if it occurs in postpredicative position.

Non-finite clauses

Action nouns, participles and converbs allow non-finite realisations of predications as embedded clauses. The suffixes of these categories function as subjunctors, marking the clauses as embedded, much like English subordinative conjunctions and relative pronouns or adverbs. As dependents, non-finite clauses typically precede their heads. The embedding is recursive in that the clauses may contain non-finite clauses themselves. They may also be coordinated with each other. Their thematic suffixes occupy positions corresponding to those of aspect-mood-tense suffixes of finite clauses, but the inventories in question are different and less elaborate. For example, not all forms carry personal markers.

Action Clauses

Clauses based on paradigmatic verbal nouns, action nouns, are complement clauses, mostly corresponding to English *that* clauses. They mostly take on possessive personal markers, and their syntactic functions in the matrix clause are marked by case markers or postpositions. The formal relationship with their subjects mostly corresponds to genitive constructions, e.g. *Orhan'ın bildiği* ('Orhan's knowing') 'that Orhan knows/knew'. The nominative is also often used, e.g. Chuvash *açasem kayni* 'the children having gone'. In cases where the subject can take on both cases, genitive-marking tends to express 'specificity', much like the accusative-marking dealt with above, e.g. Turkish *para/paranın kaybolduğu* 'the fact that money/the money disappeared'.

There is usually a difference between factive and non-factive action clauses. The former are construed with suffixes such as *-GAn* and *-DİK* + possessive suffixes, e.g. Turkish *Ali'nin burada oturduğunu söylüyor* '(S)he says that Ali lives/lived here', Uzbek *Kelgânini bilâmân* 'I know that (s)he has come', *Âybekniñ bu kitâbni yazğanini bilâmân* 'I know that Aybek has written this book', Kirghiz *Men sizdin ayılya bolyonuğuzdu bilem* 'I know that you have been to the village', Turkmen *Nirâ göçenini bilemôq* 'I do not know where (s)he has moved'. The suffixes used have various aspecto-temporal meanings. Thus, Uzbek *-(â)yâtgân* is intraterminal, presenting the event as current, while *-Gân* is not.

Action clauses based on nominal predicates use copulative forms such as Uzbek *ekân*, Turkish *olduk* or extended forms such as Uzbek *ekânlık* and Kirghiz *ekendik* + possessive suffixes. Forms in *-LIK* and *-DİK* go back to the copula verb *er-* 'be', compare Ottoman *idük*, e.g. in Turkmen *Onuğ ayamdiyini tanadim* 'I realised that he was my elder brother'. Suffixes of this kind can be added to participial forms, e.g. Turkmen *-A(:)ndİK* and *-yÂndİK* in *gelendiyi* 'his/her having come' or *ölyændiyi* 'his/her being dying', Kirghiz *-GAndİK* in *Bala oquyandiyin bilem* 'I know that the child reads', Uzbek *-GânlİK*, *-(â)yâtgânlık*, etc. Such extended action noun forms are not

used as attributes to form relative clauses.

Action clauses referring to non-factive actions are based on suffixes such as *-mA*, *-(I)ş* and *-(U)w*, e.g. Tatar *ěşlěw* 'work(ing)', Kazakh *žasaw* 'do(ing)', Kumyk *getiw* 'go(ing)'. They can be used to form purpose clauses such as Turkish *Ali gelmelerini söylüyor* ('Ali says their coming') 'Ali says that they shall come', or infinitival complement clauses, often governed by modal lexical items, e.g. Uzbek *Ketişimiz kerâk* ('our going necessary') 'We must go'. Such clauses often imply first-actant coreference with their verbal head without carrying personal markers, e.g. Uzbek *Mên içišni istâyman* 'I want to drink', Turkish *Ali gelmek istiyor* 'Ali wants to come'.

Relative Clauses

Relative clauses, nominal actor clauses, refer to participant entities to which actions are ascribed, and are based on participles. They can be used as the head of a nominal phrase, e.g. Turkish *gelen* '(the) one coming/having come', Kirghiz *aytqanı* 'what (s)he has said', Uyghur *eytqanliriñiz* ('your said [things]') 'what you have said', Tuvan *körgenim* 'what I have seen', Yakut *üelēbet* 'who does not work'. They can also be attributed to a head in the sense of an identity attribution (see p. 50), e.g. Uyghur *kelgen kişi* 'a/the man who has come', Azerbaijani *oxuduğum kitab* 'a/the book I read'.

The participial markers used, e.g. Turkish *-(y)An*, *-DİK*, *-(y)AcAK*, *-mİş*, *-mİş olan*, Uzbek *-Gân*, *-(â)yâtgân*, *-âdigân/-ydigân*, or Tuvan *-(V)r*, *-GAn*, are relative subjunctors comparable to English relative pronouns and adverbs. They also signal various aspecto-temporal meanings. There are thus post-terminal participles as in Kazakh *kelgen kisi*, Chuvash *kilnē šin* 'a/the person who has arrived', Tuvan *uşqan iyaş* 'a/the fallen tree', more focal intraterminal participles as in Kazakh *kele žatqan kisi*, Uzbek *kelayâtgân kişi*, Chuvash *kileken šin* 'a/the person just coming', less focal intraterminal participles as in Uzbek *yazâdiyân kişi* 'a/the (usually) writing', prospective participles as in Tuvan *kelir çil* 'a/the coming year', etc. Some subjunctors have a more general terminal use, e.g. the Turkish and Azerbaijani markers *-(y)An* and *-DİK*. Predicates such as *gören* and *gördüğü* can thus be rendered as 'who sees/saw/has seen' and 'whom [somebody] sees/saw/has seen'.

Coreference with the First Actant

If the first actant of the relative clause is coreferential with the head, it is not signalled by any personal marker, e.g. Tatar *kilgen kēsē*, Chuvash *kilnē etem* 'a/the man who has come'. In this case, Turkish and similar Oghuz varieties use the type of relative subjunctors that do not take on personal markers, e.g. *-(y)An*.

The same situation obtains if the head is coreferential with the 'possessor' of the subject of the relative clause, e.g. Turkish *evi yanan adam* 'a/the man whose house is burning/burnt', Uzbek *ânâsi kâsâl bolyân bâlâ* 'a/the child whose mother is ill', Kirghiz *qizi sulû bolyon kişi* 'a/the person whose daughter is beautiful'. Here the heads, *adam* 'man' etc., are coreferential with

the genitival attributes of the corresponding constructions *adamın evi* 'a/the man's house', etc.

There are also similar nominal constructions, e.g. Uzbek *kállasi káttá kiši* 'a/the man whose head is big', where *kiši* 'man' is coreferential with the possessor of *káttá kállá* 'big head', Kirghiz *balıyı bar köl* ('its fish existent lake') 'a/the lake full of fish'. Possessive constructions of this kind, e.g. Uzbek *ânasi kásal* 'whose mother is ill', Kirghiz *qolu açıq* ('his/her hand open') 'generous', *eli köp* ('its people numerous') 'densely populated', are sometimes called 'bahuvrihi' compounds by analogy with the pattern exemplified in Old Indian *bahu vrıhi* ('abundant rice') 'rich in rice' (with the reverse order of elements).

Coreference with Other Entities

The head may also be coreferential with an entity other than the first actant, i.e. another actant or an entity expressing time, location, direction, source, instrument, etc. In such cases most Turkic languages make use of the same participles and the same order of elements.

An isolated syntagm such as *körgeñ kiši*, Chuvash *kurnă şin* etc. may thus in principle mean both 'a/the person who has seen' and 'a/the person whom somebody has seen'. The precise relation between the modifying clause and its head noun is established by pragmatic means alone. As noted above, this allows 'impersonal' expressions that do not refer to any specific first actant referent, e.g. Uzbek *bâryân yér* 'a/the place [someone has] gone to', Kirghiz *jazyñ qat* 'a/the letter [someone has] written'. With causatives, the agent is in the dative, e.g. Kirghiz *maya soyduryñ qoy* 'a/the sheep which they [= someone] had me slaughter'. These constructions should not be confused with constructions containing passives such as Kirghiz *jazyñ qat* 'a/the written letter'.

The first actant may be expressed by a subject, e.g. a subject pronoun as in Azerbaijani *men açan gapı* 'a/the door I open(ed)', Uzbek *mên yâzyân kitâb* 'a/the book I have written', *mên bâryân yér* 'a/the place I have gone to', Tatar *sin körgeñ kěşşë*, Chuvash *esë kurnă etem* 'a/the man you have seen', *esë yulakan yurâ* 'a/the song you are singing', *epir kayakan poyezd* 'a/the train we are going in', Kirghiz *ayam jazbayan qat* 'a/the letter that my father has not written', Uyghur *men oquyan kitap* 'a/the book I have read'. This simple construction is common in many languages.

The first actant may also be expressed by a personal marker of the possessive kind. This marker may be added to the head of the construction, e.g. Uzbek *yâzyân kitâbım* 'a/the book I have written', *bâryân yérım* 'a/the place I have gone to', Kumyk *satıb alyan atım* 'a/the horse I have bought'. A corresponding subject referring to a specific entity normally stands in the genitive, e.g. Turkmen *meniñ alan kitabım* 'a/the book I have bought', Uzbek *Návâiniñ yâzyân asârlari* 'works/the works which Navâi has written', *mên(iñ) yâzyân kitâbım* 'a/the book I have written', *mên(iñ) bâryân yérım* 'a/the place I have gone to'.

The personal marker may also be added to the participle, e.g. Old Turkic *eşitmişî* 'what (s)he has heard', Turkish *aldığım* 'what I take/took/have taken'. Turkish and partly Azerbaijani use special types of subjunctors in these cases, *-DIK*, *-(y)AcAK* etc., e.g. Turkish *oturduğum köy* 'the village I live(d) in', *geldiğim gün* 'the day I arrive(d)'. The presence of personal markers requires genitive subjects, e.g. *Orhan'ın gördüğü adam* 'the man whom Orhan sees/saw/has seen'. A corresponding relativisation out of possessive constructions is possible when the head is coreferential with a 'possessor' of an entity other than the first actant, e.g. Turkish *babasını tanıdığım çocuk* 'the child whose father I know/knew/have known'.

Thus, relative constructions such as *gördüğün at* 'the horse you see/saw/have seen' and *yediğim et* 'the meat I eat/ate/have eaten' correspond to the Kirghiz type *sen körgön at, men jegen et*. However, even Turkish may use the latter type, notably if the first actant referent is non-topical/non-specific and the head refers to a place, e.g. *su akan yer* 'a place where water flows', *dondurma satılan dükkân* 'a/the shop where ice-cream is sold'. It can also, for example, use the participle in *-(y)AcAK* in the impersonal way discussed above, e.g. *okuyacak bir kitap* 'a book to read'; compare Kirghiz *oqur kitep*, Bashkir *uqihi kitap*.

As noted above, the precise relation between the relative clause and its head is not specified by the syntactic construction itself. Thus, Kirghiz *men kelgeñ şâr* means 'the town I have come to' or 'the town I have come from', and Turkish *taşındığım ev* may mean 'the house I move/moved have moved into' or 'the house I move/moved have moved out of'. If necessary, the relation may of course be specified, e.g. by means of possessive postpositions such as *içine* 'into its interior' and *içinden* 'from its interior', e.g. Turkish *içine taşındığım* 'into which I move(d)' and *içinden taşındığım* 'from which I move(d)'.

Non-propulsive Relative Clauses

Due to the left-branching subordination, prepositive Turkic relative clauses are not propulsive, i.e. cannot advance the plot beyond the point of time represented by the head clause unless the sequence of the events is specified by other means, e.g. Turkish *Taşındığım evi buldum* '*I found the house, which I moved into'.

Converb Clauses

Converb clauses, adverbial action clauses, are based on converb markers. The latter are mostly monofunctional, i.e. do not serve to form other clause types. Some are morphologically simple, e.g. Turkish *gelerek* 'coming', Uzbek *yâzyâş* 'having written', whereas others may be rather complex, for example, consisting of combinations of participles with case markers, postpositions or nouns, e.g. Kirghiz *kelgende* 'when arriving', Turkish *geldikten sonra* 'after coming', *geldiği zaman* 'when (s)he comes/came'. Subjects of converb clauses are usually in the nominative. Some markers require identity between the first

actants of the converb clause and the head clause, e.g. Yakut *-A(:)n*, Turkish *-(y)Ip* (with few exceptions). Most markers do not require this identity, e.g. Turkish *Ali gelince Osman şaşırdı* 'When Ali came, Osman was surprised'.

Semantic Relations

Converb markers express various semantic relations, a wide range of aspectual, temporal and other relational concepts: intraterminality, post-terminality, terminality, anteriority, posteriority, temporal inclusion ('while'), abtemporality ('since'), terminativity and limitation ('until', 'as long as'), reason ('because'), means, instrumental ('by'), purpose ('in order to'), contrast ('whereas'), accordance ('as'), substitution ('instead of'), preference ('rather than'), condition ('if'), etc.

Conditional clauses are usually based on suffixes of the type *-sA*, with personal markers of the possessive kind, and copula particles of the type *ése* 'if ... is'. The conditional forms may contain particles and auxiliary verbs to express various temporal and other distinctions, e.g. Uyghur *yazyan bolsaq édi* 'if you had written'. In many languages, conditional clauses may be introduced by copies of the Persian conjunction *agar*, e.g. Uyghur *eger yazsam* 'if I (should) write'.

Subjunctors of the types *diye, dep* 'saying' (see pp. 48, 117) form other kinds of adverbial clauses, e.g. purposive and motivating causal clauses containing optatives or prospectives, e.g. Turkish *ev yansın diye* ('saying: the house may burn') 'in order for the house to burn', Chuvash *saxăr iles tese* ('saying: will buy sugar') 'in order to buy sugar'.

Non-modifying Converb Clauses

Not all converb clauses are used in a propositionally restrictive sense, modifying the head clause semantically. Those based on markers of the *-(V)B* type, Yakut *-(A)n*, Chuvash *-sA*, Tuvan *-GAš*, etc. may, though they are syntactically dependent clauses, represent events of equal narrative value with the event of the head clause. They can thus often be rendered with English coordinative constructions, e.g. Güllüp *odadan çıktı* '(S)he laughed and left the room', Kirghiz *Oqup tüşündü* '(S)he read and understood', Chuvash *Kilse kurčë* '(S)he came and saw it'. These non-modifying converbs are terminal and thus propulsive, capable of ordering events linearly, e.g. Turkish *Ali kayıp düştü* 'Ali slipped and fell'. Beside their purely joining function, these markers may get various contextual readings such as anteriority, cause, concession, manner and condition. Ambiguity between converb clauses of this kind and the postverb constructions mentioned above (p. 42), e.g. Kirghiz *Oqup turdu* '(S)he read it and stood up' versus '(S)he kept reading', are prevented by prosodic features, notably pitch patterns.

Non-modifying terminal converbs of this type are ideal for the discourse function of narrative linking and serve as central text-constructing units in traditional narrative styles. They allow periodic chain sentences of considerable length to be constructed, corresponding to whole text paragraphs in other styles:

extended sequences of propulsive clauses, representing thematically closely interconnected narratively equal events, e.g. Kirghiz *Men erteñ menen turup zaryadka žasap kiyinip žünup mektepke baražatam* 'I get up in the morning, do gymnastics, dress, wash myself and go to school' (see pp. 353–4). No link in such chains necessarily modifies the next one, and only the final head clause is fully equipped to permit a definitive personal, aspectual, modal, temporal and illocutional interpretation of the chain. Different converb types may be used as text-subdividing devices signalling different degrees of thematic closeness between the events of the chain. These old and widespread narrative patterns are vanishing in more Europeanised styles.

Questions of Scope

Since finite clauses are asserted as a whole, embedded modifying converb clauses are outside the scope of illocutionary markers signalling assertion. They are also outside the scope of a negation of the head clause, and must take on negative markers of their own. Negative converb markers such as Turkish *-mAdAn*, Tuvan *-BAyn* etc. are modifying ('without doing/having done'), e.g. Turkish *Selam vermeden gittim* 'I left without greeting'. Non-modifying converb clauses may, however, be included in the scope of negation, e.g. Turkish *Gelip görmedin* 'You did not come and [did not] see'. There are also non-modifying negative markers of the type *-mAyIB*, e.g. Turkish *Selam vermeyip çıktım* 'I did not greet, but left'.

A final interrogative particle questions the content of a whole sentence, e.g. Turkish *Ali gelince gittin mi?* 'Did you leave when Ali came?' As with all modifying adverbial constituents, the content of a modifying converb clause may be interrogated separately by shifting the position of the interrogation marker, e.g. Turkish *Ali gelince mi gittin?* 'Was it when Ali came that you left?'

Other Forms of Junction

There are also other forms of clause junction, commonly regarded as copies of Indo-European subordinative structures, e.g. postpositive clauses which modify the preceding clause or a constituent of it. They are provided with predicative markers typical of finite clauses and preceded by junctors such as *kim* or *ki* 'who, that', *qačan* 'when' etc.

Though such clauses may resemble English constituent and relative clauses, they often differ considerably from them. The junctor signals a close semantic connection with the subsequent clause, but normally does not introduce it, e.g. Turkish *Sanırım ki, gelecek* 'I think (s)he will come', Uzbek *Mën bildim ki, siz këlğän édigiz* 'I got to know that you had come'. Clauses of this type are mostly not embedded as postpredicative constituents of the preceding clause, but appended to it in a looser way. In certain modern languages such as Turkish, most imitations of this kind are not stylistically acceptable any more.

On the other hand, languages which have been under strong Persian or Slavic influence tend to reproduce the originals in a more genuine way. Some

of them make use of native junctors modelled on foreign ones, e.g. Gagauz *ani* 'who, which, that'. There are even tendencies to use conditionals and optatives in the sense of Indo-European subjunctives.

Free junctors plus conditional or optative predicate cores ('whoever', 'whenever', etc.) are very often used for generalising relative and adverbial constructions, e.g. Old Uyghur *kim qorqsar* 'whoever fears', Turkish *ne olursa olsun* 'whatever may happen'. Languages under predominantly Slavic influence prefer indicative forms in such cases, e.g. Chuvash *Käm eşlemest, vâl şimest* 'Who does not work, does not/should not eat' (present tense).

Further Reading

Since existing general descriptions of Turkic structure have a more or less diachronic orientation, most of the relevant literature will be listed in the bibliography added to Chapter 5. References to descriptions of Turkish structure will be given in Chapter 11. The following list includes titles in which the terminology and the analyses adopted in this chapter are elaborated on.

- Csató, É. Á. & Johanson L. (1994) 'On gerundial syntax in Turkic', *Acta Orientalia Hungarica* 46, 133–41.
- Johanson, L. (1971) *Aspekt im Türkischen. Vorstudien zu einer Beschreibung des türkeitürkischen Aspektsystems*, Studia Turcica Upsaliensia I, Uppsala: Almqvist & Wiksell.
- (1990a) 'Subjektlose Sätze im Türkischen', in B. Brendemoen (ed.) *Altaica Osloensia*, Oslo: Universitetsforlaget, 193–218.
- (1990b) 'Studien zur türkeitürkischen Grammatik', in G. Hazai (ed.) *Handbuch der türkischen Sprachwissenschaft* 1, Budapest: Akadémiai Kiadó, 146–278.
- (1991a) 'On syllabic frontness oppositions in Turkic', in *Varia Eurasistica*, Szeged: University of Szeged, 77–94.
- (1991b) *Linguistische Beiträge zur Gesamtturkologie*, Bibliotheca Orientalis Hungarica 37, Budapest: Akadémiai Kiadó.
- (1992a) *Strukturelle Faktoren in türkischen Sprachkontakten*, Sitzungsberichte der Wissenschaftlichen Gesellschaft an der J. W. Goethe-Universität Frankfurt am Main 29: 5, Stuttgart: Steiner.
- (1992b) 'Periodische Kettensätze im Türkischen', *Wiener Zeitschrift für die Kunde des Morgenlandes* 82, 201–11.
- (1993) 'Typen türkischer Kausalsatzverbindungen', *Journal of Turkology* 2, 213–67.
- (1995) 'On Turkic converb clauses', in M. Haspelmath and E. König (eds) *Converbs in Cross-linguistic Perspective. Structure and Meaning of Adverbial Verb Forms – Adverbial Participles, Gerunds*, Empirical Approaches to Language Typology 13, Berlin, New York: Mouton de Gruyter, 313–47.
- (1996) 'Terminality operators and their hierarchical status', in B. Devriendt, L. Goossens and J. van der Auwera (eds) *Complex Structures: A Functionalist Perspective*, Functional Grammar Series 17, Berlin, New York: Mouton de Gruyter, 229–58.
- (1998) 'Viewpoint operators in European languages', in Ö. Dahl (ed.) *Tense and Aspect in the Languages of Europe*. Berlin, New York: Mouton de Gruyter.

4 The Reconstruction of Proto-Turkic and the Genetic Question

András Róna-Tas

The Concept of Proto-language

A proto-language is a reconstructed language. It is neither intended to be identical with the spoken language of a given period, nor to be a self-contained description of it. It is rather a model in the form of a set of hypotheses. If the supposed features existed in the reconstructed language, features of known languages can be deduced by assuming relatively simple regular historical changes, which can be described as dominant tendencies and in terms of regular correspondences. The proto-language is a formula which helps understand the differences of the later varieties of a language. On the other hand, it can itself only be reached by analysis and comparison of later documented languages. An older proto-language such as Proto-Turkic can be helpful in reconstructing a more recent one such as Proto-Kipchak. But Proto-Turkic can only be reconstructed by taking the Kipchak languages, thus also Proto-Kipchak, into consideration. The reconstruction is a two-way traffic, going from existing languages to the proto-language and back again.

Proto-Turkic

Proto-Turkic and Ancient Turkic

While Proto-Turkic is a reconstructed language, Ancient Turkic is a language that has existed and has been spoken by a language community. Ancient Turkic lasted from its formation until the consolidation of the First Türk Kaghanate in the middle of the sixth century AD. Thus, Ancient Turkic can be investigated empirically, while Proto-Turkic is a linguistic construct.

Chronological Limits

Even if a proto-language has not been a living language, it has had a certain time depth and a certain geographical relevance. Proto-languages have an

overall upper chronological limit, the end of the Mesolithic Period or rather the beginning of the Neolithic Period. The changes connected with the Neolithic Revolution created the sociolinguistic conditions for a system of communication that led to the type of languages we know. In the region concerning us here, the Neolithic Period started at about 4500–4000 BC. If Proto-Turkic reflects a language of the first period, it began around this time. If there was an earlier proto-language, e.g. a language in the sense of a reconstructed Proto-Altaic, Proto-Turkic came into existence only after its dissolution. The lower limit of Proto-Turkic is the time of the appearance of the first direct data from existing Turkic languages, in fact after the separation of the branches of Turkic, i.e. about the middle of the first millennium BC.

Proto-Turkic must be conceived of as a successive set of synchronous systems in the Ancient Turkic period, a period from which we have no direct data, but which we can reconstruct by comparing existing Turkic languages and their historical documents. The Proto-Turkic we can reconstruct best is the layer nearest to our sources. We denote this system as Proto-Turkic₀, or simply Proto-Turkic, and earlier systems as Proto-Turkic₋₁, Proto-Turkic₋₂ etc. As for Ancient Turkic, it can be divided into early Ancient Turkic, the period in which no stabilised dialects existed, and late Ancient Turkic, when the dialect boundaries were stabilised.

The Question of the Original Homeland

The term *urheimat* 'original homeland' denotes the region where a language has come into being, in our case the place of the formation of Turkic. However, it is not necessary to suppose that the ancestors of the Turks remained at the place where Turkic had been formed up to the dissolution of Ancient Turkic. If the entire Turkic-speaking group moved from there to relatively distant areas, a second, a third and even more *urheimats* can be implied.

There are several methods of reconstructing the place of an *urheimat*. The lexical stock of the proto-language can give hints through words denoting parts of the natural environment peculiar to a region: flora, fauna, climate. Early loanwords from other (proto-)languages may be helpful, if the geographical setting of these languages is known. Borrowing mostly implies language contact, which means historical contacts among peoples. Archaeologists try to trace archaeological cultures of identified speaker groups back to earlier groups. Such methods have their limitations and must be used with great caution. Even then, the only *urheimat* we can determine is the last one, the place where the Turks lived before the dissolution of the Ancient Turkic unity. The last habitat we can reconstruct with our data and existing methods can be placed in west and central Siberia and in the region south of it.

Early Dialects and Isoglosses

Though Ancient Turkic must be considered as a successive set of systems of one language, this unity was not a homogeneous one. In early Ancient Turkic,

from about 3000 until 500 BC, no stabilised dialects can be assumed. In late Ancient Turkic, the 'ancestor dialects' of the later Turkic languages were formed. At the end of early Ancient Turkic, the formation of the isoglosses *r/l* and *z/š* had begun. This means that, in special phonetic environments, the oppositions *r* vs. *z* and *l* vs. *š* were neutralised and new sounds appeared. Later on, the change was generalised areally to the effect that certain words were pronounced with *z* and *š* in one area and with *r* and *l* in another area, e.g. **buzagu* vs. **buragu* 'calf', **tāš* vs. **tāl* 'stone'. Following the split into different Turkic branches, one bunch of isoglosses was generalised in one branch and another bunch in the other. However, certain features existed in more than one branch. Thus, Proto-Turkic initial **y-* was replaced by *ǰ-* not only in the ancestor of the *r/l* languages, but also in some of the *š/z* languages. Furthermore, the lexical isoglosses of Ancient Turkic did not always follow the phonological ones. Thus the dialect distribution of the words *tūš* and *tül* 'dream' did not coincide exactly with the isoglosses of *š* and *l*. As a result, *tül* is found in Old Uyghur texts beside the normal form *tūš*.

A Sketch of Proto-Turkic

Sources available for the reconstruction of Proto-Turkic include comparison of present-day Turkic languages, historical documents written in Turkic languages, loanwords copied by non-Turkic languages from Turkic languages and loanwords copied from non-Turkic languages by one or more Turkic languages.

While for a long time, the reconstruction of Proto-Turkic was mainly based on East Old Turkic and the Oghuz languages, a more thorough comparison of all known written and spoken languages began later on. Nevertheless, the Runic, Uyghur, Manichaean and other documents of East Old Turkic still play a key role in the reconstruction, since west Old Turkic languages such as Old Chuvash and Khazar are much less documented and thus difficult to reconstruct.

In the following outline of Proto-Turkic, special attention will be paid to features common to Chuvash and East Old Turkic. Some of them may be secondary, for instance, developed in Chuvash under Tatar influence, and some features that are absent may have been present in Proto-Turkic but lost later on. If these possibilities are kept in mind, the features common to East Old Turkic and Chuvash may serve as a solid basis for reconstructing Proto-Turkic.

Phonology

Vowels

As for vowel quantity, Proto-Turkic had two distinctive vocalic lengths which were realised in first syllables as short (or 'normal') vs. long vowels. Non-first syllables had short and reduced vowels, but it is still unclear whether there

was an opposition between a 'normal' and a reduced quantity. For a possible third Proto-Turkic vowel quantity, see p. 90.

The known Turkic languages show no traces of early diphthongs, but rather segments in which one part is a vowel and the other a consonant *w* or *y*. Still we cannot exclude the possibility that Proto-Turkic had genuine diphthongs with two vowel elements. They were then probably of transitory character, i.e. emerging from long vowels or reflecting foreign diphthongs in loanwords, e.g. Old Turkic *tōn* 'clothing' < **taunǎ* ← Indic *tauna*. Some interesting features in Khalaj can be attributed to old diphthongs (see pp. 93–5).

The shortening of long vowels was gradual and did not take place at the same time in all early Turkic languages. It is not even necessary to assume that all types of long vowels were shortened in all positions at the same time within one and the same language. Old Turkic exhibits a transitory stage in the process of shortening.

Theoretically, Proto-Turkic might have had four *e*-sounds, *e*, *ē*, *ê*, *ĕ*. It is at least safe to assume that it had a short closed *ê* beside an open *e*. It is, however, not necessary to assume a back vowel **ĕ* in order to explain Common Turkic *a* ~ Chuvash *ī* in certain words (see p. 92). In some Turkic languages, e.g. those spoken in the Volga region, *a* was rounded. The *a*-sounds that remained unrounded later changed into *ī* in Chuvash, Yakut and Tuvan (in different words). There is no reason to reconstruct a rounded *ā* for Proto-Turkic.

In non-first syllables, Proto-Turkic did not only have *-u/-ü*, but most probably also *-o/-ö*. The high vowels *u*, *ü*, *i* and *ī* were reduced word-finally and disappeared in Old Turkic, whereas the half-low ones became high: *-o* > *-u*, *-ö* > *-ü*, *-ê* > *-i*. The low *-e* was raised to *-ê*. The low *-a* partly remained as such and was partly rounded to *-ā* or raised to *ī*.

In certain Old Turkic documents, *a* is written instead of an expected *ī* in non-first syllables. While some scholars have taken this notation to represent

Table 4.1 Vowels in Proto-Turkic first syllables

Short vowels				Long vowels			
ī	u	o	ü i	ī	ū	ū	ī
		ö ê		ō	ō	ē	
		a e		ā			

Table 4.2 Vowels in Proto-Turkic non-first syllables

Short vowels				Reduced vowels			
ī	u	ö ê	ü i	ī	ū	ū	ī
		a e		ō	ō	ē	
				ā ě			

a special Old Turkic or even Proto-Turkic sound, it is probably a transcription of the still preserved back *ī*, mostly in the vicinity of *k*, *g*, *r*, *l* (cf. pp. 107–8).

It has also been claimed that traces of an old vowel gradation can be found in stem alternations such as *bi* 'I' ~ *bē-n-* (oblique stem). This is, however, a structurally conditioned sound change that seems more similar to an umlauting phenomenon.

Consonants

Proto-Turkic had a fortis (strong) vs. lenis (weak) system of obstruents, though the actual phonetic features are not known. The strong member of each opposition was unvoiced and maybe aspirated, whereas the other member was weak and most probably also unvoiced. There are no traces of a corresponding opposition of affricates, which would mean that *č-* had no weak counterpart in Proto-Turkic.

As for word-initial stops, the Proto-Turkic system represents a secondary simplification. The earlier oppositions *k* vs. *g* and *t* vs. *d* of Proto-Turkic₋₁ were neutralised and represented by *k-* and *t-*. As far as *p* vs. *b* is concerned, *p-* developed into a bilabial fricative and then into *h-*. Since the spirantisation of *p-* took place prior to the neutralisation of the opposition strong vs. weak with the other stops, *b-* was preserved as such. The initial *h-* disappeared in almost all languages. It remained unmarked in most Old Turkic scripts, even in cases where it still existed. Tibetan and Chinese transcriptions show that it was preserved at least in some Old Turkic dialects. The initial *h-* is also found in Khalaj (see pp. 96, 281). However, not all cases of *h-* in Old Turkic and Khalaj go back to Proto-Turkic *p-*. Some are of secondary and some are of foreign origin.

Proto-Turkic *y-* has three antecedents in an earlier Proto-Turkic₋₁, namely *y-*, *di-*, and *ń-*, reflected in Mongolian *y-*, *di-* and *ni-* respectively. In a large group of Turkic languages, Proto-Turkic *y-* of whatever origin changed into *j-*.

Of the two Proto-Turkic sibilants only *s-* occurred word-initially. All initial *š-* sounds in native Turkic words have developed from a primary or secondary *si-*. Proto-Turkic consonants not occurring in word-initial position are *n-*, *m-*, *ŋ-*, *l-*, *r-*, *z-*. As mentioned, however, *ń-* occurred in Proto-Turkic₋₁.

In non-initial positions, the consonants *p*, *b*, *k*, *g*, *t*, *d*, *č*, *s*, *š*, *m*, *n*, *ŋ*, *ń*, *r*, *z*, *l* occurred. Many old Turkic languages and dialects show spirantisation of weak obstruents, but we have no clear sources to determine when this change began.

In Chuvash and its earlier forms, *r* is found instead of *z* in other Turkic languages. This important feature of 'rhotacism' is reflected in early Turkic loanwords in Common Samoyed, Mongolian and Hungarian. While some scholars take it to reflect a special Proto-Turkic consonant or consonant combination, there is no need for such a reconstruction. Proto-Turkic had the opposition *z* vs. *r*, which was lost in the ancestor of Chuvash, but preserved

in all other Turkic languages. The loss was due to the change of *z* into *r*, a development that occurred twice in the history of Chuvash. The second time, it affected Proto-Turkic intervocalic and final *d*, which became *r* via an intermediate stage *z*. As shown by Tekin and others, *z* has developed into *r* in special, mainly pre- or post-consonantal, positions in other Turkic languages as well, e.g. *köküz* 'chest, breast' → *kökrek* 'chest, upper part of the body', *semiz* 'fat' → *semri-* 'become fat', *tīz* 'knee' → *tirsgek* 'elbow, knee joint' from **tiz-ge-k* (Hungarian *tér-d* from **tir* + Hungarian diminutive suffix *-d*) or Old Turkic *qoŋuz* 'beetle, insect' → *qomursga* 'ant' from **koŋuz* + *gA* > Chuvash *χāmār*. The change *z* > *r* thus began in special phonetic environments in all Turkic languages at the beginning of late Ancient Turkic. It later ceased except in the ancestor of Chuvash, where it was generalised in all positions.

In most Turkic languages, non-initial *š* corresponds to an *l* in Chuvash and its ancestors as well as in early Turkic loanwords in Common Samoyed, Mongolian and Hungarian. Scholars such as Ramstedt, Poppe and Doerfer have taken these cases of 'lambdaism' to represent a special Proto-Turkic sound or sound combination. The case is similar to that of rhotacism: Chuvash and its ancestors lost the opposition *š* vs. *l*. But not even in Chuvash was the change to *l* complete, since *š* is often found here instead of Proto-Turkic *š*, e.g. *bāš* 'head' > Chuvash *puš*. Turkic loans in Hungarian display *č* or *lč*, both going back to *č* < *š*, e.g. *bocsát-* ← **bočat-* < *bošat-* 'release', *gyümölcs* ← **jemič* < *yemiš* 'fruit'. Similar changes are found in other Turkic languages, mainly before or after a consonant, e.g. *tapšur-* 'hand over' > *tapčur-*. The next stage is rare, though witnessed, e.g. Turkic *taz baši* 'sparrow hawk' > **tar balči* ⇒ Mongolian *tarbalji*.

Word Structure

Word types of the structure V and CV are found in the earliest Turkic documents, e.g. *ī* 'vegetation', *tū* 'body hair'. The existence of the same structures in Proto-Turkic cannot be excluded. Often, however, it can be shown that cases of CVC had an additional vowel, which later disappeared. The basic types should have been CVCV̄ and CVCV. The reduced *-V̄* disappeared, e.g. Proto-Turkic **sagā-* > Old Turkic *say-* 'milk' ⇒ Mongolian *saya-*, Proto-Turkic **kertī-* 'cut a notch' > Old Turkic *kert-* ⇒ Mongolian **kerti-* > *kerči-*. The non-reduced vowels remained, e.g. Proto-Turkic *teke* 'goat' > Old Turkic *teke*, Proto-Turkic *kara* 'black' > Old Turkic *qara*. The gradual loss of the final reduced vowels in Old Turkic has been demonstrated by Johanson (1979). They first disappeared after the sonorants *l*, *r* and the nasal dental *n*. Since this phase was concluded at the beginning of Old Turkic and is reflected in Chuvash, it can also be presumed for Proto-Turkic. The existence of reduced vowels after *l* is shown by some proper names. Thus, one and the same name is found in two different forms in the inscriptions: *Köl* and *Köli*.

Though the Turkic word accent is generally on the last syllable, the

disappearance of the final reduced vowels may point to a different situation in Proto-Turkic. A similar piece of evidence is the Old Turkic alliteration.

Consonant clusters were in general not possible in initial and final positions. Old Turkic final clusters such as *-rt*, *-lt*, *-lp*, *-nt*, *-nč*, *-st*, *-rk* seem to be the result of secondary developments, e.g. *and* 'oath' < Proto-Turkic **andā* ⇒ Mongolian *anda* 'sworn brother'.

Proto-Turkic had front vs. back sound harmony. Most features in documents and present languages contradicting this, e.g. the loss of *ī* in some central Asiatic languages or the drastic changes in Chuvash, are late secondary developments and not preserved Proto-Turkic traits. While almost all suffixes had a front and a back variant, some had only one form, e.g. the front third-person singular possessive suffix *-i* and the back dative suffix *-ka*. There are also some traces of labial harmony, in a few cases even of an opposition low vs. high or low vs. reduced.

Morphology

Word Formation

Proto-Turkic had various morphological devices for word formation, the most frequent formatives being suffixes. Some word formatives became unproductive very early, e.g. those in *kö-z* 'eye', *kö-r-* 'see'. There was also partial reduplication as in *kap kara* 'very black' and full reduplication as in *adīn adīn* 'very dissimilar'.

Nouns

No traces of Proto-Turkic gender are found. The plural suffix was *-lAr*, whereas Chuvash *-sem* is a late replacement. Proto-Turkic had several suffixes denoting collectivity, e.g. *-An* as in *oglan* 'princes, bodyguards, boys' ← *ogul* 'male child'. There was no dual. Contrary to the opinion of Erdal and others, *-z* has never been a dual suffix or denoted pairs of body parts. However, an old suffix denoting plurality can be detected in the *-z* of personal pronouns, e.g. *biz* 'we' ← *bi* 'I'.

Proto-Turkic nouns probably had an oblique stem in *-n*, just as pronouns still have in Old Turkic. Four cases show traces of it: genitive **-n*, accusative **-nVG* (in the pronominal declension *-nI*), dative **-nKA* and instrumental **-nVn*. No *-n* is found in the locative **-dA*. The Proto-Turkic ablative suffix **-dAn*, an extension of the locative, exhibits an *-n*, though not as a stem extension.

The possessive suffixes in the singular were 1p. **-m*, 2p. **-ŋ* and 3p. **(s)i*. Though Chuvash has only *-i*, the existence of *-si* in the earlier history of Chuvash can be demonstrated. The plurals of the first and second persons were formed with the suffix *-z*. The third person shows no traces of a singular vs. plural opposition.

Pronouns

The personal pronouns were 1p.sg. *bi*, pl. *biz*, 2p.sg. *si*, pl. *siz*. Later on, third-person pronouns developed from demonstratives such as *ol* 'that'. Proto-Turkic may nevertheless have had a pronoun for the third person, possibly **a(n)-*, since the oblique stem of *ol* is *an-*; cf. Chuvash *un-*. The fact that Chuvash has a 3p.sg. *-ě* < **-i* in certain conjugations shows that Proto-Turkic had a third-person singular pronoun **i-* or **in-*. It developed into a suffix, but disappeared in other Turkic languages. Note that Proto-Mongolian had 3p.sg. **in-* and 3p.pl. **an-*.

Instead of the Common Turkic demonstrative *bo ~ bu*, Chuvash has *ku* < **kü*. Proto-Turkic has no traces of an inclusive vs. exclusive opposition in the first-person plural. The interrogative pronoun was **kem* 'who'. It is unlikely that Old Turkic *ne* 'what' reflects a Proto-Turkic form, since it would be the only native Turkic word with initial *n*.

Numerals

The Proto-Turkic numeral system was decimal. The numerals for digits were *bir* 'one', *eki* 'two', *üç* 'three', *dört* 'four', *běş* 'five', *altı* 'six', *yeti* 'seven', *sekiz* 'eight', *tokuz* 'nine'. The tens were *on* 'ten', *yegirmi* 'twenty', *otuz* 'thirty', *kirk* 'forty', *elig* 'fifty', *almış* 'sixty', *yetmiş* 'seventy', *sekizon* 'eighty', *tokuzon* 'ninety'. 'Sixty' and 'seventy' are thus formed with the unproductive suffix *-mış*, but 'eighty' and 'ninety' by compounding with *on* 'ten'. Since this system is common to Old Turkic and Chuvash, it can also be supposed for Proto-Turkic. Compounds with *-on* in the lower tens, as in some Siberian dialects and Yellow Uyghur, are products of later developments. The word *yüz* 'hundred' is also Proto-Turkic; cf. Chuvash *şer*. Even if Chuvash *pin* 'thousand' is a loan, Old Turkic *biñ* may still go back to Proto-Turkic; cf. Mongolian *mingan* 'thousand'. Old Turkic *tümen* 'ten thousand' is a late loanword, most probably from Tokharian. It is not present in Chuvash, but has been borrowed by Hungarian.

Some Turkic languages exhibit long consonants in some numerals, e.g. *ikki* 'two', *sekkiz* 'eight'. Chuvash has double forms, the longer form occurring in non-attributive and in 'determined' attributive function, e.g. *pillëk këneke* 'the five books' vs. *pilëk këneke* 'five books'. It is uncertain whether the longer forms have etymological relevance for the reconstruction of the Proto-Turkic numerals.

In Old Turkic and in present-day Yellow Uyghur, the cardinal numerals 11–19 are formed according to the pattern 'one of the second ten', e.g. Old Turkic *bir yigirmi* 'eleven'. The numerals 21–29 are constructed as 'one of the third ten', e.g. *bir otuz* 'twenty-one'. The absence of this system in Chuvash may mean that there has been a secondary development here, or that the system did not develop outside east Old Turkic.

The Proto-Turkic ordinal suffix seems to have had both a shorter form *-m* and an extended form *-Vmči*.

Verbs

Already in Proto-Turkic, the second-person singular simple imperative was identical with the verb stem, e.g. *kör* 'see!'. It is not clear which of the later suffixes for the other persons existed in Proto-Turkic.

As for voice, Proto-Turkic had a cooperative, e.g. *körüş-* 'see another', a middle, e.g. *körün-* 'become visible', a passive, e.g. *körül-* 'be seen', and a causative, e.g. *körtür-* 'show'. Negation was expressed with *-mA* added to the verb stem, e.g. *körmem* 'I do not see'.

The personal markers developed from personal pronouns. Even Old Turkic has structures such as *Ölteçi sen* 'You will die'. The suffixes were 1p.sg. *ben* > *-men* > *-m*, 2p.sg. *sen* > *-n*, 3p.sg. **i* > \emptyset , 1p.pl. *biz* > *-biz/-miz* and 2p.pl. *siz* > *-siz*. Besides, *-ŋ* is found in the second person and *-k* in the first person plural.

Verbal nominals played an important role in Proto-Turkic. Nominal stems of verbs functioned as attributives, adverbs and finite predicates. The past marker was *-dl*. The aorist marker *-(V)r* can also be traced in Chuvash (see pp. 116, 445). However, no Proto-Turkic future marker can be detected.

The Lexical Stock

Though there is disagreement on how to define a basic lexical stock, we may determine a group of words denoting concepts of everyday life which may have existed in the proto-language. A comparison of East Old Turkic data with Chuvash words denoting the same concepts shows that only about 65 out of 100 words can be compared without any problem. There is a surprisingly great number of basic words without a regular correspondence. Words with some additional elements in Chuvash are easiest to explain, e.g. Chuvash *uyax* < *ayiq* ← Old Turkic *ay* 'moon'. Another group includes Chuvash loans from other Turkic languages, e.g. *pisäk* ~ Old Turkic *bedük* 'big, great', *sarä* ~ Old Turkic *sariy* 'yellow'. Though the expected counterpart of Old Turkic *tiş* 'tooth' would be Chuvash **čäl*, we find *šäl* instead, which can only be explained by the development *tiş* > **čiş* > **šiš* > *šil*. This early shift is reflected in Hungarian *süllő* 'lucio-perca sandra, the fish with teeth' ← **šilliy*. Instead of **šerě* as the regular correspondence of Old Turkic *yürek* 'heart' we find Chuvash *čērě*, which may be a loan from a different Chuvash dialect in which the change *č* > *š* was slower. For Chuvash *mäy* ~ Old Turkic *boyun* 'neck', a Proto-Turkic *boñü(n)* may be implied. Chuvash *puyan* and Old Turkic *bay* 'rich' go back to Proto-Turkic **bayä(n)*. Chuvash *mäyraka* ~ Old Turkic *bügüz* 'horn' presupposes a Proto-Turkic *büñürke* with a palatalised *ñ* which became *ñ* > *y* in Chuvash, but was depalatalised in other Turkic languages. The high percentage of the seemingly irregular Chuvash forms in the basic lexical stock can only be due to the fact that the dissolution of the Proto-Turkic unity, the departing of the ancestors of the Turks speaking a Chuvash-type language from those speaking a common Turkic language, was a long procedure.

With the help of the lexical stock of the reconstructed Proto-Turkic language some conclusions can be drawn concerning the cultural level of the people that spoke Turkic prior to the split of the later Turkic languages. They existed in the Iron Age, were familiar with agriculture and, to some extent, horticulture. Animal husbandry was important to them and they lived in tents. They were organised in clans and tribes, and had shamans.

Proto-Turkic and the Genetic Question

Languages are genetically related if and only if their basic linguistic features can be considered the result of continuous and regular changes of the features of a common proto-language.

The Turkic–Mongolic Relationship

Scholars who believe there existed a common Altaic proto-language and those who deny its existence accept that there was close contact between the early forms of Turkic and Mongolic and that borrowings in both directions had taken place since early times. It is generally agreed that, after the rise of the Chingisid Empire in the early thirteenth century, different Turkic languages came under strong Mongolian influence. A less strong influence on Turkic languages spoken in west Asia can be observed from the end of the tenth century onwards, mainly due to the migration to the west of the Qara Khitai, who spoke an Old Mongolian type of language and later became Turkicised. From the fifteenth century onwards, Mongols became missionaries of the type of Buddhism that developed in Tibet and Mongolia. This impact can be traced in the Turkic languages of Siberia.

As for contacts prior to the tenth century, there was, as has been argued, very close contact between several Turkic- and Mongolic-speaking groups from the middle of the first millennium BC until the end of the ninth century AD. Many structural traits of Mongolic developed under Turkic influence or came into being through convergent changes and mutual impact of the two language groups. The impact of the Turkic branch of which only Chuvash has survived was of essential importance. Many changes typical of Chuvash began in the yet undifferentiated Turkic linguistic territory. Because of the migration of one section of the people, changes ceased in the ancestor dialects of the later Common Turkic languages and were accomplished and generalised in Chuvash only. If Middle Mongolian has *boyorla-* 'cut the throat', which comes from a Turkic **boyorla-* ← Old Turkic *boyaz* 'throat', the Turkic language from which it was copied was not necessarily of the Chuvash type, since the rhotacistic change $z > r$ before a consonant may have taken place in other Turkic languages as well. Mongolian *agurkai* 'mine, pit' is related to Proto-Turkic *agiz* 'mouth, opening', but may have been borrowed from any branch of Turkic. However, it is certain that the donor language was of the Chuvash type if Mongolian has *ikire* 'twin' but *koyar* 'two', whereas Common Turkic has *ikiz* 'twin' < Proto-Turkic *iki* 'two'. The Old Chuvash

form has also been copied into Hungarian as *iker* 'twin'. Mongolian *kuyur* 'a type of lute', a wandering cultural word, has also been borrowed from a language of the Chuvash type; cf. Common Turkic *kopuz*, etc.

The fact that Mongolian has early layers of loanwords from several Turkic languages does not permit the conclusion that all correspondences between the two groups are due to borrowing. It is possible that they once had a common proto-language, but the material is very scanty, if early loans are put aside.

The Altaic hypothesis

In earlier studies, it was generally accepted that Turkic, Mongolic, Tungusic and Uralic constitute one language family. Some authors called this supposed family 'Altaic'. Though, in a second phase, this hypothesis was seriously doubted, scholars such as Ramstedt and Poppe later tried to demonstrate that at least Turkic, Mongolic and Tungusic belong to one and the same 'Altaic' family. Their arguments were increasingly accepted until Clauson challenged this 'Altaic theory' in 1956. This started a lively discussion on certain features in Chuvash and Mongolian which some scholars considered a common heritage from an Altaic proto-language. Other scholars claimed that these features had been borrowed into Mongolian from a language of the Chuvash type.

Ramstedt was the first scholar to try to demonstrate in a coherent way that Korean is also a member of the Altaic language family. His comparisons have been heavily criticised in more recent studies, though the idea of a genetic relationship has not been totally abandoned. Later on, Miller (1971) claimed Japanese to be one of the branches of the Altaic family. If such a relationship had existed, it must be even more remote than the relationship of any two of the Indo-European languages.

The Nostratic Hypothesis

According to the Nostratic hypothesis, advocated by Illič-Svityč and his school, the six language families Altaic, Uralic, Indo-European, Dravidian, Semito-Hamitic and Kartvelian go back to a common proto-language and are thus genetically related. Illič-Svityč observed that the most problematic family was Altaic, its members being more remote from each other than any two members of other families.

The hypothesis has been questioned because the semantic range of most comparisons in the suggested etymologies is too wide, because the differences in the phonetic shape of words are overbridged with more than one asterisked form, and because the Nostratic proto-language would have existed in a period when the sociolinguistic background of language formation and preservation was totally different from the one known to us. It is thus a methodological mistake to project linguistic laws back to the supposed Nostratic proto-language.

Proto-Turkic and Proto Indo-European

Turkic languages have always had various contacts with languages of the Indo-European family. The question is whether there also existed a common proto-language for Indo-European and Turkic. Ramstedt excluded the possibility of a genetic relationship between Indo-European and Altaic. The parallels include a small group of wandering cultural words that came into Turkic very early. In Mongolic and Tungusic, however, it is not possible to detect any words borrowed directly from Indo-European. A few words of ultimate Indo-European origin have been copied into Mongolian through Turkic mediation and from there into Tungusic. The Indo-European borrowings include Proto-Turkic words such as *absa(k)* 'asp' ← (*H*)*osp*^h, *alma* 'apple' ← *ablu* or *amlu* and *burč* 'pepper', cf. Sanskrit *marica*. The words in question have certainly been borrowed in different periods. Research into early Indo-European words in Ancient Turkic is interesting since these words can be used to reconstruct the early history of the Turkic-speaking peoples.

Proto-Turkic and Proto-Uralic

Some scholars have implied a common proto-language for Uralic and Altaic. If all elements due to later contacts are removed from the material it becomes obvious that there is no reason to suppose such a proto-language. The valid comparisons are loans belonging to different periods and geographical areas. Revising a list of 27 words considered by Németh to belong to the earliest layer of Uralic and Turkic parallels, it has been shown that certain comparisons must be abandoned, while many others can be determined as early loanwords. Very few words remain that cannot with certainty be determined as loans. Only six to eight words do not exclude the possibility of originating in a common proto-language. This is not enough to establish a Uralo-Turkic proto-language.

Proto-Turkic and Other Language Families

Proto-Turkic also exhibits some cultural words from remote languages. The East Old Turkic word *böz* 'linen cloth' is ultimately of Egyptian origin and came through Turkic mediation to China and even twice to the Mongols, namely as *büse* and *bös*. The word thus probably existed in some Turkic languages along the Silk Route before the first Türk Kaghante. The form of the earlier loan *büse* contains an interesting final vowel added to the original Turkic **böz* to comply with the morphophonological structure of Proto-Turkic.

There are important early loans from Tokharian, e.g. *öküz* 'ox', *tümen* 'ten thousand', and from several Iranian languages, e.g. *bor* 'wine', *baltu* 'axe'. Some of them are not genuine Iranian words, but ultimately of foreign origin. Turkic *balta* and Mongolian *aluka* < *haluka* < **paluka* can, together with Chuvash *purta*, be traced back to a Semitic word which occurs in Akkadian as *pilakka* (stem *p-l-k* 'cut'); cf. Greek *πέλεκυς* (*pelekus*) 'axe'. The word,

which denotes the most important tool and weapon of the Bronze Age, was brought to the Turks and Mongols through the mediation of different Iranian languages at different times.

The early connections between Turkic and Chinese are also very instructive. Though many of the Chinese etymologies of Turkic words suggested earlier, e.g. *altun* 'gold', must be refuted, there are early Chinese words that can help reconstruct Proto-Turkic. Imitations of one and the same Chinese word are found in Chuvash *män* 'great' < West Old Turkic *mon*, in the Kipchak and Oghuz element *man* in place names such as *Mangišlaq* < *Man kišlak* or *Man Kermen* 'Kiev, the great city', and in the East Old Turkic numeral *ban* 'ten thousand'. The Chuvash word has, however, been copied from a different Chinese dialect and at an earlier time than the forms found in East Old Turkic, Oghuz and Kipchak.

Words of identical origin that have been borrowed into Turkic more than once may help us differentiate early chronological layers and dialects of Turkic. At the same time, they reflect old contacts with non-Turkic peoples and thus help reconstruct the history of the community that spoke early forms of Turkic.

References and Further Reading

- Clauson, G. (1956) 'The case against the Altaic theory', *Central Asiatic Journal* 2: 181–87.
- Doerfer, G. (1975–6) 'Proto-Turkic: Reconstruction problems', *Türk Dili Araştırmaları Yıllığı Belleten*, 1–50.
- Erdal, M. (1991) *Old Turkic Word Formation*, Turcologica 7, Wiesbaden: Harrassowitz.
- Illič-Svityič, V. M. (1971–84) *Opyt sravnenija nostratičeskix jazykov, vvedenie, sravnitel'nyj slovar'* 1–3, Moscow: Nauka.
- Johanson, L. (1979) *Altürkisch als 'dissimilierende Sprache'*, Abhandlungen der Akademie der Wissenschaften und der Literatur, Mainz, Geistes- und sozialwissenschaftliche Klasse 3, Wiesbaden: Steiner.
- Ligeti, L. (1980) 'A propos du rhotacisme et du lambdaïsme', *Central Asiatic Journal* 24: 222–50.
- Miller, R. A. (1971) *Japanese and the Other Altaic Languages*, Chicago, London: The University of Chicago Press.
- Németh, J. (1942–7) 'Probleme der türkischen Urzeit', in L. Ligeti (ed.) *Analecta Orientalia memoriae Alexandri Csoma de Körös dicata*, Budapest, 57–102.
- Poppe, N. (1960) *Vergleichende Grammatik der altaischen Sprachen I. Vergleichende Lautlehre*, Wiesbaden: Harrassowitz.
- Ramstedt, G. J. (1946–7) 'The relation of the Altaic languages to other language groups', *Finnisch-Ugrische Forschungen* 53: 15–26.
- (1957) *Einführung in die altaische Sprachwissenschaft 1–2*. Mémoires de la Société Finno-Ougrienne 104, Helsinki.
- Róna-Tas, A. (1986) *Language and History. Contributions to Comparative Altaistics*, Studia Uralo-Altaica 25, Szeged.
- (1988) 'Turkic influence on the Uralic languages', in D. Sinor (ed.) *The Uralic Languages. Description, History and Foreign Influences*, Handbuch der Orientalistik, 8. Abteilung, Leiden: Brill, 742–80.

- (1990) 'Altaic and Indo-European. Marginal remarks on the book of Gamkrelidze and Ivanov', *Acta Orientalia Hungarica* 42: 391–404.
- Tekin, T. (1969) 'Zetacism and sigmatism in Proto-Turkic', *Acta Orientalia Hungarica* 22: 51–80.
- (1986) 'Zetacism and sigmatism: main pillars of the Altaic theory', *Central Asiatic Journal* 30: 141–60.

5 *The History of Turkic*

Lars Johanson

Introduction

The following is a short outline of the main paths of diachronic development of the Turkic languages during the 1,200 years of their history which we can follow. The earliest stage clearly documented is the language of the East Old Turkic inscriptions of the eighth century AD. Attempts to interpret earlier materials as Turkic, e.g. the Hsiung-nu elements found in Chinese sources, have not been successful.

Development

Unity and Divergence

The Turkic languages are clearly genetically interrelated, showing striking similarities in phonology, morphology, and syntax. In spite of considerable structural changes in many of them it seems justified to speak of a certain conservatism of the whole group. There is, for example, even an amazing resemblance in many basic features between the language of the East Old Turkic inscriptions and modern Turkish.

Leaving aside the difficult question of pre-Turkic stages and the Altaic language family, we may safely assume that the unity of Turkic was dissolved by an early split of what is often called 'Oghur' (or 'Bulghar') from the rest, mostly called 'Common Turkic'. Some Turcologists refer to Oghur as the western branch, and to Common Turkic as the eastern branch of Turkic. The Proto-Bulghar language, reportedly similar to the language of the Khazars, belonged to the Oghur type. Its only modern representative is Chuvash in the Volga region. Oghur differs from Common Turkic by regular phonetic representations such as *r* and *l* instead of *z* and *š* in certain words, e.g. Chuvash *šēr* 'hundred', *šul* 'year' vs. Turkish *yüz* 'hundred', *yaş* 'age'.

A second split is represented by Khalaj, spoken today in central Iran and going back to an old Turkic dialect called Arghu. It is closer to Common Turkic than Chuvash, but both are strongly deviant languages, exhibiting both highly specific archaic features and numerous innovations due to language contact.

Divergence due to dialect splitting and complicated contact processes has led to further differentiation of Common Turkic into the three primary

branches Oghuz, Kipchak and Uyghur. Corresponding to their gradual diffusion, these branches later underwent further divergence and differentiation into secondary branches.

From the inner Eurasian steppes, Turkic-speaking groups early penetrated other regions in different directions: Oghur and Kipchak groups went westwards to the Pontic steppes and beyond; other old Turkic groups went westwards into central Eurasia; the Uyghurs moved southwards to Eastern Turkestan, Oghuz groups migrated southwestwards, towards Iran, Anatolia, etc. Due to the unique mobility of such groups, certain varieties of Turkic had an amazing expansion. The breakthrough in their diffusion occurred from the thirteenth century onwards, when Turkestan and Tatarstan were extensively Turkicised. Thus, of the Iranian languages of Central Asia, practically only Tajik has survived.

Classification

The dynamic history of the Turkic-speaking peoples makes it difficult to set up a classification of modern Turkic languages that combines geographic and genetic criteria. The following rough scheme represents an attempt to combine the current area distribution with genetic and typological features. It primarily distinguishes six branches, some of heterogeneous origin:

- 1 A southwestern (SW) branch, Oghuz Turkic;
- 2 A northwestern (NW) branch, Kipchak Turkic;
- 3 A southeastern (SE) branch, Uyghur Turkic;
- 4 A northeastern (NE) branch, Siberian Turkic;
- 5 Chuvash, representing Oghur or Bulghar Turkic;
- 6 Khalaj, representing Arghu Turkic.

Subdivision

The six branches may be subdivided as follows:

Southwestern

SW^W, a West Oghuz group comprising Turkish, Gagauz and Azerbaijanian.

SW^E, an East Oghuz group comprising Turkmen and Khorasan Turkic.

SW^S, a South Oghuz group comprising dialects of Iran (Kashkay, Sonqorī, Aynallu etc.) and Afghanistan (Afshar).

Northwestern

NW^W, a West Kipchak group comprising Kumyk, Karachay, Balkar, Crimean Tatar and Karaim.

NW^N, a North Kipchak or Volga–Ural group comprising Tatar (Kazan Tatar, Mishar, West Siberian) and Bashkir. West Siberian Tatar (Tura, Baraba, Tomsk, Tūmen, Ishim, Irtysh, Tobol, Tara and other dialects) is partly of different origin.

NW^S, a South Kipchak or Aralo–Caspian group comprising Kazakh,

Karakalpak, Kipchak Uzbek and Noghay. In spite of strong genetic bonds with Altay Turkic of the NE^S group, modern Kirghiz may also be placed in the NW^S group, since recent changes have brought it close to Kazakh. Though the so-called Fu-yü language of Manchuria is considered a Kirghiz dialect, it is closely related to South Siberian Turkic, e.g. Yenisey Turkic.

Southeastern

SE^W, a western group comprising modern Uzbek and its various dialects, i.e. Oghuz Uzbek.

SE^E, an eastern group comprising modern Uyghur, Taranchi and different Eastern Turki dialects of Kashgar, Yarkand, Khotan, Kerya, Cherchen, Aksu, Kucha, Turfan, etc. Deviating languages belonging to the SE^E area are Yellow Uyghur, probably of NE^S origin, and Salar, historically developed from the SW branch.

Northeastern

NE^N, a North Siberian group comprising Yakut (Sakha) and Dolgan.

NE^S, a heterogeneous South Siberian group comprising three subgroups:

- 1 Sayan Turkic consisting of Tuvan (Soyot, Uryankhai) and Tofa (Karagas);
- 2 Yenisey Turkic comprising Khakas, Shor and related dialects (Sayay, Qača, Qizil);
- 3 Chulym Turkic comprising dialects such as Küerik;
- 4 Altay Turkic comprising Altay (Oyrot) and dialects such as Tuba, Qumanda, Qū, Teleut, Telengit.

Some Altay Turkic dialects are rather close to Kirghiz (NW^S). As noted above, Yellow Uyghur may be of NE^S origin, and Fu-yü displays similarities with Yenisey Turkic.

Traditional Classificatory Criteria

One traditional classificatory criterion is the treatment of the final consonant in the word for 'nine'. Its representation as *r* in Chuvash *tāḫār* separates Oghur from Common Turkic, e.g. Noghay *toyiz*.

The treatment of the intervocalic consonant in the word for 'foot' divides most NE languages from the rest, e.g. Tuvan *adaq*, Yakut *ataḫ*, Khakas *azaḫ* vs. SW, NW, SE *ayaq*. Note, however, Khalaj *hadaq*, Yellow Uyghur *azaq*, Fu-yü *aziḫ*. This criterion is less important than traditionally assumed.

The Oghuz languages are divided from the rest by the loss of suffix-initial *-G*, e.g. *qalan* instead of *qalyan* 'remained'. Devoicing of suffix-final *-G* separates SE from NW, e.g. Uyghur *tayliq* vs. Tatar *tawli*, Kirghiz *tōlū* 'mountainous'.

Language Contacts

The massive displacements of Turkic-speaking groups throughout their history have led to a considerable mixture and a certain levelling of Turkic varieties.

One marked contact area has been the Crimea with its SW + NW encounters of Crimean Ottoman + Crimean Tatar, Crimean Noghay, Crimean Karaim. Another area is the territory of today's Uzbekistan with its NW + SE contacts. The Amu Darya region has been a real melting pot with SE + NW + SW contacts of Uzbek + Kipchak Uzbek + Oghuz Uzbek, Khorasan Turkic and Turkmen. The heterogeneous South Siberian contact area has already been mentioned.

Languages spoken in the central area of the Turkic-speaking world have undergone a good deal of interactive levelling, whereas those spoken in the periphery, e.g. Turkish, have preserved many older features. Certain languages such as Chuvash, Yakut, Salar, Yellow Uyghur, Khalaj, Karaim and Fu-yü have for centuries developed in relative isolation from the branches or groups they once belonged to, thus preserving old features and acquiring new ones in their respective environments.

Part of the divergence of the Turkic languages is due to contacts with Iranian, Slavic, Mongolic, Uralic, etc. In some languages, massive foreign influence has caused considerable deviation from the normal Turkic type. Turkic and Iranian have interacted intensely for many centuries, notably in Central Asia, which has led to a profound Iranian impact on Uzbek and an even stronger Turkic impact on some Tajik dialects. The Persian influence on Turkic dialects of Iran and Afghanistan is still highly intense.

The influence has been of various types. Speakers of Turkic have copied lexical, phonetic, morphological and syntactic elements from foreign languages into their respective varieties. There have also been cases of substratum influence, when non-Turkic groups shifting to Turkic have imposed native elements on their new language.

Periodisation

It is impossible to find linguistically meaningful criteria for a periodisation of the development of the Turkic group as a whole. Written sources do not give direct information on the spoken languages, and very often even written records are missing. For example, many older Turkic languages spoken in the Pontic-Caspian steppes, early Bulghar, Khazar, or predecessors of West Kipchak such as Pecheneg and Kuman, are only known from titles and names of tribes, persons and places.

The periodisation of the Turkic literary idioms which have emerged in different cultural centres is rather controversial. The reason is that it mostly depends not only on linguistic criteria, but also on extralinguistic – political and cultural – ones. Sometimes the periodisation is even based on the writing systems applied. The following sketch can only give a rough approximation of which main periods can be discerned in the history of written Turkic.

The Older Period

A so-called 'Old Turkic' period assumed by some Turcologists embraces no less than six centuries, beginning in the eighth century and extending up to Mongol rule. This heterogeneous period includes texts in different scripts, from the East Turkic runiform script through Uyghur and Manichaean scripts to Arabic script, and comprises at least three stages:

- 1 East Old Turkic proper, documented from the eighth century onwards. The earliest known records are inscriptions on stone steles in present-day Mongolia's Orkhon valley, in honour of rulers of the Second Türk Kaghanate. The language mirrored here is also represented in other inscriptions and in some manuscripts. Though it exhibits some features that are later typical of Oghuz, it may well be taken to represent a Common Turkic that has not yet split into Oghuz, Kipchak and Uyghur. The Old Kirghiz inscriptions found in the Yenisey basin are linguistically similar.
- 2 Old Uyghur is first recorded in runiform script from the period of the Uyghur power over the Eastern Kaghanate. From the ninth century onwards, it is documented in manuscripts in a rich literature of a mostly religious (Buddhist and Manichaean) nature, developed in the Tarim Basin up to the thirteenth century and written in Uyghur, Manichaean, Brāhmī and other scripts. Though the texts represent several chronological stages, they reflect a language that basically is similar to East Old Turkic proper.
- 3 Karakhanid, the first Islamic Turkic literary language, written mainly in Arabic script, developed in the eleventh century under the Karakhanid dynasty in Eastern Turkistan. It is known from Yūsuf of Balasayun's poem 'Wisdom that brings good fortune', and Maḥmūd al-Kāšyarī's 'Compendium of the Turkic Dialects' (see pp. 29, 138). The language is close to Old Uyghur, though its vocabulary is influenced by Arabic and Persian.

The Middle Period

The middle period, from the thirteenth century onwards, comprises various regional written languages. The Oghur or Bulghar branch is represented by Volga Bulghar, preserved in inscriptions on tombstones of the thirteenth and fourteenth centuries. The written Common Turkic varieties can be broadly determined as Uyghur, Kipchak and Oghuz, or as mixtures of elements from these branches.

There is much controversy regarding this period. For example, the term 'Middle Turkic' is often used for the Islamic Turkic languages of the eleventh to fourteenth centuries, i.e. also for Karakhanid.

East Middle Turkic written languages comprise:

- 1 Khorezmian Turkic, used in the thirteenth and fourteenth centuries in the Golden Horde, a preliminary stage of Chaghatay. It is based on the written languages of the older period, though interspersed with more specific Oghuz and Kipchak elements.
- 2 Early Chaghatay Turkic, used in the fifteenth to sixteenth centuries in the Timurid realm. It is based on Karakhanid and Khorezmian traditions, but contains more local elements of the Chaghatay area.

West Middle Turkic written languages comprise:

- 1 Kipchak Turkic, documented from the fourteenth century onwards in the *Codex Cumanicus*, compiled by Christian missionaries, and in dictionaries and grammars written in Egypt and Syria under the Mamlūk dynasty. It remained in use in the khanates succeeding the Golden Horde. Many sources that are nominally Kipchak are of a linguistically heterogeneous nature.
- 2 Oghuz Turkic, represented by Old Anatolian Turkish from the thirteenth century onwards, was initially influenced by eastern traditions. Its Old Ottoman variety later developed into Ottoman Turkish. A specific Azerbaijani literary language different from Ottoman Turkish began to develop from the fifteenth century onwards. Turkmen, written from the fourteenth century onwards, was later on subject to strong Chaghatay influence. There are also texts that seem to reflect Khorasan Turkic and South Oghuz.

The Pre-modern Period

From the sixteenth century onwards, a pre-modern period may be postulated as a preparatory stage of the formation of the modern languages. More distinct literary languages, more strongly influenced by regional spoken varieties, now develop in various areas.

- 1 Chaghatay continues, as Middle and Late Chaghatay, to play a major literary role, adopting local elements, e.g. in Eastern Turkistan, in the Khanate of Kazan and also among the Turkmens. It was later on ousted by more local varieties referred to as 'Turki', in Eastern Turkestan, the Volga region and the Crimea.
- 2 Ottoman Turkish develops, as Middle and Late Ottoman, into the leading Turkic written language with an abundantly rich literature comprising a variety of forms and styles.
- 3 Literary Azerbaijani reaches a high level of development, beginning with the Middle Azerbaijani period from the sixteenth century onwards.
- 4 Kipchak Turkic vanishes as a major literary language. Small languages such as Karaim and Armeno-Kipchak develop written forms.

The Modern Period

The modern period is represented by the following 24 written languages:

- 1 Southwestern: Turkish, Gagauz, Azerbaijani, Turkmen;
- 2 Northwestern: Kazakh, Karakalpak, Noghay, Kirghiz, Tatar, Bashkir, Crimean Tatar, Kumyk, Karachay, Balkar, Karaim;
- 3 Southeastern: Uzbek, Uyghur;
- 4 Northeastern: Yakut, Altay, Khakas, Tuvan, Shor, Tofa;
- 5 Chuvash.

Validity and Variation

The major literary languages emerging after the earlier period were trans-regional. For example, Chaghatay had, until a century ago, enormous prestige, serving as a written medium over all Central Asia, among West and East Turkestanian Turks, Kazakhs, Tatars, etc. Many works in Chaghatay were read and even written outside its basic area of distribution. Even literary products with relatively strong regional admixtures were widely understood.

Early Turkic literary languages exhibit considerable variation, variable norms that allow a choice of linguistic elements according to individual, situational and stylistic needs. Texts are often mixed in the sense of carrying features of more than one branch. Thus, many works by old Anatolian and Mamlūk authors are of a hybrid nature, showing both Kipchak and Oghuz characteristics. Many texts are intended to pursue Chaghatay norms, but still contain Oghuz or Kipchak additions from the author's and/or the copyist's own linguistic environment. Varieties of Chaghatay often carry features typical of regional languages. Thus, the old 'Volga Turki' texts show clear-cut Kipchak features.

In spite of regional specifics, the areas of validity of the written varieties never coincide with those of the branches of the spoken varieties. Their use does not imply linguistic uniformity in the regions covered. Very often they lack a direct relation to a specific spoken language. Thus, neither Volga Turki nor the eighteenth-century 'törki töl', a kind of Chaghatay with Tatar characteristics, mirror the spoken varieties of the region in a direct way.

Further Differentiation

The formation of regional written languages began in the second half of the nineteenth century. Languages intended to replace Chaghatay eventually emerged, much as a result of increasing Russian influence. Tatar, Uzbek, Kazakh and Turkmen intellectuals began to create literary languages for their own 'nations'. Since these languages used regionally limited words and forms, their radius of intelligibility was more limited. Thus, the Tatars developed a written norm that was rather close to spoken Kazan Tatar. Nevertheless, languages such as Ottoman, Azerbaijani, Uzbek and Tatar still had a considerable transregional validity up to the twentieth century.

At the same time, Tatar and Turkestanian intellectuals began to discuss a common Turkic written language intended to replace the different regional ones. For example, the journal *Terjiman*, published in the Crimea from 1883 on and intended to address all Turkic-speaking groups of Russia, was deliberately written in a language based on Ottoman but with a widely intelligible vocabulary.

Not only were these efforts resultless, but the differentiation went much further in the twentieth century. Many new written languages emerged. In Turkey, modern Turkish replaced Ottoman. In the Soviet Union, a set of new limited 'national' languages were established. The older established literary languages were restricted to their national territories. Some languages whose foundations as written media had been laid more recently, e.g. Yakut, Chuvash, Kazakh and Kumyk, were developed further. The new languages, created in the 1920s and 1930s, included Kirghiz, Bashkir, Karakalpak, Karachay-Balkar, Noghay, Tuvan, Altay (originally called Oyrot), Khakas and Shor. A 'New Uyghur' language was proclaimed the written language of Eastern Turkestan.

Lack of coordination in the construction of the regional languages and the introduction of different script systems restricted their mutual intelligibility. Language reforms had similar effects. Ottoman, which had almost lost its Turkic character in favour of innumerable Arabic-Persian elements, gave way to an essentially new Turkish literary language. It was not easily understood outside Turkey, since the basis of its vocabulary did not only consist of genuinely Turkic words, but was to a certain extent artificial. The languages of the Soviet Union also underwent changes, copying numerous Russian elements. Thus new differences emerged, and many real similarities became less visible. The political division of the Turkic world during most of the twentieth century made contact between Turkic-speaking groups difficult. In default of common language planning and close contact situations, the Turkic languages continued to grow apart.

The social importance of many Turkic languages was reduced. In some areas, e.g. in Iran, they were seriously hindered in their development. After political barriers have largely been overcome, the significance of the Turkic-speaking groups is, however, rapidly increasing.

The last century has seen many ideologically motivated attempts to underrate or overrate the similarities between Turkic varieties and endless disputes concerning the question which of them were 'languages' and which ones 'dialects'. As for the mutual intelligibility of Turkic languages, it is rather limited within the family as a whole, but may of course be rather high between neighbouring groups.

Diachronic Phonology

The following is a brief account of the main sound changes in the Turkic languages known to us. Reference will also often be made to a hypothetical

Proto-Turkic stage of development which may be reconstructed on the basis of data from known languages, notably Common Turkic and Oghur as documented in East Old Turkic and Chuvash, respectively.

Developments in Primary Stems

The developments in primary stems differ considerably from those in suffixes. Non-first syllables of primary stems which go back to old suffixes may therefore also display special features.

Vowels and consonants in primary stems have not developed individually, but as part of phonological systems. One distinction that concerns whole systems is the front vs. back distinction. For a realistic judgement upon its earlier stages of development, the graphic distinction of frontness and backness in older texts is crucial. The basically syllabic East Old Turkic runiform script system is very clear in this respect. It has a number of sign pairs, in which one letter signals a front consonant and the other a back one, both with the possible inclusion of an adjacent corresponding front or back vowel, mostly *e* and *a*. The functions of the different -, <D>-, <T>-, <G>-, <K>-, <L>-, <Y>-, <N>- and <R>-signs are obviously not to mark two kinds of *b*, *y*, etc., but to distinguish front and back syllables, regardless of the distribution of front vs. back articulation over individual segments. Traditional statements on segments used in 'back vocalic words' and 'front vocalic words' are misleading, since the intersyllabic front vs. back harmony is far from fully developed in East Old Turkic. There are obviously words in which not all syllables are harmonised (see pp. 108–9).

The phonetic realisations of the front vs. back distinctions vary across languages. Most Central Asiatic languages have maintained a velar vs. deep velar contrast in the dorsal obstruents *k* vs. *q*, whereas others, e.g. Turkish, have shifted to palatal vs. velar contrasts. Some languages show tendencies towards general fronting, often under Iranian or Slavic influence. Fronting of vowels is especially favoured in varieties such as Taranchi and Khalaj. Palatalisation of front consonants is observed in Gagauz, Karaim, Azerbaijanian dialects, etc. Here, the distinction palatalised vs. non-palatalised can signal the front vs. back quality of a syllable.

Vowels in Primary Stems

It is often assumed that at least the vowels *i*, *ü*, *i*, *u*, *e*, *ö*, *a*, *o* occurred in Proto-Turkic primary stems, just as in Turkish and many other modern languages. This assumption needs some additions and reservations.

First, it has traditionally been supposed that there also was a short phoneme *é*, a so-called 'closed *é*', an unrounded mid-high front vowel articulated between *e* and *i*. Most modern languages have *é* sounds as variants of *e*, though not as phonemes. It is still unclear whether the reflexes of a second *e* sound in later languages really go back to a Proto-Turkic *é*.

Secondly, though it is often taken for granted that an opposition *i* vs. *i̇* existed and that *i̇* was lost later in certain languages, this is far from certain.

Some scholars argue that *i* vs. *ī* coincided when Proto-Turkic *k* split into *q* and *k* (see p. 97). In default of clear graphic distinctions, we cannot decide whether East Old Turkic, Old Uyghur and Karakhanid had corresponding phonemes or just subphonemic variants. Some Turcologists suppose an *i* phoneme that was neutral, like the Mongolian *i*, with respect to the front vs. back distinction.

Thirdly, in addition to the short vowels, Proto-Turkic had a set of long vowels phonemes, *ā*, *ū*, *ū̄*, *ō*, etc. This is shown by corresponding length distinctions in Turkmen, Yakut, Khalaj and sporadic reflexes in other languages, e.g. Turkmen *būz*, *mūs* 'ice', Khalaj *bu^uz* 'cold'. The long vowels may at least partly have been diphthongs, like many of their modern reflexes, e.g. Yakut *on* 'ten', *k^uöl* 'lake', Khalaj *tā^r* 'narrow'. According to Doerfer, there was even a third quantity degree of half-long vowels, mirrored by plene writing in Maḥmūd al-Kāšyārī's treatise on older Turkic dialects (see p. 138) and preserved in Khalaj (see p. 281). Though Khalaj may well reflect more than one old quantity distinction, it is not yet clear whether it itself really exhibits three quantitatively distinguished phonemes of each vowel type.

It is even doubtful whether Proto-Turkic had as many as four *e*-sounds, *e*, *ē*, *ē̄* and *ē̄̄*. It might rather only have had *e* and *ē̄*, the latter possibly of diphthongic nature, *i^e*, and corresponding to the 'closed *e*' found at later stages. It would then be reflected in words such as Yakut *ki^eŋ*, Khalaj *kī^eŋg*, Turkmen *gīŋ*, 'wide', Yakut *ki^ehe* 'evening', Khalaj *kī^eče*, Turkmen *gīje* 'night', Yakut *bīl*, Khalaj *bī^el*, Turkmen *bīl* 'waist'.

The old long vowels were later on shortened in most Turkic languages. The shortening was obviously a protracted process occurring at different rates in different languages and positions.

Proto-Turkic long stem-final vowels have been preserved in later languages, e.g. *ara-* 'search', *bāšla-* 'lead', *aŋla-* 'understand', *teke* 'goat', *qara* 'black'. Short stem-final vowels, however, were reduced and lost, e.g. *> *erē* *> er* 'man', **atā-* *> at-* 'throw', **sagā-* *> say-* 'milk'. Their loss is attested by Mongolic parallels such as *saya-* 'milk', *ere* 'man'. They have also left traces in the aorist vowels *a*, *e*, *i*, *ī* of conservative languages such as Turkish, e.g. *sağar* 'milks', *atar* 'throws', *alır* 'takes', *sever* 'loves', *gelir* 'comes' (see p. 116). The gradual process of the loss of short stem-final vowels can be observed in East Old Turkic texts. The loss of final vowels resulted in final clusters such as *-nd*, *-rk*, etc., e.g. *and* 'oath', cf. Mongolian *anda* 'sworn brother'.

However, the final reduced vowels found in certain Chuvash words, e.g. *külē* 'lake', *uyā* 'arrow', *uđā* 'straw', *yiđā* 'dog' (cf. Turkish *göl*, *ok*, *ot*, *it*) or in words of the Eastern Turki Khotan dialect, e.g. *otā* 'fire', *bešī* 'five', *butū* 'thigh', *üčū* 'three', *biri* 'one' (cf. Turkish *od*, *beş*, *buđ*, *üç*, *bir*) are mostly considered elements added secondarily.

The old short vowels were reduced before consonants of non-first syllables in primary stems, e.g. *oy^l* 'son', *ay^z* 'mouth', *qarⁿ* 'stomach'.

They were lost in front of suffix-initial vowels ('middle syllable loss'), e.g. *oyli* 'his/her son', *ayzi* 'his/her mouth', *qarni* 'his/her stomach'. The syncopation was systematic only in syllables with the vocalic entity written here as *°*, probably a *schwa*-like centralised vowel. Doerfer has suggested that an <a> found in non-first syllables in certain Old Uyghur texts may represent this sound.

Another question is whether Proto-Turkic possessed initial diphthongs such as *'a*, *'e*, *'o*, *'ö*. Many Common Turkic words with initial vowel have Chuvash counterparts with a *y*-preceding the vowel, e.g. *yal* 'village', *yuχ-* 'flow', *yat* 'name', *yiđā* 'dog', cf. Azerbaijani *el* 'country', *aχ-*, *ad*, *it*. Diphthongic pronunciation of initial *e* is typical of the NW^S group, e.g. Kazakh *'el* 'people', Noghay *'er* 'man', and is also found in some other languages such as Karachay-Balkar. A parallel case is offered by the initial *v-* in Chuvash words such as *vut* 'fire', *vil-* 'die', *vāđār* 'thirty'; cf. Turkish *ot*, *öl-*, *otuz*. A similar diphthongic pronunciation of *o-*, *ö-* is regular in NW^S and Karachay-Balkar, e.g. Kazakh *'otiz* 'thirty', *'oraq* 'sickle', *'öt-* 'pass', Karakalpak *'oq* 'arrow', Noghay *'orta* 'middle', *'öl-* 'die', Karachay-Balkar *'ot* 'fire', *'on* 'ten'. However, it is impossible to determine all these cases as remains of initial diphthongs, particularly in view of similar phenomena in Slavic languages.

Another problem is whether vowel alternations in pronominal stems such as *bēn* 'I' → *baŋa* 'to me' are traces of an old vowel gradation or umlauting.

Later Tendencies

Vowel Lengthening

Secondary vowel lengthening has arisen through contractions with loss of consonants, mostly voiced fricatives, typically *VG(V) > V̄*, e.g. Khakas *azām* 'my foot' < **azayim*, *tōp* < *toŋip* 'frozen', Tuvan *ār* 'heavy' < **ayir*, *bōr* 'becomes' < *bōlur*, Kirghiz *tō* 'mountain' < **tay*, *zōt* 'factory' ⇐ Russian *zavod*, *qištō* 'winter quarters' < **qišlay*. There are often long vowels in words of foreign origin, e.g. Turkish *hāl* 'situation, state' ⇐ Arabic.

Foreign Influence on Vowels

Iranian phonetic influence has affected the front vs. back system of vowels in Standard Uzbek, which is based on the urban dialects of Tashkent and Ferghana. Though *i*, *o*, *u* and *a* are often claimed to be the only Uzbek vowel phonemes, the differences between their front and back realisations are mostly clear-cut. Thus, the type *a* has its front realisations, transcribed here as *ā*, and its back realisations, transcribed as *a* and *ā̄*, the latter written <o> in the current orthography.

Rounding of *a*, observed in Uzbek, Khalaj, Tatar, Bashkir, Chuvash, etc. is probably secondary. It is typical of Uzbek, e.g. *bāš* 'head', *āt* 'horse', *āt-* 'throw', *ārqa* 'back', *barmāq* 'finger', where it often corresponds to an *ā* of

old Turkic or foreign origin and is most frequent in first and closed non-first syllables. Though it represents an areal phonetic type developed under Iranian influence, its distribution is different from that of Tajik *â*. Khalaj displays very similar rounding phenomena under Persian influence, e.g. *bâš* 'head', *hât* 'horse'. Unknown substrata may have led to similar rounding in the Volga region. Tatar and Bashkir *â* is mainly observed in first syllables, e.g. *ât* 'horse', *bâlâ* 'child'. A recent further development in Chuvash has produced the high rounded vowel *u*, e.g. *ut* 'horse', *ura* 'foot', *uldâ* 'six'.

Variation *a* ~ *ï*

A set of primary stems in which most modern languages exhibit *a* have equivalents with *ï* in Tuvan, e.g. *tïp-* 'find', *tïrt-* 'pull', *sïp* 'handle', *çït-* 'lie', *çïrïn* 'stomach', Yakut, e.g. *iy* 'moon', *sit-* 'lie', *tïmir* 'vein', *iy-* 'say', and Chuvash, e.g. *sïr-* 'write', *pil* 'honey', *ildân* 'gold'. It is not clear whether *a* has become *ï* in certain cases, or if some other vowel has developed to *ï* and *a* under different conditions. The *a*-sounds that remained unrounded may have changed into *ï* in these cases (see p. 70). Doerfer assumes a back unrounded *ê* between *a* and *ï*, a back counterpart of the 'closed *ê*'. It normally turned into *a*, but became *ï* in Chuvash, Yakut and (word-internally) Tuvan. Word-finally it developed into *-ï*, e.g. *altï* 'six', but became *-a* in Yakut and Khalaj and *-ä* in Chuvash, e.g. Yakut *alta*, Khalaj *altâ* 'six', Chuvash *uldâ*.

Systematic Vowel Shifts

One interesting phenomenon is the systematic raising of low vowels, notably *e* > *i*. Tatar and Bashkir display a shift that began with *e* > *i*, e.g. Tatar *kil-* 'come', and spread to *o* > *u* and *ö* > *ü*. There is a corresponding centralisation of high vowels. The NW^S subgroup, except Kirghiz, tends towards the shifts *i* > *ï*, *ï* > *ĩ*, *u* > *u* and *ü* > *ü*. In Tatar and Bashkir, high vowels have been centralised and further shortened: *i* > *ẽ*, *ï* > *ã*, *u* > *õ*, *ü* > *õ*. The tendency goes further in the reduced Chuvash vowels *ä* and *ẽ*, e.g. *tär-* 'stand', *për* 'one'.

Both tendencies fit into general patterns of the Volga–Kama area with phonetic parallels in neighbouring Fennic languages. Independently of the controversial question of their origin, their presence in the Turkic languages of the region is clearly motivated by internal phonological factors. The centralising and shortening of high vowels ensured the distinctions endangered by the raising of low vowels.

Vacillations of the high vs. low distinction in rounded vowels occur in several other Turkic languages. Reduction and loss of high vowels, e.g. *b'r*, *b'p* 'one', is frequent in many languages of Central Asia. Retraction of rounded vowels, *ö* > *o* and *ü* > *u*, is often observed in West Kipchak.

Assimilatory Rounding

A widespread kind of rounding is the intersyllabic labial harmony (see p. 33), in which the assimilating factor is the vowel of the preceding syllable, e.g.

Kirghiz, Altay Turkic, Yakut *orto* 'middle' < **orta*.

In Uyghur, old *a* and *e* sounds in open first syllables have been rounded to *o* and *ö* due to an *u* or *ü* in the second syllable, e.g. *yoruq* 'light' < **yaruq*, *oçuq* 'open' < **açuq*, *ötük* 'boot' < **etük*, *tömür* 'iron' < **temür*. Similar cases are noted in Yenisey Turkic dialects. Shifts *a* > *o* due to an *u* of the following syllable have also taken place in Yakut, e.g. *çotun* 'lady' < **çatun*.

Vowels are often rounded due to a neighbouring labial consonant, *p*, *b*, *m*, *v* or *f*, e.g. Turkmen *θöy-* 'love', *öy* 'house', Azerbaijanian *ovçu* 'hunter', *dovşan* 'rabbit'; cf. Turkish *sev-*, *ev*, *avcı*, *tavşan*.

Assimilatory Fronting and Backing

A widespread kind of fronting is the intersyllabic front vs. back harmony, in which the assimilating factor is the quality of the preceding syllable (see pp. 32–3). Vowels may also be fronted due to a neighbouring consonant, mostly *y*, e.g. Tatar *jeş* 'young' < **yāš*, *jey* 'summer' < **yāy*, Azerbaijanian *yər* 'place' < *yer*, also with loss of *y-*, e.g. *il* 'year', *ilan* 'snake'; cf. Turkish *yer*, *yil*, *yılan*. There are also variants such as Bashkir *ađ* ~ *eđ* 'little'. Fronting of vowels in Chuvash has often occurred in the neighbourhood of *š*, *ç* and *l*. Kumyk displays *a* > *á* before a syllable with an original *l*, e.g. *árıw* 'beautiful', *sári* 'yellow'. Cases of fronting umlauting are also found in Yenisey Turkic, e.g. *palezi* 'his/her child' < **palazi*.

Backing of vowels is a much less frequent phenomenon, except as a result of the intersyllabic front vs. back harmony. Chuvash has a regular shift **e* > *a* and some other cases of vowel backing, often in the neighbourhood of *š*, *y* and *k*.

Vowel Raising and Uyghur Umlauting

Modern Uyghur displays several cases of regressive vowel assimilation, perhaps due to Iranian substratum influence. First, *a* and *e* are raised to *i* in open syllables, e.g. *balılar* 'children' ← *bala* 'child'. Secondly, there is the phenomenon called Uyghur umlauting, which also seems to have begun as raising. In Eastern Turki it occurs before both *i* and *ï*, producing a vowel which, according to Jarring (1933), is phonologically back, e.g. *æti* 'its horse' ← *at* 'horse'. The additional fronting typical of Taranchi dialects (pp. 83, 395) was taken over in the standard Uyghur language. The change *a* > *é* is thus secondary. Both *a* and *e* developed into a mid-level *é* in unaccented open syllables before a high unrounded vowel of the following syllable, e.g. **baliq* > *béliq* 'fish'.

Development of Individual Vowels

High Unrounded Vowels

Long *ï* is preserved in Yakut, Turkmen, Khalaj *qï'z*, e.g. Yakut *kïs*, Turkmen *gïđ*. Short and shortened *ï* is mostly preserved. It has been centralised in Kazakh, Karakalpak and Noghay, e.g. Kazakh *qïš* 'winter', centralised and further shortened in Tatar and Bashkir, e.g. *qış* 'winter', and centralised and

reduced in Chuvash, e.g. *χěl* 'winter' (beside other representations).

Long *ī* is preserved in the languages just mentioned, e.g. Yakut *tīs*, Turkmen *dīs*, Khalaj *tīz* 'tooth'. Short and shortened *i* is mostly preserved. It has been lowered in Kazakh, Karakalpak, Noghay and Khakas, e.g. Kazakh *bil-*, Khakas *pil-* 'know', centralised and further shortened in Tatar and Bashkir, e.g. Tatar *běl-* 'know', centralised and reduced in Chuvash, e.g. *pēr* 'one' (beside other representations).

The contrast *i* vs. *ī* is weak in certain languages, a phenomenon often ascribed to Iranian influence. Though Uyghur and Khalaj are claimed to lack *ī* as a phoneme, they clearly display back vs. front realisations of high unrounded vowels. The history of Turkic has also seen several cases of class shifts in the sense that certain primary stems with *ī* have shifted to *i*, e.g. Turkish *iş* 'work' < *īš*, *iz* 'trace' < *īz*, *dil* 'tongue' < *tīl*.

High Rounded Vowels

Long *ū* is preserved like other long vowels, e.g. Yakut *tūs*, Turkmen *dūd*, Khalaj *tūz* 'salt', Yakut *būt*, Turkmen *būt*, Khalaj *būt* 'thigh'. Short and shortened *u* is mostly preserved. It has been lowered in Kazakh, Karakalpak and Noghay, e.g. Kazakh *uč-* 'fly', *tut-* 'hold', *tur-* 'stand', centralised and further shortened in Tatar and Bashkir, e.g. *öč-* 'fly', *töt-* 'hold', *tör-* 'stand', and centralised and reduced in Chuvash, e.g. *tär-* 'stand' (beside other representations).

Long *ū* is preserved like other long vowels, e.g. Yakut *ūt*, Turkmen *θūyt* 'milk', Yakut *kūs*, Turkmen *gūyč* 'strength'. It is mostly unrounded in Khalaj, e.g. *sīt*, *kīč*. Short and shortened *i* is mostly preserved. It has been lowered in Kazakh, Karakalpak and Noghay, e.g. Kazakh *tūs-* 'fall', centralised and further shortened in Tatar and Bashkir, e.g. Tatar *töš-* 'fall', centralised and reduced in Chuvash, e.g. *kěl* 'ash' (beside other representations). It has been slightly retracted in Crimean Tatar and Trakai Karaim, e.g. *yüz* 'hundred', and unrounded in Halič Karaim, e.g. *ic* 'three', *yiz* 'hundred'.

Low Unrounded Vowels

Long *ā* has preserved its length like other long vowels, e.g. Turkmen *yāz* 'summer', *dāš* 'stone', *gān* 'blood', *gār* 'snow', Yakut *sās*, *tās*, *χān*, *χār*, Khalaj *yāz*, *tāš*, *qān*, *qār*. Short and shortened *a* is mostly preserved. It corresponds to *ī* in certain Tuvan, Yakut and Chuvash words, e.g. Tuvan *čit-* 'lie', Yakut *sīt-* 'lie', Chuvash *pil* 'honey'. It is often rounded in Uzbek, Khalaj, Bashkir, Chuvash, e.g. Uzbek, Tatar, Bashkir *ât*, Khalaj *hât*, Chuvash *ut* 'horse'; with originally long *ā*: Uzbek *sâl* 'raft', Chuvash *χur* 'goose', *yul-* 'remain'.

The old long *ē*, probably a diphthong *i^e*, is reflected in Yakut *kī^eη*, Khalaj *kī^eηg*, Turkmen *gī^eη* 'wide', Yakut *kī^ehe* 'evening', Khalaj *kī^eče*, Turkmen *gī^eje*, Azerbaijanian *gēje* 'night', Khalaj *bī^el*, Turkmen *bīl*, Yakut *bīl*, Azerbaijanian *běl* 'waist'. In most languages it has coincided with the short *e*, e.g. Turkish *geniş*, *gece*, *bel*. Short *e* is normally preserved, mostly as a mid-

level [e], e.g. Turkmen *kes-* 'cut'. It may also be lowered to [æ] or raised to [e], e.g. Azerbaijanian *kæs-* 'cut' (here transcribed as *kes-*), *yēr* 'place'. It has been backed to *ā* in Khalaj under Iranian influence and to *a* in Chuvash (after the old *a* had been rounded), e.g. *kas-* 'cut', *tar* 'sweat', *sas* 'voice'. It has been raised in Tatar, Bashkir and Khakas, e.g. *it* 'meat', *kil-*, Khakas *čil* 'wind', *pil* 'waist'. If there ever was a Proto-Turkic distinction *e* vs. *ē*, it is not reflected in either Common Turkic or Chuvash.

Low Rounded Vowels

Long *ō* is preserved in Yakut, Turkmen and Khalaj, e.g. Turkmen *yōl*, Yakut *s^ool*, Khalaj *yū^ol* 'way'. Short and shortened *o* is mostly preserved. It is raised in Tatar, Bashkir and partly Chuvash, e.g. Tatar *ut* 'grass', *tuq* 'full', Chuvash *uđā* 'hay'.

Long *ō* is preserved in Yakut, Turkmen and Khalaj, e.g. Yakut *r^oöt*, Turkmen *dört*, Khalaj *tū^ort* 'four'. It is often unrounded in Khalaj, e.g. *kī^ok* 'green'. In Chuvash it is represented in some words by *va-* and *-äva-*, e.g. *var* 'inside, centre', *tāvata* 'four' *kāvak* 'blue'. Short and shortened *ö* is mostly preserved. It is raised in Tatar, Bashkir and partly Chuvash, e.g. Tatar *küp* 'much', *kür-* 'see', Chuvash *pül-* 'divide', and retracted in West Kipchak, e.g. Crimean Tatar *köz* 'eye', *öz* 'self', Trakai Karaim *k'öz*. It is often unrounded in Khalaj, e.g. *kez* 'eye', and in Halič Karaim, e.g. *k'ez* 'eye'.

Consonants in Primary Stems

Proto-Turkic is assumed to have had the stops *p*, *t*, *k*, *b*, *d*, *g*, the affricate *č* (possibly also *ǰ*), the fricatives *s*, *z*, *š*, and the sonorants *r*, *l*, *m*, *n*, *ŋ*, *ñ*. As we will see, it is also possible that *ñ*, *z* and *š* have emerged from combinations with a palatal element, i.e. < *nⁱ*, *rⁱ* and *sⁱ*.

The stops may be assumed to have formed the following oppositions with respect to the features tense vs. lax:

Fortes (strong consonants) Lenes (weak consonants)

* <i>p</i>	* <i>b</i>
* <i>t</i>	* <i>d</i>
* <i>k</i>	* <i>g</i>

Initial Stops

The distribution of consonants in initial position seems to have been rather limited. Of the six hypothetical stops mentioned above, only *b-*, *t-* and *k-* are assumed to have occurred word-initially. These cases will be briefly discussed here, whereas word-initial *č-*, *s-* and *ñ-* will be dealt with on pp. 105–6.

An initial **p-* has sometimes been presumed because of the voiceless word-initial labials found in South Siberian and other dialects, namely *ḫ-* on a par with the other mediae lenes *q-*, *g-*. But East Old Turkic *b-* is clearly indicated by the runiform script, and parallels such as Mongolian *bura^u*, Hungarian *borjú* point to *b-* in old Turkic words such as *buzāyū* 'calf'.

According to the Altaic theory, Altaic **p*- has developed into Tungusic *p*-, *f*-, *h*-, and into Mongolic and Turkic *h*- > \emptyset . Today, not only Altaicists believe that **p*- developed very early over a bilabial fricative into an *h*. The latter is taken to have disappeared in almost all Turkic languages, whereas it is found in Khalaj, e.g. *hadaq* 'foot', *hat* 'horse', *hu't* 'fire', *häv* 'house', and has left traces in East Old Turkic and some modern languages.

As for *g*-, it is thought to have disappeared due to a neutralisation of a word-initial opposition *k*- vs. *g*-.

Initial *d*- is supposed to have disappeared in favour of *t*- as a result of corresponding neutralisation. The East Turkic runiform script cannot prove this, since its <T>-signs denote both *t* and *d*. Neither do Mongolian words such as *dalay* 'sea', *dörben* 'four', *durusun* 'birch bark' prove an original *d*-, since they might represent an early development *t*- > *d*-. Sayan Turkic exhibits, like Oghuz, both *t*- and *d*-. In spite of certain distributional similarities, e.g. Tuvan *dolu* 'full', *day* 'mountain', *dört* 'four', *düş*- 'fall', *tut*- 'seize, hold' vs. Turkish *dolu*, *dağ*, *dört*, *düş*-, *tut*-, this phenomenon is not likely to reflect an older opposition *t*- vs. *d*-.

An early **d*- may also have developed via δ - into Proto-Turkic *y*-, just as an intervocal *-d*- did at a later stage. Greek transcriptions of the second and sixth centuries exhibit initial < δ > when rendering the Turkic word known as *yōy* 'funeral feast' and the name of the Ural river, **Yayiq*. A Danubian Bulghar *dilom* might reflect an early **dilān* 'snake'; cf. Khalaj *gilān*, Chuvash *sēlen*, Turkmen *yilān*. Turkic *yayiz* 'brown' and *yayī* 'enemy' correspond to Mongolian *dayir* and *dayi(n)*, which either reflect an Altaic **d*- that developed into *y*- in Turkic, or have been copied from an early Turkic variety possessing *d*-. We shall return to this problem when discussing the relationship between Turkic *y*- and *-j*.

Initial *y*- It is generally assumed that Proto-Turkic had an initial **y*-, though there is little consensus on its origins. In modern languages, *y*- is generally found in SW and SE, but also in the NW languages Kumyk, Karaim and Bashkir. There is sometimes a development *y* > \emptyset , notably in Azerbaijanian, e.g. *il* 'year', *ulduz* 'star', *üz* 'face', *ürek* 'heart'. Some languages have *j*-, e.g. Kirghiz, Kipchak, Uzbek *jat*- 'lie', *jol* 'way', and some have *ž*- e.g. Kazakh *žat*-, *žol* 'way'.

The dialectal variation is often considerable. We find *y*- in southern and western Kazakh dialects, in northwestern Tatar dialects, West Siberian dialects, Qara-Noghay, *j*- in southwestern Tatar dialects, eastern Kazakh dialects, southwestern Karakalpak dialects close to Turkmen, Central Noghay, *ž*- in Karakalpak northwestern dialects close to Kazakh, Aq-Noghay and sporadically in Uyghur, e.g. *žiraq* 'far', *žitiq* 'torn'. Some Tatar dialects have *y*- before back vowels and *j*- before front vowels, e.g. *yat*- 'lie', *jil* 'wind'. Kumyk has *j*- in a few words, e.g. *jiw*- 'collect', *juw*- 'wash'.

Most South Siberian (NE^S) languages display *č*-, e.g. Khakas *čaxsi*, Tuvan *čaxši* 'good'. Older texts show *d*'- and *t*'-. Altay Turkic has a weak palatalised *d*'-, e.g. *d'at*- 'lie', *d'ol* 'way', *d'eti* 'seven', *d'üs* 'hundred', in northern dialects *t'at*- etc. Yakut has *s*-, e.g. *suol* 'way', *sūs* 'hundred'. Chuvash rather regularly displays the palatalised dental fricative *ś*-, e.g. *śul* 'way', *śēr* 'hundred', *śul* 'tear', *śil* 'wind', *śēlen* 'snake'. Note that this is also the way **č*- has developed in both Yakut and Chuvash.

The picture is additionally complicated by the fact that certain Turkic dialects, both in the northeast (Siberia, Manchuria) and in the southwest (Iran) display *n*'- or *n*- instead of expected *y*- in some words that contain a nasal; see pp. 105–6.

The prevalent opinion among Turcologists is that *y*- is primary in relation to the other variants, i.e. has developed into *j*, *ž*, *d*'-, *č*-, *s*-, etc. NE forms such as *d*' and *t*' are explained as *d*- and *t*-prothesis, dentalisation of *y*-, due to Samoyedic substratum, etc. The option *j*- > *y*- is seldom considered, though there are several indications of such shifts, e.g. in Noghay. The fact that *j* and *ž* may also occur in recent loanwords, e.g. Aq-Noghay *žadra* 'bullet' ← Russian *jadró*, does not prove that they have developed from *y*-.

One conceivable possibility is that **d*- has developed over *j*- into *y*-. Turkic *y*- may, as we have seen, correspond to Mongolic *d*-, but it sometimes also corresponds to Mongolic *j*-, e.g. *jil* 'year'; cf. Common Turkic *yil*. The initial consonant either reflects an Altaic **j*- that has coincided with *y*- in Turkic, or the words in question are copies of Turkic words with initial *j*-; cf. **jüz* 'hundred' in Volga Bulghar inscriptions. It is thus in principle possible that *j*, *ž*, *d*'-, *č*-, *s*-, etc. have not developed from *y*-. Words such as Altay Turkic *d'üs*, Yakut *sūs*, Chuvash *śēr* 'hundred' and *d'ildis*, Yakut *sulus*, Chuvash *śıldār* 'star' might rather go back to forms such as **düz* and **duldaz*. Note that there are even similar forms in various modern languages, e.g. dialectal Kazakh *düz* 'hundred', *düz*- 'swim', Balkar *duldaz* 'star'.

Non-initial Consonants

Consonants in non-initial position are assumed to have included the six stops mentioned above, the affricate *č*, (possibly also *j*), the fricatives *s*, *z*, *š*, and the sonorants *r*, *l*, *m*, *n*, *ŋ*, and *ń*. The lenes *b*, *d*, *g* are likely to have tended towards spirantisation. This tendency is obvious in East Old Turkic, where the variation of oral dentals is indicated by runiform <D>-signs denoting δ and <T>-signs denoting *d*. The presence of the weak affricate *j* is shown by the global sign <NJ>, which denotes the sequence *nj*.

Front and Back Dorsals It is mostly assumed that the stops **k* and **g* were realised as front or back according to their environment and that they later split into *k* vs. *q* and *g* vs. *ɣ*, respectively. The front lenis *g* is also likely to have had fricative realisations, as, for example, it also has in modern Tuvan and Khakas.

The runiform script clearly distinguishes front and back *k*- and *g*-sounds. Though the graphic systems based on the Uyghur, Manichaean and Arabic scripts are less differentiated, they usually distinguish front and back dorsals. Most modern Turkic alphabets use distinct signs for *k* vs. *q* and *g* vs. *γ*.

Secondary Weak Obstruents The later development of the old lenes is characterised by further weakening. However, a new generation of lenes developed secondarily in certain positions.

Oghuz languages display five such weak obstruents, *b*², *d*², *g*², *ğ*², *j*², developed from the fortes *p*, *t*, *k*, *q*, *č*. They basically occur in monosyllabic primary stems after originally long vowels and in non-first primary stem syllables, e.g. *ād*² 'name', e.g. *qanad*² 'wing'. In most other languages, the phenomenon is more limited or less clearly manifested.

In the tense vs. lax oppositions thus established, the old fortes remain voiceless, whereas the lenes may be more or less voiced or devoiced. In modern languages, the new lenes have partially followed the old ones and developed into fricatives or \emptyset , notably between vowels.

Secondary weak obstruents are, for instance, found in Azerbaijani *od*² 'fire', *bud*² 'thigh', *güj*² 'strength', *dib*² 'bottom', *ağ*² 'white', *dodağ*² 'lip', *gög*² 'heaven', *aj*² 'hungry', *uj*² 'tip, end'. They are observed most clearly if followed by a vowel, e.g. *od*²*u* 'its fire', *bud*²*u* 'its thigh', *güj*²*ü* 'its strength', *dib*²*i* 'its bottom', *ay*²*ar*- 'whiten', *doday*²*i* 'its lip' etc. The Turkish counterparts are written *adı*, *odu*, *budu*, *gücü*, *dibi*, *ağar*- etc. In syllable-final position, they are mostly devoiced, at least partially, e.g. *ađ*² 'name', *ođ*² 'fire', *bud*² 'thigh', *güj*² 'strength', *dib*² 'bottom', *ağ*² 'white', *gög*² 'heaven'. The Turkish counterparts are written *but*, *güç*, *dip*, *gök*, etc.

Examples of *g*², *ğ*² and *b*² in other languages are Uzbek *ēšig*² 'door' → *ēšig*²*i* 'its door', Kazakh *esig*² 'door' → *esig*²*i*, *etig*² 'boots' → *etig*²*i*, Uzbek *ortâğ*² 'friend' → *ortâğ*²*i*, Kazakh *tarâğ*² 'comb' → *taray*²*i*, Uyghur *ayâğ*² 'foot' → *ayiy*²*i*, Uyghur *qab*² 'sack' → *qab*²*i*.

Obstruents in lexemes copied from Arabic, Persian, Russian etc. may also appear as new lenes, e.g. Azerbaijani *mektub*², Uzbek *âzâd*² 'free', Uyghur *zawud*² 'factory', *mekteb*² 'school'.

The Sayan Turkic languages Tuvan and Tofa devoice their syllable-final obstruents, but may still distinguish fortes from lenes by means of a glottal element ʔ , e.g. Tuvan *aʔt* 'horse' vs. *at* 'name', *oʔt* 'grass' vs. *ot* 'fire'. The same element may mark intervocalic voiced obstruents as fortes, e.g. Tuvan *aʔdim* 'my horse' vs. *adim* 'my name'. Even if the glottal element is, as has been suggested, due to Samoyedic substratum, it has an internally motivated function as an anticipating fortis signal (p. 98). There are similar devices in Salar and Yellow Uyghur (p. 339).

Other Tendencies Voicing of intervocalic *-p-* > *-b-*, *-k-* > *-g-*, *-q-* > *-γ-* takes place in Kipchak languages, e.g. Bashkir *hiğēð* 'eight'; cf. Turkish

sekiz. The same shifts plus *-t-* > *-d-* occur in South Siberian, e.g. Khakas *ada* 'father'; cf. Turkish *ata*. The Sayan Turkic marking of intervocalic strong obstruents by means of glottalisation has just been mentioned. Chuvash displays at least partial voicing of consonants between vowels and between a sonorant and a vowel, i.e. *-b̥-*, *-d̥-*, *-g̥-*, *-γ̥-*, *-z̥-*, *-ž̥-*, written <p>, <t>, <k>, <x>, <s>, <ś>, e.g. *văđār* 'thirty', *ađā* 'boot', *suγān* 'onion', *užā* 'open'. In most other languages, intervocalic fortes are preserved as such, e.g. SW *ata* 'father'.

West Kipchak languages of the Caucasus region display strongly aspirated voiceless stops, e.g. Karachay *tʰapʰ-* 'find', and glottal stops, e.g. Balkar *atʔa* 'father', southern Kumyk dialects *qʔirq* 'forty'. Similar tendencies are found in Salar and Yellow Uyghur. Both South Siberian languages and Fu-yü display aspirated initial fortes *pʰ-*, *tʰ-*, *kʰ-* in loanwords, whereas the corresponding initial stops in native words are *b̥-*, *d̥-*, *g̥-*.

General fronting of the articulation basis, often under foreign influence, has affected the phonetic realisation of consonants in certain languages. Thus, in Istanbul Turkish, in the extreme southwest of the Turkic world, the articulation of *q*, *k*, *ğ*, *g* etc. has shifted to the effect that the front consonants are generally palatalised, probably a substratum phenomenon under the influence of Mediterranean languages. General fronting in Chuvash has led to oppositions of palatalised and non-palatalised consonants. Gagauz and Karaim have, under Slavic influence, developed palatalised variants of all consonants, e.g. Karaim *m'en* 'I'. Palatalisations are also observed in Azerbaijani dialects.

Development of Individual Consonants

q Initial *q-* is preserved in most languages, e.g. Turkish *kuş* 'bird'. Many languages display fricatives, e.g. Khakas *χar* 'snow', *χarax* 'eyeball', Chuvash *χura* 'black'. Tuvan has fricatives in some words, e.g. *χar* 'snow', but *qara* 'black'. Kirghiz and Altay do not display *χ-* in native words. Yakut has *k-* before front vowels and high back vowels, e.g. *kīs* 'girl', but *qʰ-*, written <x>, before *a* and *o*, e.g. *χāl-* 'remain'. Khalaj exhibits *χ-*, *qʰ-*, *qʰ-*, etc. The interrogative element **qa-* has developed into Oghuz *ha-*, e.g. Turkmen *χayθi*, Azerbaijani *hansı*, Turkish *hangi* 'which'. Voicing has occurred in eastern and southeastern Anatolian dialects, Azerbaijani, Turkmen, Kumyk dialects, sometimes Khalaj, etc., e.g. Azerbaijani *ğara* 'black', *ğuş* 'bird', Turkmen *ğīð* 'girl', *ğān* 'blood'. Though this development is commonly ascribed to Persian influence, it is not present in Uzbek. There is also loss of *h-* in some Turkish dialects, e.g. *ačan* 'when' and development into a glottal stop in Mishar Tatar, e.g. *ʔāl-* 'remain'. In certain Chuvash words before originally long vowels we do not find *χ-*, but *y-*, e.g. *yun* 'blood', *yur* 'snow', *χulān* 'thick', *yul-* 'stay'; cf. Common Turkic *qān*, *qār*, *qalın*, *qāl-*. This *y-* may go back to diphthong **i̯a*, e.g. **k'ān* 'blood', **k'ar* 'snow'.

Final *-q* is mostly preserved, e.g. Turkish *tok*, Kazakh, Kirghiz, Uzbek, Uyghur *toq*, Turkmen *doq*, Tatar *tuq* 'full, satiated'. Spirantisation is found in Azerbaijani, Khakas, Trakai Karaim, e.g. *toχ*. The final secondary lenis g^2 is mostly devoiced, e.g. $a\check{g}^2$ 'white', Turkish *ak*. The corresponding sound is represented by χ in Yakut and Chuvash, e.g. *suox*, *sux* 'non-existent'. The final stop has disappeared in Chuvash bisyllabic stems, e.g. *pulä* 'fish', *ura* 'foot'; cf. Azerbaijani *balig*, Khakas *palix* 'fish', Yakut *ataχ* 'foot'.

Medial *-q-* is often preserved, e.g. Turkish *sakal* 'beard', *bakır* 'copper', *akım* 'flow'. Intervocalic spirantisation is found in Azerbaijani, Karaim, etc., e.g. Azerbaijani *yuxarı* 'up', *yuxu* 'sleep', *yaxın* 'near', Karaim *oxu-* 'read', spirantisation and voicing in Kipchak, South Siberian and Chuvash, e.g. Tatar, Kazakh, Karachay-Balkar, Kirghiz, Khakas *sayal* 'beard', Tatar, Karachay-Balkar *bayir* 'copper', Chuvash *suyal* 'beard', written <suxal>. The secondary lenis $-g^2-$ is spirantised or deleted, e.g. Tuvan *ayar-* 'whiten'.

γ Final *-γ* is often preserved, e.g. Turkmen *dāy*, Uyghur *tay*, Tuvan *day*, Uzbek *tāy* 'mountain', Yellow Uyghur *sariy* 'yellow', Khalaj *ariy* 'pure'. It is sometimes lost, leaving minor traces, e.g. Turkish *dağ* [da:] 'mountain', Yakut *ʔa* 'taiga', Salar *sāri* 'yellow', Yakut *arī* 'butter'. It may be labialised, notably in NW, e.g. Tatar, Noghay, Kumyk, Karaim, Karakalpak *taw*, Kirghiz *tō*, Altay *tū*, Chuvash *tu* 'mountain'. Some languages have two representations in non-first syllables, e.g. Bashkir *hari* 'yellow' vs. *hiliw* 'beautiful'. Fortition, which in reality means neutralisation of the opposition *q* vs. γ , is typical of SE languages, though less so of modern Uzbek, e.g. Chaghatay *sariq*, Uzbek *sariq*, Uyghur *seriq* 'yellow'.

Intervocalic *-γ-* is preserved or disappears, often with contraction length, e.g. Turkmen *ayiz*, Uzbek *āyiz* 'mouth', Tuvan, Khakas, Altay, Shor *ās*, Kirghiz *ōz* 'mouth', Yakut *os* 'lip', Altay *ūl*, Khakas *ōl* 'son', *sōχ* 'cold'. It may also be labialised, such as in Tatar *awiz* 'mouth', *bozaw* 'calf', Azerbaijani *buzov* 'calf'.

k Initial *k-* is mostly preserved, e.g. Kirghiz *kir-* 'enter', *kes-* 'cut', Khakas *köl* 'lake', Khalaj *kel-* 'come'. It is mostly voiced in SW, e.g. Azerbaijani, Turkish *gir-* 'enter', *göl* 'lake', *göz* 'eye', Turkmen *gün* 'day, sun', sporadically in some NW languages, e.g. Kumyk *giši* 'person', Karakalpak *güz* 'autumn'. SW also has cases of preserved *k-*, e.g. Turkish *kiši* 'person', Azerbaijani, Turkish *kömür* 'coal', *kül* 'ashes', Turkmen *köl* 'lake', *köngül* 'heart'. Spirantisation is noted in Sayan Turkic, e.g. Tuvan *χöl* 'lake', *χir* 'dirt', Tofa *höl* 'lake', *hir* 'dirt', though *k-* is preserved in many words. Palatalisation is found in Turkish dialects, notably in the Balkan and Black Sea areas, Azerbaijani and Uyghur dialects, e.g. Azerbaijani *k'end* 'village', *k'ümüš* 'silver', northeastern Turkish *d^hel-* 'come', northwestern Azerbaijani *č^heči* 'goat', *j^hel-* 'come' (*k* > *č^h* and *g* < *j^h*), Uyghur *čim* 'who'.

Final *-k* is mostly preserved, e.g. Turkish, Azerbaijani, Türkmen, Kazakh *kök* 'root', Turkish, Azerbaijani, Türkmen, Noghay, Uzbek *yük*, Tatar, Bashkir *yök*, Kazakh *žük*, Karachay-Balkar, Kirghiz *jük*, Tuvan *čü?k*, Chuvash *šäk* 'burden'. Palatalisation occurs in some languages, e.g. Azerbaijani *k'ök* 'root'. The secondary lenis g^2 is mostly devoiced, e.g. Turkish *gök* 'heaven', but develops into *-y* in Azerbaijani dialects, e.g. *göy* 'heaven'. Corresponding Yakut words show spirantisation, e.g. *küöχ* 'blue'. Bisyllabic words have consonant loss in Chuvash, e.g. *ine* 'cow'; cf. Yakut *inax*.

Intervocalic *-k-* is preserved in Oghuz, e.g. *yüki* 'its burden', Azerbaijani, Turkmen *ekiz*, Turkish *ikiz* 'twin'. It is voiced in many languages, notably in Kipchak, e.g. Kazakh *segiz* 'eight', Kirghiz, Uyghur *egiz* 'twin', except when going back to a long consonant, e.g. Kirgiz *eki*, Uzbek, Uyghur *ikki* 'two'. Chuvash has *-g-*, and Sayan Turkic *-g-*, e.g. Tuvan *kö?gü* 'its root', the latter form with glottalisation as a fortis signal. Intervocalic g^2 is further weakened, spirantised or deleted, e.g. Turkish *gögü* [œy] 'its heaven'.

g Final *-g* is preserved, e.g. Uzbek *tég-*, Uyghur *teg-* 'reach, touch', or spirantised, e.g. Azerbaijani *dey-*, Tuvan *dey-*, Tatar, Bashkir, Kazakh, Kirghiz, Altay *tiy-*, Yakut *tī'y-*, Chuvash *tiv-*. Bisyllabic words often have consonant loss, e.g. Turkish, Salar *elli*, Chuvash *alla* 'fifty', sometimes with labialisation, e.g. Kirghiz *elü*. Fortition, i.e. neutralisation of the opposition *k* vs. *g*, is typical of SE languages, e.g. Uyghur *ellik*, Uzbek *ellik* 'fifty'.

Medial *-g-* has mostly developed to spirantisation or loss, e.g. Azerbaijani *deyirman*, Bashkir, Kumyk, Karachay-Balkar *tirmen*, Noghay *termen*, Khakas *terben* 'mill', Bashkir *öyren-*, Kirghiz *üyrön-* 'learn'.

t Initial *t-* is mostly preserved, e.g. Uyghur *taš*, Kazakh *tas*, Yakut *täs*, Khalaj *tāš* 'stone'. Voicing has taken place in Oghuz, though irregularly, e.g. Turkmen *dāš*, Azerbaijani *daš* vs. Turkish *taš*, Turkmen *düz* 'salt' vs. Turkish *tuz*, Azerbaijani *tök-* vs. Turkish *dök-* 'pour'. Sayan Turkic shows certain similarities, e.g. Tuvan *dolu* 'full', *daγ* 'mountain', *dört* 'four', *düş-* 'fall, *tut-* 'seize, hold'; cf. Turkish *dolu*, *dağ*, *dört*, *düş-*, *tut-*. NW has sporadic voicing, e.g. Kumyk *damyä* 'seal', Tatar *di-* 'say', Bashkir *dürt* 'four', *duñiö* 'pig', *dinğëd* 'sea', Karakalpak *diz* 'knee'. Other developments include *t'i* > *k'i* in Halič Karaim, e.g. *k'iš* 'tooth', and *t-* > *č-* before front vowels in Azerbaijani, Turkish and Uyghur dialects, e.g. *čiš* 'tooth'. Chuvash *č-* corresponds to *t-* in palatal environments, e.g. *čër* 'knee', *čëřë* 'alive'; cf. Turkish *diz*, *diri*. The element *č-* in *čul*, corresponding to Common Turkic *tāš* 'stone', may mirror an old diphthong *'ā*, e.g. **tāša*; cf. Mongolian *čila(yun)*.

Final *-t* is preserved, e.g. Uzbek *āt*, Chuvash *ut*, Khalaj *hat*, Tuvan *aṯ* 'horse'. The secondary lenis $-d^2$ is mostly devoiced, e.g. Azerbaijani *ađ* 'name'.

Medial *-t-* is mostly preserved, e.g. Kirghiz *ata*, Uzbek *ātā* 'father'. It is voiced to *-d-* in Chuvash, e.g. *vāđār* 'thirty', and to *-d-* in South Siberian, e.g.

Khakas *ada* 'father', but not in the NW languages. Secondary lenis $-d^2-$ is preserved, e.g. Turkmen *ōdun*, Turkish *odun* 'wood', Turkmen *pūdaq*, Turkish *budak* 'branch'. There are no corresponding traces of old vowel length in NW and SE, e.g. Kazakh *otın*, Kirghiz *otun* 'wood', Uzbek *butāq*, Uyghur *putaq* 'branch'.

d For a possible original initial $*d-$, see p. 96.

Intervocalic $-d-$ is fricative $-δ-$ in East Old Turkic, e.g. *adaq* 'foot'. It is represented as $-d-$ in Sayan Turkic, e.g. Tuvan *adaq* 'foot', and in another peripheral language, Khalaj, e.g. *hadaq* 'foot'. It has been devoiced in Yakut, e.g. *ataχ* 'foot'. In SW, NW and finally SE, $-δ-$ has developed into $-y-$, e.g. Kazakh *ayaq*, Salar *ayaχ* 'foot'. The development $-δ- > -z-$ is represented in Yellow Uyghur *azaq*, Khakas *azaχ* and Fu-yü *aziχ*. The shift $-δ- > -r-$ has taken place in Chuvash, e.g. *ura* 'foot', *urāl-* 'sober up', cf. Karakhanid *adil-*. There have also been various other developments in medial position due to assimilations; compare, for example, East Old Turkic *eđgü* 'good' with Tuvan *eki*, Yakut *ütüö*, Turkish *iyi*.

Final $-d$ was realised as fricative $δ$ in East Old Turkic, e.g. *bōδ* 'clan, stature'. It has been devoiced in Yakut, and at least partly devoiced in Sayan Turkic and Khalaj, e.g. Yakut, Tuvan *ket-* 'put on, wear', Tuvan *bot* 'self', Khalaj *b^uod* 'figure'. The normal development has been $δ > -y-$, e.g. Turkish *boy* 'stature', *giy-* 'dress', *koy-* 'put'. The development $-δ > -z$ is represented in Yellow Uyghur *poz* 'body', Khakas *pos*, Fu-yü *bos* 'self'. The shift $-δ- > -r$ has taken place in Chuvash, e.g. *χur-* 'put'.

p For initial $p-$, see pp. 95–6. Final $-p$ is generally preserved, e.g. *alp* 'hero', *tap-* 'find', *sep-* 'pour', *ip* 'rope'. It may be strongly aspirated and even develop into f in some Azerbaijani dialects, e.g. *t^hop^h* 'cannon'. Intervocalic $-p-$ is preserved in many languages, e.g. Turkish *tepe*, Turkmen *depe*, Uyghur *töpe*, Uzbek *tēpā* 'hill'. It is voiced in some languages, notably NW, e.g. Kazakh *töbe*, Kirghiz *döbö*, Tatar, Bashkir *tübe*. The secondary lenis b^2 is devoiced in final position, e.g. Turkmen *dūp* 'bottom', *gāp* 'vessel' and preserved or spirantised intervocally, e.g. Turkish *gib²i* 'like' ← *kib²* 'shape, form', Turkmen *dūbü* 'its bottom'.

b- Initial $b-$ is mostly preserved, though it may be voiceless in certain words, e.g. Turkish *piš-* 'boil', *parmaq* 'finger', Azerbaijani *poz-* 'spoil'. Tuvan, Khakas, Shor, Chuvash, Yellow Uyghur and Fu-yü display at least partially voiceless labials, e.g. Khakas, Shor *paš*, Chuvash *puš*, Yellow Uyghur, Salar *paš* 'head'. These labials are opposed to strongly aspirated fortis of the type p^h , occurring in loanwords, e.g. Tuvan *p^haš* 'cast-iron cup'.

Nasalisation conditioned by a following nasal has long been common outside Oghuz, e.g. Yakut *murun* 'nose', Uyghur *min-* 'mount, ride'. It has spread considerably also in SW to the effect that only Turkish has $b-$ today,

e.g. *ben* 'I' vs. Turkmen *men*. Another exception is Chuvash, e.g. *eḷḷe* 'I'. There is also nasalisation in a few other words, e.g. Uzbek *muz* 'ice'.

West Oghuz tends towards spirantisation and loss, e.g. Turkish, Azerbaijani *ver-* 'give', *var* 'existent' vs. Turkmen, Khorasan Turkic *ber-*, Turkmen *bār*, Turkish, Azerbaijani *ol-* 'become, be' vs. Turkmen *bol-*.

Final $-b$ was spirantised early, maybe already in East Old Turkic. It has shifted to v , w and $θ$, e.g. Turkmen *θuw* 'water', *āw* 'hunting', Kirghiz *sū*, Turkish *su* 'water'. In some languages, $*-eb$ has yielded special labial forms, e.g. Kazakh, Kirghiz, Uzbek *üy*, Bashkir, Tatar, Turkmen, Uyghur *öy* 'house', Kazakh, Kirghiz *süy-*, Bashkir *höy-*, Tatar, Uyghur *söy-*, Turkmen *θöy-* 'love'.

Medial $-b-$ has developed correspondingly. The development to $-v-$ was typical of early Oghuz, e.g. *deve(y)* 'camel', *evet* 'yes', *yavlaq* 'bad'. Spirantisation to $-w-$ is common in NW, e.g. Tatar *yawız* 'bad'; cf. Turkish *yavuz* 'stern'. A secondarily developed lenis $-b^2-$ in words of foreign origin may also be spirantised, e.g. Turkmen *kitāwī*, Uyghur *kitiwi* 'his/her book'.

č Initial $č-$ is often preserved. It develops into $š-$ in some Kipchak and Siberian languages, Tatar, Kazakh, Karakalpak, Noghay, Tuvan, Tofa and Shor, e.g. Kazakh *šiq-* 'go out', Tuvan *šay* 'tea'. Chuvash has a strongly palatal $š-$, e.g. *šup-* 'strike'. A further development to $s-$ affects Bashkir, Khakas and Yakut, e.g. Bashkir *sěk-*, Khakas *sīχ-* 'go out', Yakut *sax-* 'strike fire'. There are also shifts to c and similar sounds in Tatar dialects, Halič Karaim, northeastern dialects of Turkish, etc., e.g. Halič Karaim *cek'ic* 'hammer'.

Medial $-č-$ is preserved, e.g. Turkish *ačiq* 'open', Azerbaijani *ačiqχ-*, or undergoes the changes mentioned above: e.g. Kazakh *ašiq*, Bashkir *āšiq*, Mishar *ačiq*, Halič Karaim *aci* 'bitter'. Chuvash has partial voicing, e.g. *kužar-* 'transfer' and South Siberian Turkic has voiced forms such as Tuvan *ažiq* 'open', Khakas *azil-* 'be opened' ← *as-* < *aš-* 'open'. Yakut displays the development $č > -s- > -h-$, e.g. *bihax* 'knife'. A secondary lenis $-j^2-$ is preserved as such, e.g. Turkish *actik-* 'feel hungry'.

Final $-č$ is preserved or undergoes the shifts mentioned above, e.g. Noghay *üş* 'three', *qaš-* 'flee', Chuvash *ěs-* 'drink', Bashkir *ěs*, Yakut *is* 'inside', Khakas *as-* 'open'. A secondary lenis $-j^2$ is devoiced, e.g. Turkish *aç*, Azerbaijani *aj*, Turkmen *āj*, 'hungry'.

s Initial $s-$ is mostly preserved, e.g. Turkish *sarı* 'yellow'. It is spirantised in Turkmen, e.g. *θarı* 'yellow', *θid* 'you', and develops into h in Bashkir, e.g. *hâqal* 'beard', *hõt* 'milk', *hüđ* 'word'. Yakut represents the $θ$ stage, e.g. *ūt* 'milk', *en* 'thou', *ari* 'butter', *ū* 'water'. The normal Chuvash representation is $s-$, e.g. *sēt* 'milk'. Cases with $š$, e.g. *šiv* 'water', *šurā* 'white', may possibly be due to a palatal combination such as $*s^i$, which developed into $s-$ in Common Turkic. Medial $-s-$ is mostly preserved. It develops into $-θ-$ in Turkmen and Bashkir, e.g. Turkmen *iθī* 'heat', Bashkir *kēθek* 'piece'. Chuvash has $-z-$, e.g. *užā* 'profit, use', cf. Karakhanid *asiy*. South Siberian

languages (except Shor) exhibit voicing, e.g. Altay *kezek*, Khakas *kizek* 'piece, segment'. Final *-s* is spirantised in Bashkir and Turkmen, e.g. Turkmen *θeθ* 'voice'.

r Initial *r-* is assumed not to have existed in Proto-Turkic, and it still does not occur in native words. Medial *-r-* is usually preserved, e.g. Kirghiz *qara*, Chuvash *χura* 'black', but may disappear in combination with consonants, e.g. Uyghur *qaya* < *qarya* 'crow'. Final *-r* is preserved, e.g. Azerbaijani *ğar*, Chuvash *yur* 'snow'. There is often loss of *-r* in Uyghur, e.g. *qa* 'snow', *ye* 'place'; cf. Turkish *kar*, *yer*.

z Initial *z-* is assumed not to have existed in Proto-Turkic, and it still only occurs in loanwords. Medial *-z-* and final *-z* have undergone spirantisation in Bashkir and Turkmen, e.g. Bashkir *hëð*, Turkmen *θið* 'you' (pl.). Final *-z* is often devoiced, e.g. Yakut *kīs* 'girl'. In a few cases, Yakut *-t* corresponds to *-z*, e.g. *otut* 'thirty'.

As noted above, Chuvash *r* may correspond to Common Turkic *r*, e.g. *yur* 'snow' vs. *qār*. In certain words, however, it corresponds to Common Turkic *z*, e.g. *χur* 'goose' vs. *qāz*, *tāχχār* 'nine' vs. *toqquz*, *yëker* 'twin' vs. *ekiz*, *paru* vs. *buzāyu* 'calf', *sër* 'hundred' vs. *yūz*, *čër* 'knee' vs. *tīz*. This *r* is also reflected in synonyms such as Mongolian *ikire*, Manchu *ikiri*, Hungarian *iker* 'twin', Hungarian *borjú* 'calf'.

It possibly represents an early **rⁱ*, which became *r* in the Oghur group, and *z* through 'zetacism' in Common Turkic. Altaicists suppose that the corresponding Mongolic and Tungusic words developed like the Oghur ones. Others explain them as copies from early Proto-Turkic and the Hungarian words as copies from Old Chuvash. In both cases, Mongolian *bora* 'grey' might go back to **bōrⁱa*, which developed into Turkic **bōrⁱ* (loss of short unstressed final vowels) and further to *bōz*.

Other scholars claim that Chuvash has developed this second *r* through 'rhotacism', a shift *z > r*, and that even certain copied lexical elements have taken part in this change, e.g. *pir* 'linen' < **bōz*, *χir* 'pine' < Mordva *kuz*. One opinion is that an old shift *z > r* in certain phonetic environments was generalised in the ancestor of Chuvash, whereas it stopped in other Turkic languages (see pp. 71–2).

l Most Turkic languages display one front and one back *l*, e.g. *yel* 'wind', *yot* 'way'. This difference may be neutralised, notably in Iranised varieties. Initial *l-* is assumed not to have existed in Proto-Turkic, and it still only occurs in copied words or after loss of an initial vowel, e.g. Chuvash *lar-* 'sit'. Medial *-l-* is mostly preserved, e.g. Azerbaijani *bulud* 'cloud', but is less stable in combinations with consonants, e.g. Turkish *getir-* 'bring', *otur-* 'sit' (< *lt*), Uyghur *bosa* 'if it is' (< *ls*). Final *-l* is stable, e.g. Turkmen *yöl*, Kazakh *žol*, Tatar *yul*, Chuvash *sul*.

š Initial *š-* is supposed not to have existed in Proto-Turkic, and it is still very rare except when going back to *č-*. Certain other cases of *š-* in native words may go back to *sⁱ-*. A few demonstratives have *š-*, e.g. Uyghur (*a*)*šu* 'that', Kirghiz (*u*)*šul*, 'this here', (*o*)*šol* 'that', Turkish *šu* 'this here', cf. *s-* in Noghay, Kazakh, Yakut *sol*. Final *-š* is mostly preserved, e.g. Tatar *baş* 'head', Turkmen *ğāš* 'eyebrow'. It has shifted to *-s* in Kazakh, Karakalpak, Noghay, Khakas, Yakut and Halič Karaim, e.g. Kazakh *qīs* 'winter', Khakas *tas* 'stone', Yakut *χās* 'eyebrow'. Medial *-š-* is mostly preserved, e.g. Uzbek *kīši* 'person'. South Siberian Turkic languages exhibit voiced forms, e.g. Tuvan, Altay, Shor *kīži*. Kazakh, Karakalpak, Noghay have *-s-*, e.g. Kazakh *kīši*. This *-s-* is voiced in Khakas, e.g. *kīzi*. Yakut displays the result of a development *-š- > -s- > -h-* in, e.g., *kīhi* 'person', *ihit-* 'hear', *kīhīn* 'in winter'.

Chuvash *l* may, as noted above, correspond to Common Turkic *l*, e.g. in *sul* 'way' vs. *yöl*. In certain cases, however, it corresponds to Common Turkic non-initial *š*, e.g. *čul* 'stone' vs. *tāš*, *kēmēl* 'silver' vs. *kümüš*, *χēl* 'winter' vs. *qīš*, *sul* 'year' vs. *yaš* 'age', *alāk* 'door' vs. *ēšik*, *pil(ēk)* 'five' vs. *bēš*. The thirteenth to fourteenth century Volga Bulghar inscriptions suggest forms such as *biⁱl* 'five' and *jāl* 'year (of age)' for the ancestor of Chuvash. Certain synonyms in other languages also have *l*, e.g. Mongolian *čila(yun)* 'stone' vs. *tāš*, Mongolian *göligē* 'puppy', Hungarian *kölyök* 'whelp' vs. *köšek* 'young animal'.

The relation of this second *l* to *š* is similar to that of the second *r* to *z*. The second *l* possibly represents an old *lⁱ*, which has developed into Chuvash *l* and Turkic *š*. Altaicists take the corresponding Mongolic words to be Altaic cognates, whereas others consider them early copies from Turkic. Another solution involves 'lambdaism', which means that an old *š* or *ž* has developed into Chuvash *l*. One opinion is that, due to neutralisation of the Proto-Turkic opposition *š* vs. *l* in special phonetic environments, *š* was generalised in Common Turkic, and *l* in the ancestor of Chuvash, albeit with some exceptions.

An exception that complicates the picture is the fact that Chuvash sometimes has *š*, e.g. Chuvash *puš* 'head' vs. Common Turkic *bāš*. This may be one of the cases where Turkic *š* goes back to an old **lē*; compare, for example, e.g. *bēšek* 'cradle' with Hungarian *bölcső*, or *ešgek* 'donkey' with Mongolian *eljigen*. Some scholars, however, explain such cases as a Turkic shift *š > č*.

n Initial *n-* presents another controversial issue. It may have occurred at an early stage. Altaicists take it to have existed in Proto-Altaic and, like some other consonants, to have developed into Common Turkic *y-*. Thus, Turkic *yāz* 'spring' is compared with Mongolian *niray* 'fresh' and Hungarian *nyár* 'summer', the latter a possible loanword.

South Siberian Turkic languages, some Turkic dialects of Iran etc. display an initial *n* or *n* that is commonly not explained as a preserved old feature, but

as a product of *y*-assimilating to a following nasal, e.g. Tofa *ñan-*, Shor *nan-* 'return', Shor *nañis*, Tofa *ñuñus* 'solitary', Khakas *nañmir* 'rain', Khakas, Fu-yü *nāχ* 'cheek' (cf. *yañaq*), Khakas *nā*, Shor *ñā*, *nā* 'new' (cf. *yañi*), Fu-yü *nomurtga* 'egg'; cf. Turkish *yumurta*.

Non-initial *n*, which seems to go back to an element *n'*, is found in East Old Turkic proper, e.g. *kön-* 'burn', *qōñ* 'sheep', *an'γ* 'sin'. Old Uyghur notations with *y*- instead may be of purely graphic nature, or indicate a dialectal difference. The Yakut counterpart is *ñ*, a nasalised *y*, e.g. *añi* 'guilt'. Most Turkic languages display *y*, e.g. Kirghiz *küy-* 'burn', *qoy* 'sheep'. However, *n* was already typical of the eleventh century Arghu dialect, and is still retained in its descendant Khalaj, e.g. *qā'naq* 'cream', *ki'n-* ~ *kön-* 'burn', *qo'n* 'sheep'. Oghuz preserves the nasal in the shape of *yn* etc., e.g. Azerbaijani *ğoyun*, Turkish *koyun* 'sheep' (with vowel insertion), Turkish *kaymak* 'cream'. The interrogative stem *qāñ-* has developed correspondingly, e.g. Kazakh, Uzbek *qaysi*, Yakut *χaya*, Khalaj *qā'ni*, Turkish *hangi*, Azerbaijani *hansı* 'which'.

ŋ Final *-ŋ* is preserved, e.g. Uyghur, Kirghiz, Karachay-Balkar, Altay *keŋ*, Yakut *ki'ŋ*, Turkmen *gīŋ* 'wide', or has developed to *-n*, e.g. Turkish *geniŝ* 'wide', *tan* 'dawn'. The type *biŋ* 'thousand' seems to go back to **biŋan*; cf. Mongolian *miŋgan*. Medial *-ŋ-* is also preserved, e.g. Tatar *yaŋa*, Yakut *saŋa*, Kazakh *zaŋa*, Uyghur *yeŋi* 'new', or has become *-n-*, e.g. Turkish *yeni*, Azerbaijani *yeni*, Chuvash *šēñě*. West Oghuz sometimes has *-m-*, e.g. Azerbaijani *sümü̇k* 'bone', Turkish *domuz*; cf. Turkmen *doñuđ*. Other languages often exhibit a shift to *-y-* in front environments, e.g. Kazakh *süyek*, Kirghiz, Altay, Khakas *sök* 'bone'.

n Initial *n-* does not occur in native words, which raises the question whether *ne* 'what' is originally Turkic. Medial *-n-* is stable, e.g. *ana* 'mother'. Final *-n* is mostly preserved, e.g. Turkmen *yān*, Karakalpak *žan*, Kirghiz *jan*, Uzbek *yān*, Altay *d'an*, Khakas, Tuvan *čan* 'side'. The final *-n* in words such as *titiün* 'smoke', *uzun* 'long', *sān* 'number' seems to go back to a final *-m*, judging from forms such as Chuvash *tēđēm* 'smoke', *vārām* 'long', Hungarian *szám* 'number', the latter a possible loanword. An old **-mÁ* has not developed into *n*, e.g. Common Turkic *ēm* 'trousers'; cf. Chuvash *yēm*.

m Initial *m-* is supposed not to have existed in Proto-Turkic, and it still only occurs in copied words and in forms assimilated due to a following nasal, e.g. *men* 'I' < *ben* (pp. 112–13).

Medial *-m-* is normally preserved as such. Note that South Siberian Turkic sometimes has *-b-* instead, e.g. Khakas *kebe* 'ship', *tebir* 'iron'. Suffixes which begin with *m-* in other languages often have *b-* and *β-* after vowels, e.g. Tuvan *Bižiβedim* 'I did not write' < *biži-* 'write'.

Irregularity and Coincidence of Stems

Due to the changes dealt with above, stem irregularities have emerged in many languages. Alternation may be caused by secondary lenes, e.g. Turkmen *git-* 'go' vs. *gidyār* 'goes', *gīč* 'late' vs. *gījirāk* 'later', *āq* 'white' vs. *āyiraq* 'whiter', vowel raising and umlauting, e.g. Uyghur *kitap* vs. *kitiwi* 'its book', *baş* 'head' vs. *bēšim* 'my head', *al-* 'take' vs. *ēlip* 'taking', etc., contraction and assimilations, e.g. Tuvan *sayin-* 'think of', *saqtir* 'thinks of', or consonant shifts, e.g. Yakut *kīs* 'daughter' vs. *kīha* 'to a/the daughter', *bīs-* 'cut' vs. *bīhabin* 'I cut'.

All such changes and shifts of segment combinations may cause more or less drastic cases of coincidence. Yakut *as-* does not only represent **ač-* 'open', but also **sanj-* 'pierce'. The *-t-* in Yakut *utaχ* 'thirst' probably reflects the cluster in **sūbsa-* 'be thirsty'. Remarkable sets of homonyms have emerged in Khakas since *č* and *š* have become *s*, and final *-z* is devoiced. The form *as* can mean 'open!', 'pass over!', 'go astray!', 'food', 'weasel', 'little', 'hungry'. Compare the Turkish lexemes *aç-*, *aş-*, *az-*, *aş*, *as*, *az*, *aç*.

Other phenomena that have affected stems include metathesis, e.g. Tuvan *ökpe* < *öpke* 'lung', Azerbaijani *ireli* 'forward' < *ileri*, and dissimilation within long sonorants or sonorant clusters, e.g. Kirghiz *molđo* 'mullah' < *molla*, Trakai Karaim *uglu* 'big' < *ullu*.

Suffix Phonology

The phonological development in suffixes differs considerably from the development in primary stems.

Vowels

Of the vowels occurring in Proto-Turkic suffix syllables, the long ones later remained as full vowels. The short ones were reduced and lost in word-final position, and partly preserved as reduced vowels in medial position.

East Old Turkic texts show that the vowel oppositions valid in first syllables are largely neutralised in suffixes. They clearly mirror the full vowels *a*, *e*, *i* and two types of rounded vowels, *o*, *ö* and/or *u*, *ü*. They also suggest the existence of a particular, probably reduced, vowel type which will be denoted here as *°*. East Old Turkic had four classes of suffix vowels:

- 1 low vowels *A*, realised as *a*, *e*, e.g. in the plural suffix *-lAr*;
- 2 high unrounded vowel(s) *I*, realised as *i*, according to some scholars also as *ĩ*, e.g. in the participle suffix *-miš*;
- 3 rounded vowels, *O* and/or *U*, realised as *o*, *ö* and/or *u*, *ü*, e.g. in the deverbal suffix *-DWK*.
- 4 The vowel(s) *°*, e.g. in the 1p.sg. possessive suffix *-(°)m*, passive suffix *-(°)l*, accusative suffix *-(°)G*.

The vowels of the last class had developed from short vowels in unstressed word-internal position, e.g. *oy'l* 'son'. It is a matter of dispute whether this

class consisted of a phonetically reduced *schwa* (ə), a twofold alternation of reduced vowels (ǎ, ě), a fourfold alternation of reduced vowels (ǎ, ě, ő, ő̃), or even a fourfold alternation of full vowels (i, i, u, ü). The latter option would mean that East Old Turkic already had the type of alternation represented by modern written Turkish, a system which has developed there in the last three centuries.

It is obvious from the graphical conventions of the runiform script that East Old Turkic had intersyllabic front vs. back harmony. But there were also, as at all stages of development, non-harmonic suffixes with one invariable form. The dative suffix *-qa* was a back suffix, while the possessive suffix *-(s)i*, the post-terminal suffix *-miš*, the nomen actoris suffix *-či* were front ones. There are thus disharmonic sequences in the oldest Turkic words known to us.

Further Development

The further development of the suffix classes is characterised by the following:

- 1 The A class is preserved.
- 2 The I class mostly displays front vs. back harmony.
- 3 The U class is established, which involves raising of earlier *-o*, *-ö* to *-u*, *-ü*.
- 4 The ° class, developed from Proto-Turkic short vowels, develops front vs. back harmony.

Though the front vs. back harmony is thus widened, many exceptions remain until rather late. As is clearly shown by the West Oghuz development, there are 'indifference stages' preceding the stage of vowel harmony and displaying preliminary symptoms such as neutral vowels, which tend towards a phonetically central position.

Rounded vs. unrounded harmony was not developed in East Old Turkic. At least the variant operating on high vowels has developed to some degree in all subsequent languages and is still gaining ground.

The establishment of both front vs. back and rounded vs. unrounded harmony has, in some languages, led to coincidence of the classes I, U and ° into X (i, i, u, ü). The development to this fourfold harmony in Ottoman Turkish was a long and complicated process not concluded before the eighteenth century. In languages such as Tatar and Crimean Tatar, it has not yet been carried through.

In a number of other languages, the labial harmony also began to operate on the low vowels of the A class, e.g. Kirghiz *köllördö* 'in the lakes', Altay *bolboyondor* 'those who have not come into being', Turkmen *oylonlormuđđo* 'in our young men'. This harmony is highly developed in Turkmen, Yakut, Kirghiz, Altay Turkic (originally in the southern dialects), but also occurs in Kazakh, Karakalpak, Noghay, Kipchak, Uzbek, Balkar and Uyghur.

Later Cases of Non-harmonic Vowels

Non-harmonic suffixes represent an early stage of suffix development and are not necessarily signs of foreign influence. Suffix harmony is often claimed to have ceased to function in dialects where it has not even started. Suffixes such as Chaghatay and Kuman *-če* may represent a stage prior to assimilation, just like the Chuvash plural suffix *-sem* still does. Ottoman preserved non-harmonic suffixes with *i* until very late.

On the other hand, strong foreign influence has infringed upon the development of harmonic suffixes in certain languages, e.g. Uzbek and Khalaj. This does not, however, mean a breakdown of the phonological distinction front vs. back (see pp. 31, 32–3, 89, 91–2). The intersyllabic front vs. back harmony may even function in default of phonetically clear-cut stem vowels. In Karakhanid, where no distinction *i* vs. *ı* is detectable, there is a clear tendency for a given primary stem to prefer front or back suffixes. Thus, *il-* 'stick' takes front suffixes, and *il-* 'come down' takes back suffixes. Uyghur primary stems containing the neutral *i* have similar preferences, e.g. *išlar* 'works' ← *iš* 'work'.

The sound changes discussed above have their effects on suffixes. Uyghur raising of *a*, *e*, *é* in unstressed position produces suffix vowels as in *baliliri* 'their children' < **balalari*, *Alisen* 'You take' < **Alasen*, *Boludu* 'It becomes' < **Bola turur*. South Siberian intervocalic voicing yields suffix variants such as in Altay *adaži* 'his/her father', Khakas *pozim* 'myself', *pazim* 'my head'. The Kumyk loss of *r* in certain cases produces plural forms such as *-leni* (-LAR + genitive-accusative), *-lege* (-LAR + dative).

Consonants

Assimilations due to contact of stem-final dentals, nasals and sibilants with suffix-initial liquids and nasals are very common in the eastern area, notably in Siberia. They are also much more developed in the Aralo-Caspian group (NW^S) than in West Kipchak (NW^W) and SE. The southern and western dialects of Kazakh exhibit fewer assimilations than the northern one. Note that the dental of locative suffixes is often not affected, since assimilation might lead to confusion with the dative suffix of the third person, e.g. Kirghiz ablative *atınan* 'from his horse', but locative *atında* 'on his horse', cf. dative *atına* 'to his horse'.

In Oghuz, except in part of Khorasan Turkic, suffix-initial *G-* has shifted to *y-* after originally long stem-final vowels and has otherwise disappeared, e.g. *başla-yan* 'beginning', *gel-en* 'coming', *iç-eri* 'inside'. There is thus a dative in *-(y)A*, an optative in *-(y)A*, an abtemporal converb in *-(y)All*, etc. Forms with *G-* such as *-GIn*, *-GAč* and *-GAN*, e.g. Turkish *yorgun* 'tired', *sürgün* 'exile', *utangaç* 'bashful', *çalışkan* 'diligent', go back to different structures, e.g. *çalışkan* < **çalış-ayan*.

Suffix-final *-G* has developed in many ways. It has been preserved in NE^S, except in Altay, e.g. Tuvan *dayliŋ* 'mountainous', *a?ttiŋ* 'having a horse',

küştüy 'forceful'. Yakut has *-LĀχ*, e.g. *küstēχ* 'forceful', Khalaj *-luy*, e.g. *hatluy* 'horseman'. The consonant has disappeared in Oghuz and Chuvash, e.g. Turkish *diri* 'alive', *yazıcı* 'writer' < **yaz˘yçi*, *-li* in *dağlı* 'mountainous', *sütlü* 'milky', etc., Chuvash in *-LĀ*, e.g. *tullā* 'mountainous'. It has disappeared, often producing rounded suffix vowels in NW, e.g. Karachay-Balkar *tawlū*, cf. Kazakh *tawlı* 'mountainous', Kirghiz *attū*, Altay *attu*, 'having a horse', Kirghiz *küčtū*, Altay *küčtū* 'forceful'. It has become *-K* in SE^E and thus coincided with the original *-K*, e.g. Uyghur *sütlük* 'milky', *taylıq* 'mountainous', *uluq* 'big'. Thus Uzbek has, like Chaghatay, lexicalised forms such as *quruq* 'dry', *tirik* 'alive', *ölük* 'dead', whereas it tends towards consonant loss in productive suffixes such as *-li*, e.g. *küçli* 'forceful'.

Phonological Adaptation of Lexical Copies

Lexical elements copied from other languages have mostly been phonologically adapted to the Turkic systems and reshaped. It must, however, be mentioned that, in most Soviet Turkic languages, recent Russian loanwords were written and often also pronounced without noticeable adaptations.

Foreign segments have been replaced by those native segments which were felt to be their closest equivalents. Thus, in many Central Asian and neighbouring languages, *p* has replaced *f*, e.g. Kazakh, Kirghiz, Uyghur, Kumyk *payda* 'profit, use' ← Arabic, Turkmen, Kirghiz *tarap*, Uyghur *terep* 'side' ← Arabic; cf. Turkish *fayda*, *taraf*. In Tatar and Bashkir, *γ* has replaced Arabic glottal stops, e.g. *γömr* 'life', *γadet* 'habit', *seyet* 'hour'; cf. Turkish *ömür*, *adet*, *saat*. In Bashkir and some other languages, *y* has replaced *j*, e.g. *yen* 'soul' ← Persian; cf. Turkish *can*. In some NW^S languages, *q* and *θ* have replaced *χ* and *h*, e.g. Kirghiz *qabar* 'message' ← Arabic, *ar* 'each' ← Persian; cf. Turkish *haber*, *her*.

Unallowed consonant clusters have often been dissolved, according to Turkic phonotactic principles, with prothetic, epenthetic or epithetic vowels, deletion (syncope, apocope) and metathesis, e.g. Turkish *gireyfurut* 'grapefruit' ← English, Kirghiz *dos* 'friend' ← Persian, Kumyk *fikru* 'thought' ← Arabic; cf. Turkish *dost*, *fikir*.

Short vowels have mostly replaced long vowels. Arabic and Persian vowels are rendered differently with respect to the front vs. back distinction, e.g. Azerbaijani *xeber* 'message', *teref* 'side', Turkish *haber*, *taraf*.

When loanwords are adapted to front vs. back harmony, foreign syllables are classified according to the phonology of the copying language. The result may vary considerably. Persian and Russian syllables with *k* and *g* are mostly interpreted as front, e.g. Uzbek *göšt* 'meat', Karakalpak *kelliyktip* 'collective'. German *k* is mostly classified as back in Turkish, since the Turkish front counterpart is more fronted, e.g. *kirimi* '(crime) thriller'. Intrasyllabic disharmony is often found in not fully integrated foreign lexemes. Such cases must also have existed in older loanwords, e.g. in Old Uyghur copies of Indo-Iranian words, which may well have contained syllables with *k* or *g*

co-occurring with a back vowel such as *a*, *o*, *u*. The widespread practice of transcribing Old Uyghur by just marking the vowels for front- and backness is, however, based on the assumption that front vowels always co-occur with front consonants, and vice versa.

Intersyllabic harmonising, e.g. Turkish *madalya* 'medal' ← Italian, very often plays a less prominent role and occurs later than other adaptation processes. Languages such as Uzbek, Uyghur, Azerbaijani and Turkish are relatively modest in this respect, e.g. Turkish *günâh* 'sin' ← Persian, whereas other languages are more consistent, e.g. Turkmen *günä*, Kirghiz *künö*.

Accent and Intonation

The assumption of a certain accent prominence of the first syllable at early stages of Turkic would help to explain phenomena such as the reduction and partial loss of the final short vowels, the emergence of sound harmony and Old Turkic alliteration patterns. As for intonation, there is no evidence of its historical development. It is, however, clear that certain languages have taken over strong intonation patterns from dominant contact languages, e.g. Khalaj from Persian.

Morphology

The morphological structure of Turkic has remained relatively stable through the centuries. There are considerable similarities between modern languages, but also often sizeable functional differences between etymologically related forms. The morphology of languages such as Chuvash and Khalaj deviates a great deal from that of the rest. In what follows, only a few selected examples of morphological development will be given.

Declension

The case forms have been rather stable. The dative suffixes normally have the shape *-GA*, in Oghuz *-(y)A*, Chuvash (dative-accusative) *-(n)A* etc., Khalaj *-KA*. As for the genitive, Orkhon Turkic has the archaic suffix *-(°)ŋ*, of which there are traces in Chuvash as well. The type *-nlŋ* is found in Old Uyghur, Karakhanid, Chaghatay, with later forms such as *-nlŋ*, *-(n)lŋ*, *-(n)ln*. Yakut has lost its genitive suffix, but retains a rest *-n-* in possessive constructions with third-person suffixes. As for accusative suffixes, Orkhon Turkic has *-(°)G*, Old Uyghur and Karakhanid *-IG* and *-nl*, Chaghatay and later stages *-nl*, SW *-(y)l*.

Orkhon Turkic has a locative-ablative in *-DA*, also found in Karakhanid. It is later on mostly used as a locative, though in Yakut as a partitive. The ablative function is typical of older Oghuz and present-day Khalaj, which displays a locative in *-čA*, developed from the case normally used as equative. The ablative suffix is *-DIn* in Old Uyghur, Chaghatay, Uyghur and Uzbek dialects, elsewhere *-DAn* (except for Altay Turkic *-DAŋ*).

The introduction of accusatives and genitives in *-nl* and *-nlŋ* from Old

Uyghur onwards is considered an influence of the pronominal flexion. Old Turkic personal and demonstrative pronouns have a declension distinct from that of nouns, e.g. an oblique stem in *-n*. The personal pronouns *ben* 'I' and *sen* 'thou' are formed from **bi* and **si* and exhibit plurals in *-z*: *biz* 'we', *siz* 'you'. Cf. Chuvash *eḫē* 'I', *ežē* 'thou', *eḫir* 'we', *ežir* 'you', all with a proclitic *e-*. Forms based on oblique stems include *bağa* 'to me', *saga* 'to thee', etc. Demonstratives such as *ol* 'that' have oblique stems such as *an-*. A specific pronominal declension is preserved in most modern languages, e.g. Turkish *bana*, *sana*, but some languages such as Uzbek and Uyghur have given it up in favour of the noun inflection, e.g. Uzbek *mengä*, *sengä*.

Personal markers of the pronominal type have – with many exceptions – developed along the lines *ben* > *-mAn* > *-m*, *sen* > *-n*, *biz* > *-blzl/mlz*, *siz* > *-slz*. The Oghuz languages normally exhibit 1p.sg. *-vAn*, *-(y)Am*, *-(y)Im*, 1p.pl. *-(v)Uz* > *-(y)Iz* and *-(y)IG*, *-(y)IK* etc. The pronoun *ol* is also used as a third-person copula in older Turkic. Possessive suffixes are usually *-m*, *-ŋ*, *-(s)l*, plurals of the first and second persons being formed with *-z*.

In Old Uyghur, Karakhanid and Khorezmian Turkic, third-person possessive suffixes exhibit the so-called 'pronominal *n*' in front of case suffixes, e.g. *atıŋa* 'to his/her horse'. The corresponding accusative ending is usually only *-n*, e.g. *atın* 'his/her horse' (accusative). The 'pronominal *n*' was abandoned in Chaghatay and is absent in modern Uyghur and Uzbek, e.g. Uzbek *atıgä* 'to his/her horse', *änäsida* 'at his/her mother', whereas it is still used elsewhere, e.g. Turkish *atına*, *annesinde*. In Yakut, nouns carrying possessive markers have very specific declension paradigms.

Formation of Postpositive Markers

Turkic postpositive markers have been formed in systematic developmental processes of several discernible stages. Possessive constructions developing into postpositions and verbal composition developing into postverbs are examples of the shift from the 'right-to-left' modification order outside the word, e.g. Turkish *bizim köy* 'our village', to the 'left-to-right' order inside the word, e.g. *köy-ümüz* 'our village'. The tension between the two orders becomes especially obvious when analytic constructions develop into synthetic ones, thus changing the word boundaries.

Postpositions

One example is the development of postpositions. At a first stage, two lexical items combine in a genitive construction, e.g. Turkish *ev-in ön-ün-de* 'on the front side of the house', where *önü* is the head of the preceding nominal. By way of reanalysis, the second item, a spatial noun with a possessive and a case suffix, develops into a grammatical relator with the more general content 'in front of'. This reanalysis means a reversal of the relationship of modification. Though still a declined noun, *önü* is no more a nominal modified by a genitive attribute *evin*. Postpositions of late origin such as *önünde* can still occur freely as adverbials. It may then become possible to drop the genitive suffix, e.g.

ev(in) önünde, or impossible to use it (except for pronouns). The case suffix may also become standardised (see pp. 221–2), e.g. *hakkında* ('in its right') 'concerning'.

Though the details are unclear, old unanalysable Turkic postpositions such as *üçün* 'for' seem to be of similar lexical origin, probably having a career as case forms behind them. Postpositions at this stage do not occur freely as adverbials any more. They may, however, still govern a case, e.g. Turkish *sen-in için* 'for you' (with the genitive of a personal pronoun). This type plays the role of an unaccentable enclitic particle.

The next step is the development into a peripheral case suffix. It may at first be non-harmonic, and later on assimilate to the stem. At a transitory stage, both free and bounded variants may occur, such as with Turkish *ile* ~ *-(y)lA* in *uçak ile* ~ *uçakla* 'with aircraft'. Several other markers vacillate between non-suffix and suffix status, displaying free enclitic and bound variants, e.g. copula particles such as Turkish *idi* ~ *-(y)di* 'was'. The development is far from predictable. Thus, the possessive marker *içün* 'for' had a suffixed variant *-çün* in Ottoman, while the modern standard Turkish *için* cannot be suffixed.

It is unclear whether such non-accentable suffixed elements may develop into accentable suffixes of the kind found in core cases (dative, etc.). At any rate, accentability indicates the high historical age of a suffix more clearly than sound harmony. Many non-accentable suffixes can be traced back to compositional structures. Personal markers of the pronominal type clearly originate in postpositive personal pronouns, e.g. *-mAn* < *ben* 'I'. The non-accentable verbal negation suffix *-mA* also seems to suggest a compositional origin.

Each form has its own history along such paths. It is impossible to determine which Proto-Turkic formants were already suffixes. Elements known to us as suffixes are likely to have been far behind on the path at those stages.

Postverbs

Postverbs emerge, like postpositions, through grammaticalisation of lexical items. They consist of a converb marker followed by an auxiliary verb and express modifications of the actional content of the preceding main verb, e.g. durativity, iterativity, phase specification, transformativity, non-transformativity. Transformative verbs denote telic actions, implying a natural turning point, whereas non-transformatives lack this feature (cf. p. 42). Non-transformatives and duratives are often formed with stative verbs such as 'stand', 'lie', etc., e.g. Khakas *çat-* 'lie', *tur-* 'stand', *odir-* 'sit', *çör-* 'go'. Transformatives are formed with auxiliaries meaning 'send', etc., e.g. Kumyk *yiber-*, Khakas *is-*, Tuvan *it-*.

The point of departure of the development is a verbal composition. In a combination such as *yaz-a tur-* 'stand writing', the converb is originally

syntactically subordinated to the last verb. However, the last verb may develop into a functional item, which reverses the modification relation. In *yaz-a tur-* 'write continuously', the converb is no more an adverbial modifier of *tur-*, but is itself modified by it. The postverb can also fuse with the preceding item, developing into a suffix, e.g. *yazadur-*. It may also become subject to sound harmony, e.g. Khakas *kizibis-* < *kisip is-* 'cut off', Tuvan *biživit-* 'write down' < *bižip it-*.

The postverb system is highly developed in certain languages, e.g. in the South Siberian ones, which seldom use simple verb forms. Khalaj displays a number of remarkable imperative forms going back to postverb constructions. A few postverb constructions have developed further into aspecto-temporal categories, e.g. Turkish *Gele yorir* ('runs coming') > *Geliyor* '(S)he/it is coming, comes' (see p. 115).

Versatility of Thematic Suffixes

Turkic thematic suffixes, expressing aspectual, modal and temporal notions, are, particularly in older languages, rather versatile in syntactic respect. They are often not strictly bound to only one of the three functions of marking finite sentences, complement clauses and relative clauses. There are even trifunctional items, e.g. *-GAN*, which can mark all these predication types, functioning as finite form, action noun and participle, e.g. *Kelgen* '(S)he/it has come', *kelgen kişi* 'a/the person who has come', *kelgeni* 'his/her/its having come'.

Many modern languages have developed more differentiated systems. In Turkish, *-mlš* cannot function as an action noun, and *-DİK* not as a finite form. The prospective marker *-(y)AcAK*, however, is trifunctional: *Gelecek* '(S)he/it will, shall come', *gelecek* 'upcoming, to come', *geleceği* 'his/her/its anticipated or required coming'.

Aspect Formation

Aspect categories signalling different perspectives on events are basic to the Turkic verb. As already noted, certain postverb constructions have developed into aspectual markers.

Both intraterminal items, i.e. presents and imperfects, and post-terminal items, e.g. perfects, have shown permanent tendencies to become less 'focal', to express a less narrow perspective on the event. This means that they have come to focus less on the intraterminal or post-terminal state obtaining at some orientation point, as is typical of progressives ('is going') and stative or resultative forms ('is gone'). There have therefore been repeated renewals of focal intraterminality by new 'progressive' presents and imperfects, and of focal post-terminality by new stative and resultative items.

Renewal of Intraterminality

The Old Turkic 'aorist' present in *-(V)r* and the corresponding imperfect in *-(V)r erdi* were general intraterminals. Their focality was weakened, and in

most later languages, new, highly focal periphrases involving postverb auxiliaries meaning 'stand', 'lie', 'sit' or 'go' were introduced (see p. 43). The general pattern in NW, SE and Chuvash is the originally highly focal *Yaza turur* ('stands writing') '(S)he is writing'. The shape of *turur* has been reduced to *-dUr*, *-dl* or \emptyset , e.g. Kazakh *Keledi* '(S)he comes', Uzbek *Išlaydi* '(S)he works', Chuvash *Pirađap* 'I go', Kumyk *Bara* '(S)he goes', Uyghur *Oquymen* 'I am reading'. Corresponding imperfects exhibit the pattern *-A tur-* + simple past, e.g. Uyghur *Kirettim* 'I used to enter'. Some languages prefer the pattern *Yazip turur*, e.g. Tuvan *Körüp tur* '(S)he sees', *Ap tur* '(S)he takes'.

Oghuz languages prefer a postverb meaning 'move' and the pattern *Yaza yorir* ('moves writing' > 'is writing' > 'writes'), e.g. Turkish *-(Ø)Iyor* in *Yazıyor*, Azerbaijani *-(y)Ir in Yazır*, Turkmen *-yAr in Yađyār* '(S)he writes'. The corresponding imperfects are *Yazıyordu* '(S)he was writing', etc. Some languages prefer the postverb *yat-* 'lie', e.g. Khakas *parča* 'goes'. There are corresponding imperfects such as Altay Turkic *Brätti* '(S)he used to go'.

Some languages have produced a second renewal of focal intraterminality with the periphrasis *Yaza tura[turur]*, i.e. *-A* converb + present tense of *tur-*, e.g. Kumyk *Oxuy turaman* 'I am reading', Karachay *Išley turadi* '(S)he is working', *Ala turadi* '(S)he is taking'. A few languages prefer the pattern *Yazip tura[turur]*, e.g. Kirghiz *Oqup turat* '(S)he is reading'. Other languages prefer the auxiliary *yat-* 'lie' and the pattern *Yazip yata[turur]*, e.g. Uyghur *Yeziwatimen* 'I am writing'. Corresponding imperfects follow the pattern *yazip yata tur-* + simple past, e.g. Uyghur *Kiriwatattıq* 'We were entering'. The pattern *Yaza yatip[turur]* is found in Uzbek, e.g. *Išlayapmān* 'I am working'. Other focal forms are built with *yat-* 'lie', e.g. Uzbek *Kelāyātirmān* 'I am coming', Altay Turkic *Qičirip-d'attim* 'I was reading'.

Renewal of Post-terminality

Simple post-terminality markers include *-mlš* and *-GAN*, e.g. Azerbaijani *Almišam* 'I have taken'. Focal post-terminality has been renewed, in certain languages, by means of the periphrasis *Yazip turur* ('stands having written') '(S)he is in the state of having written'. Again, the shape of *turur* has been reduced to *-dUr*, *-dl* or \emptyset , e.g. Turkmen *Alipđir*, Azerbaijani *Alib* '(S)he has taken'. This form has later on developed into a perfect ('has written'), or into an indirective past ('wrote, has written' + 'obviously, as it turns out').

A few languages have produced a second renewal of focal post-terminality by means of the periphrases *Yazip tura[turur]*, i.e. *-(V)B* converb + present tense of *tur-*, e.g. Karachay *Ketib turadi* '(S)he has gone', Kumyk *Gelip tura* '(S)he has come'.

Apparent Ambiguity and Contractions

A high number of verb forms have emerged as a result of all possible postverb constructions of actional and aspectual nature. While some forms may, on the surface, seem to be ambiguous between actionality and aspecto-temporality,

they are made less ambiguous by different accent contours, e.g. Noghay *Kelē turādī* '(S)he comes' (durative actionality + general present tense) vs. *Kele turādī* '(S)he is coming' (high-focal present tense). Many forms are rather complex, but many have also undergone reduction processes. This has in some cases yielded results that look untypical of the agglutinative type, e.g. contractions such as Uzbek *Qilyāppān* 'I am doing' < **qil-a yatip [turur] men*.

Aorist

The so-called 'aorist' in *-(y)Vr*, the first known intraterminal form of Turkic, functions as the only present tense in Old Turkic, as it still does in Yakut. It is probably a combination of the old converb ending in *-(y)V* and a form of the old copula verb *er-* 'be'. In Old Turkic, the suffix variant *-yUr* is preserved after stem-final vowels, e.g. *bāšlāyur* 'lead', but it was later on contracted, e.g. *bāšlāyur* > *bašlār* > *bašlar*. There had probably been a similar development after originally short stem-final vowels, forms such as **atāyur* and **kelīyur* being contracted to *atar* 'throws', *kelir* 'comes', etc.

Turkish has preserved the old variation of high and low suffix vowels, e.g. *bilir* 'knows', *gelir* 'comes', *durur* 'stands', *verir* 'gives', *okur* 'reads', *güler* 'laughs', *atar* 'throws'. Turkmen has unified the marker to *-A(:)r*, e.g. *biler*, *geler*, *durur* (written *durar*), *berer*, *oqār*, *gölör* (written *güler*), *atar*. Azerbaijanian has a correspondingly unified suffix *-(y)Ar*, which differs from the present tense *-(y)Ir* only by the height of the vowel, e.g. aorist *biler*, *geler* vs. present tense *bilir*, *gelir*.

Aorists which have lost their focality mostly express modal meanings of disposition and prospectivity, e.g. Noghay *Berer* '(S)he will, may (is inclined to) give'. In Chuvash, the old aorist has developed into the so-called future in *-(θ)Ā* (see p. 445).

Future Markers

There are no special future markers in older Turkic. Prospective functions are observed in the use of optative markers such as *-GA(y)* and necessitative markers such as *-(y)AsI*, e.g. Mamlūk Kipchak *Kelesimen* 'I shall come'. A more explicit future marker is the Ottoman form *-(y)IsAr*, originally an aorist of the desiderative stem in *-(V)GsA-* 'want to (do)', e.g. *Geliser* '(S)he will come'. The type *-(y)AjAK* emerges rather late in Ottoman Turkish. It is also present in Azerbaijanian and has spread further in somewhat modified forms, e.g. Turkmen, Oghuz Uzbek *-JAK*, Kumyk *-(A)žAK*, Noghay *-AyAK*, Tatar *-(y)AčAK*, Bashkir *-(y)AsAK*, Uzbek *-(y)ājāk*.

Participles

The Old Turkic participle suffix *-miš* has its modern equivalents in West Oghuz *-miš* and Yakut *-BIT*. In other languages, this type has been replaced by *-GAn*, which also covers a good deal of the functions of the West Oghuz suffix *-DIK*. The suffix *-GAn* is represented by Turkish, Azerbaijanian,

Gagauz *-(y)An*, Turkmen *-A(:)n*, e.g. Turkmen *bilen* 'knowing, having known', *bilmān* 'not knowing, not having known'. A similar form is found in Khalaj, e.g. *kelen* 'coming', and in the unproductive Chuvash type *-An*, e.g. *tāvan kun* 'birthday'.

There is also an intraterminal, often habitual, participle suffix type that contains a vowel preceding *-GAn*. It includes the Chuvash present participle suffix *-(A)GAn*, e.g. *kilegen* 'coming', Khalaj *-AKAn*, *-AGAn*, the Karachay-Balkar and Kumyk habitual participle suffixes *-(A)GAn*, e.g. *alayan* '(usually) taking' and also Khakas *-IḡAn*, e.g. *kiligen* 'coming'. The Turkmen intraterminal participle is *-yĀn*, e.g. *bilyān* 'knowing', *bilmeyān* 'not knowing'. Turkish only has remnants of this participle, intensive adjectives in *-(A)GAn*, with *-GAn* after polysyllabic and *-AGAn* after monosyllabic stems, e.g. *çalışkan* 'diligent', *çekingen* 'shy', *olağan* 'usual'. All these forms seem to go back to **-A ergen*, which would correspond functionally to the periphrastic type **-A turyan* found in some other languages, e.g. Tatar *uqıy toryan*, Uzbek *oqiydiḡan* 'reading'.

Converbs

Converb markers have mostly been monofunctional in syntactic respect, only forming verbal adverbs. Some alleged exceptions have proved false or dubious: Yakut *-A(:)n* does not go back to *-GAn*, but to *-UbAn*, with intervocalic loss of the labial consonant. Tuvan *-vİšĀn* does not go back to *-miš* + *-GAn*, but to a postverb construction *-(V)p + it-* and a similar adverbial element. The Khalaj converb marker *-di*, which corresponds functionally to *-(V)B* in other Turkish languages, differs from the simple past marker with respect to accent, e.g. *āldi* 'taking, having taken'.

Converb markers are present in the oldest Turkic texts known to us. The subsequent history of Turkic has seen an increasing number of such markers, simple and complex, expressing manifold relational ideas of aspect, temporal sequence, manner, circumstance, etc. Numerous converbs are combinations of thematic suffixes with adverbial cases and postpositions.

The old intraterminal converb marker *-(y)V*, originally with a broad allomorphic variation, *-A*, *-I*, *-yU*, etc., has later counterparts of a more unified shape, e.g. modern Turkish *-(y)A*. As for the marker *-(V)B*, West Oghuz displays uncontracted forms *-(y)Ip*, e.g. Turkish *başlayıp* 'starting', *okuyup* 'reading', *oynayıp* 'playing', whereas the others have *-(I)p*, e.g. Turkmen *başlap*, *oqop*, *oynop*.

Converb forms of individual verbs have played an important role in forming relators of various kinds. They have often been reanalysed and grammaticalised as postpositions and connectives, e.g. Turkish *gör-e* 'according to' ← *gör-* 'see', *olarak* 'being', 'as', 'in the quality of' ← *ol-* 'become, be', *diye* 'saying', 'that' ← *de-* 'say', a junctor marking complements of verbs of utterance, thinking, knowing, with causal, purposive and similar meanings (cf. p. 64).

Syntactic Change

The following section will be limited to a few remarks on contact-induced syntactic changes. The syntactic structure of Turkic has proved relatively open to foreign influence. This becomes particularly obvious in older and more recent kinds of translation literature with its copies of various non-Turkic structures. There is Indo-Iranian influence on Old Uyghur, Persian influence on Chaghatay and Ottoman, Slavic influence on Karaim and Gagauz, etc. In developing the new literary languages of the Soviet Union, attempts were made to create a syntax especially adequate for translations from Russian.

Features of a 'loan syntax' emerge through selective copying of structural properties, e.g. combinability patterns. According to Persian and Russian patterns, morphologically marked adjectives have been introduced as new distributional classes. All Turkic languages have used pro-verbs such as *et-*, *qil-*, *yap-* 'do' to incorporate copies of foreign verbs (see p. 38). Languages in the Iranian neighbourhood have copied Persian *izāfat* constructions for nominal compounds (see pp. 175, 200–1, 219, 244). Most languages have copied foreign relators, e.g. conjunctions, sometimes also clause types typical of such elements. Foreign prepositions and more complex prepositional expressions are easily copied, though they are often used as postpositions in the receiving language. The peripheral languages Karaim, Gagauz and Khalaj have gone particularly far with respect to foreign syntactic influence. Karaim has copied Slavic prepositions governing certain cases, and Khalaj uses copies of Persian prepositions together with native words, e.g. *bī sen* 'without you'. It can even connect native words by means of *izāfat* constructions.

Copying of Word Order

Combinatory patterns can be copied within the boundaries of words, as the changed order of possessive and plural suffixes in Chuvash demonstrates (see pp. 38, 437). There are also many examples of contact-induced word order, e.g. in Gagauz, which has developed into a SVO language. Such calques often copy the arrangement of constituents according to pragmatic principles of sentence perspective, beginning from structures which are possible, but more marked in Turkic.

Decreasing use of Non-finite Clauses

Due to Iranian and Slavic contact influence, the use of non-finite clause types such as converb clauses has been reduced in certain languages. They have partly been replaced by imitations of Indo-European subordinative constructions with free prepositive junctors, conjunctions. This syntactic influence, especially the Russian one, has often been regarded as beneficial, since Turkic converbs have been claimed to be semantically vague. Needless to say, this is a groundless contention.

As a result of particularly long and intensive contact influence, the use of converbs has diminished radically in some varieties, e.g. Afghanistan Uzbek

and Khalaj. The highly Iranised Kashkay dialect has even lost the *-(V)B* converb and all constructions typical of it, using copies of Persian patterns instead.

The old Turkic narrative sentence patterns with periodic chain sentences (see pp. 64–5) have been increasingly abandoned in many Turkic languages. The use of *-(V)B* converb clauses with a first actant different from that of the matrix clause has diminished. There has been a gradual assimilation of the dominant written registers to European ones, where syntactic embedding is not typically used to express series of narratively equal events. Under predominantly Russian and French influence, subordination has come to be used more in accordance with European textbuilding patterns. The native narrative styles have disappeared from the norm registers of the more Europeanised languages. The extensive periodic chain sentences of Ottoman Turkish have long been considered 'abusive' in modern Turkish.

Lexicon

The vocabulary of Turkic languages exhibits a basic lexical stock which is likely to have existed in the common proto-language. Later on, a certain differentiation of the vocabulary of the major branches of Turkic took place. In addition to the inherited items, numerous lexical items have been copied from non-Turkic languages. Altogether, the changes in lexical structure appear to have been more comprehensive than the changes at other levels of the language. In what follows, a few examples of specific features in the basic lexical fields will be given.

Differences in the Basic Stock

As an illustration of the lexical differentiation in the major branches, some examples of specific Oghuz and non-Oghuz, notably Kipchak, words within the basic lexical stock are given in Table 5.1 (p. 120). Certain differences are also due to loanwords, e.g. Oghuz *baq-* 'look' vs. *qara-* ← Mongolian, Oghuz *alin* 'forehead' vs. *maglay* ← Mongolian.

Due to linguistic contacts, certain languages exhibit both Oghuz and non-Oghuz variants. Many Turkmen words are of non-Oghuz origin, e.g. *erin* 'lip' vs. *dōdaq*, *bōri* 'wolf' vs. *gürt*, *tal* 'willow-tree' vs. *sövüt*. Crimean Tatar displays both Oghuz and Kipchak elements, e.g. *baq-*, *qara-* 'look'. Karaim has a rather pure Kipchak vocabulary, e.g. *quyaş* 'sun', *yoluq-* 'meet'. Khalaj possesses, in spite of many borrowings, a non-Oghuz basic stock with words such as *erin* 'lip', *bīri* 'wolf' and *kindik* 'navel'. Some typically Oghuz words have spread far outside their own branch. Thus, *aq* 'white' has a very wide distribution, but is not found in isolated languages such as Yakut (*ürüŋ*), Khalaj (*hürün*) and Chuvash (*šurā*).

SE languages possess many words that are lacking or infrequently used in other branches, e.g. Uyghur *tola* 'much'. There are a number of regional words, e.g. for 'sparrow' Uyghur *quşqač*, Altay Turkic *boro quşqaš*, Khakas

Table 5.1 Oghuz and non-Oghuz lexical differences

	<i>Oghuz</i>	<i>non-Oghuz</i>
'bite'	išir-	tišle-
'dog'	köpek	it
'door'	qapı	eşik
'felt'	keče	kiyiz
'find'	bul-	tap-
'forward'	ileri	burun, murun
'good'	eyi, iyi	yaqşı
'hand'	el	qol ('hand' + 'arm')
'hare'	tavšan	qoyan
'lip'	dudaq	erin
'much'	çoq	köp
'navel'	göbek	kindik
'other'	başqa	özge
'rest'	dinlen-	tin-
'return'	dön-	qayt-
'send'	yolla-	yiber-
'speak'	söyle-	sözle-
'sun'	güneş	quyaş
'willow-tree'	söyüt	tal
'win'	yen-	ut-
'wolf'	qurt	böri

χusχaçaχ, Oghuz *serče*, Chuvash *šerši*, Kazakh, Noghay, Karaim *torγay*, Kirghiz *tarāncī*, Tatar *čipčiq*, Uzbek *čumčuq*, Karakalpak *šimšiq*.

Lexical Copies

All known varieties of Turkic have liberally adopted lexical copies. The oldest texts known to us already exhibit borrowings from Indo-Iranian and Chinese. There are early copies from Tokharian and several Iranian languages. For the various contacts with Indo-European and Chinese, see Chapter 4.

Arabic and Persian loanwords are numerous in all Islamic-Turkic languages, notably in those of the older sedentary groups. Karakhanid was the first written language to be lexically influenced by the Islamic vocabulary. This lexicon has later on played a considerable role in the SW, NW and SE branches, but is, for evident cultural reasons, absent in NE. Arabic and Persian lexical elements have been abundant in literary languages such as Chaghatay and Ottoman. There is a particularly strong Iranian lexical influence on Uyghur, Uzbek, and Turkic varieties of Iran and Afghanistan.

Many Turkic languages have been subject to a considerable Mongolic impact. There are very early Turkic and Mongolic parallels which are assessed differently by Altaicists and other scholars. The former consider the old lexical similarities as a common heritage from the Altaic proto-language. Non-Altaicists mostly ascribe the similar features to lexical copying from

Turkic into Mongolic due to very early contacts. The rise of the Chingisid world empire in the thirteenth century led to a massive contact-induced Mongolian impact on many Turkic languages and to numerous lexical copies, especially in the varieties of nomadic groups. Centuries later, the Turkic dialects of southern Siberia were exposed to a new strong Mongolian political and cultural influence, which led to copying of a great amount of lexical items.

Many Turkic languages have developed under Russian dominance during the modern period, partly under bilingual conditions, and thus acquired numerous loanwords and loan translations. Naturally, the influence is deeper in those languages which had early contacts with Russian, e.g. Tatar and Bashkir. The lexical influence on the Turkic literary languages developed in the Soviet period has been particularly strong. The result of a profound Slavic lexical impact is found in languages such as Karaim and Gagauz.

The development of Chuvash has been rather complicated. Its basic lexical stock shows numerous correspondences with Common Turkic, and there are also some interesting Oghuz-Chuvash isoglosses. Later on, Chuvash has acquired loanwords from Finno-Ugric, from other Turkic languages, notably Tatar, and from Russian. It even displays a number of early Arabic and Persian loanwords. The origin of many elements is still unknown.

The influence of western European vocabulary has become very important in the modern period. As far as Turkish is concerned, adoption of Italian, French and other western loanwords began in Ottoman. In the languages of the Russian influence area, the modern vocabulary has been introduced through the mediation of Russian. The proportion of loanwords from English has so far been rather modest in all Turkic languages, but is currently increasing. The modern Turkic languages of China are strongly influenced by old and recent Chinese vocabulary.

Substratum Influence

Many originally non-Turkic-speaking groups, Iranian, Greek, Finno-Ugric, Samoyedic, Yeniseian, Tungusic and others, have exerted substratum influence on Turkic languages. Such cases represent a special kind of copying of a linguistic code. A certain portion of the old lexicon has mostly been imposed on the new Turkic mother tongue. For example, Armenian words in Armeno-Kipchak texts are imposed items of this kind, and not normal 'borrowings'.

An extreme form of lexical substratum influence is found in some 'hybrid' varieties of groups which have adopted the grammatical system of a Turkic language, but retain the lexicon of their original language. One such case is the language of the Eynu in Xinjiang, who have copied a predominantly Persian vocabulary into an Uyghur basic code. In south Central Asia and Anatolia, there are similar so-called Abdal groups of settled nomads, who combine a local Turkic morphosyntax with a basic lexicon of partly Persian and partly unknown origin.

Foreign Vocabulary and Puristic Reforms

The readiness of Turkic languages to adopt foreign lexical material has often been appreciated and even admired. The fifteenth century poet 'Alī Šēr Navā'ī considered Turkic superior to other languages, since it potentially also had room for their stocks of words. On the other hand, the dominance of non-Turkic elements in certain varieties has created problems and called for puristic measures. For instance, the overload of Persian and Arabic in literary Ottoman eventually caused strong reactions. For the puristic efforts to replace this unwanted vocabulary by so-called *Öztürkçe* (Pure Turkish), see Chapter 13. The resulting neologisms have partly reduced the mutual intelligibility with those Turkic languages that have not undergone corresponding reforms.

The lexical development of Turkic is currently, as the development in general, in a rather unsurveyable stage. Very little is known about social functions, socially conditioned lexical variation, standards and substandards in the new varieties emerging on the threshold of the twenty-first century.

References and Further Reading

- Adamović, M. (1985) *Konjugationsgeschichte der türkischen Sprache*, Leiden: Brill.
- Baskakov, N. A. (1969) *Vvedenie v izučenie tjurkskix jazykov*, second edition, Moscow: Vysšaja škola.
- (1975) *Istoriko-tipoložičeskaja xarakteristika struktury tjurkskix jazykov*, Moscow: Nauka.
- (ed.) (1966) *Jazyki narodov SSSR, 2, Tjurkskie jazyki*, Moscow: Nauka.
- Bazin, L. (1994) *Les Turcs: des mots, des hommes*, Budapest: Akadémiai Kiadó.
- Benzing, J. (1953) *Einführung in das Studium der altaischen Philologie und der Turkologie*, Wiesbaden: Harrassowitz.
- Berta, Á. (1989) *Lautgeschichte der tatarischen Dialekte*, Studia Uralo-Altaica 31, Szeged: University of Szeged.
- (1996) *Deverbale Wortbildung im Mittelkiptschakisch-Türkischen*, Turcologica 24, Wiesbaden: Harrassowitz.
- Brands, H. W. (1973) *Studien zum Wortbestand der Türksprachen. Lexikalische Differenzierung, Semasiologie, Sprachgeschichte*, Leiden: Brill.
- Clark, L. (1966) 'The early Turkic and Sarig Yugur counting systems', in R. E. Emmerick et al. (eds) *Turfan, Khotan und Dunhuang*, Berlin: Akademie-Verlag, 17–49.
- Clauson, Sir G. (1962) *Turkish and Mongolian Studies*, London: The Royal Asiatic Society.
- (1972) *An Etymological Dictionary of Pre-Thirteenth-Century Turkish*, Oxford: Oxford University Press.
- Dankoff, R. and Kelly, J. (eds) (1982–1985) *Mahmūd al-Kāšyārī: Compendium of the Turkic Dialects (Dīwān Luyāt at-Turk)*, 1–3, Sources of Oriental Languages and Literatures 7, Harvard University.
- Deny, J. et al. (eds) (1959) *Philologiae Turcicae Fundamenta 1*, Wiesbaden: Steiner.
- Doerfer, G. (1963–75) *Türkische und mongolische Elemente im Neupersischen*, 1–4, Akademie der Wissenschaften und der Literatur, Veröffentlichungen der Orientalischen Kommission 16, 19, 20, 21, Wiesbaden: Franz Steiner.
- (1971) *Khalaj Materials*, Indiana University Publications, Uralic and Altaic series 115, Bloomington: Indiana University, The Hague: Mouton.

- (1971) *Grammatik des Chaladsch*, Turcologica 4, Wiesbaden: Harrassowitz.
- (1975–6) 'Proto-Turkic: reconstruction problems', *Türk Dili Araştırmaları Yıllığı Belleten*, 1–50.
- (1990) 'Die Stellung des Osmanischen im Kreise des Oghusischen und seine Vorgeschichte', in G. Hazai (ed.) *Handbuch der türkischen Sprachwissenschaft*, 1, Budapest: Akadémiai Kiadó, 13–34. (With a classification different from the one given above.)
- and Tezcan, S. (1980) *Wörterbuch des Chaladsch (Dialekt von Charrab)*, Budapest: Akadémiai Kiadó.
- Ercilasun, A. B. (1984) *Kutadgu Bilig grameri. Fül*, Gazi Eğitim Fakültesi 3, Ankara.
- Erdal, M. (1991) *Old Turkic Word Formation. A Functional Approach to the Lexicon 1–2*, Turcologica 7, Wiesbaden: Harrassowitz.
- Gabain, A. von (1974) *Altürkische Grammatik*, third edition, Wiesbaden: Harrassowitz.
- Hazai, G. (1978) *Kurze Einführung in das Studium der türkischen Sprache*, Budapest: Akadémiai Kiadó.
- (1990) *Handbuch der türkischen Sprachwissenschaft*, 1, Wiesbaden: Harrassowitz, Budapest: Akadémiai Kiadó.
- Jarring, G. (1933) *Studien zu einer osttürkischen Lautlehre*, Lund: Borelius, Leipzig: Harrassowitz.
- Johanson, L. (1979) *Altürkisch als 'dissimilierende Sprache'*, Akademie der Wissenschaften und der Literatur Mainz, Abhandlungen der Geistes- und sozialwissenschaftlichen Klasse 1979: 2, Wiesbaden: Steiner.
- (1989) 'Aorist and present tense in West Oghuz Turkic', *Journal of Turkish Studies* 13: 99–105.
- (1990) 'Zu den Grundfragen einer kritischen Altaistik', *Wiener Zeitschrift für die Kunde des Morgenlandes* 80: 103–24.
- (1991) *Linguistische Beiträge zur Gesamturkologie*, Bibliotheca Orientalis Hungarica 37, Budapest: Akadémiai Kiadó.
- (1992a) *Strukturelle Faktoren in türkischen Sprachkontakten*, Sitzungsberichte der Wissenschaftlichen Gesellschaft an der Johann Wolfgang Goethe-Universität Frankfurt-am-Main 29: 5, Stuttgart: Franz Steiner. (English edition, London: Curzon.)
- (1992b) 'Zur Isochronie im Türkischen', in G. Bethlenfalvy et al. (eds) *Altaic Religious Beliefs and Practices*, Budapest: Hungarian Academy of Sciences and Eötvös Loránd University, 183–8.
- (1993) 'Code-copying in immigrant Turkish', in G. Extra and L. Verhoeven (eds) *Immigrant Languages in Europe*, Clevedon: Multilingual Matters, 197–221.
- (1994) 'Funktion, Kompetenz und Etymon. Bemerkungen zu einer ostaltürkischen Wortbildungslehre', *Central Asiatic Journal* 38: 160–78.
- (1995a) 'Wie entsteht ein türkisches Wort?', in B. Kellner and M. Stachowski (eds) *Laut- und Wortgeschichte der Türksprachen*, Turcologica 26, Wiesbaden: Harrassowitz, 97–121.
- (1995b) 'Mehrdeutigkeit in der türkischen Verbalkomposition', in M. Erdal and S. Tezcan (eds) *Beläk Bitig*, Turcologica 23, Wiesbaden: Harrassowitz, 81–101.
- and Schönig, C. (eds) (1988) *Kritische Beiträge zur Altaistik und Turkologie. Festschrift für Johannes Benzing*, Turcologica 3, Wiesbaden: Harrassowitz.
- Korkmaz, Z. (1995) *Türk dili üzerine araştırmalar 1–2*, Türk Dil Kurumu Yayınları 629, Ankara: Türk Dil Kurumu.
- Kormuşin, I. V. (1984) *Sistemy vremen glagola v altaikskix jazykax*, Moscow: Nauka.

- Ligeti, L. (1971) *Studia Turcica*, Bibliotheca Orientalis Hungarica 17, Budapest: Akadémiai Kiadó.
- Menges, K. H. (1995) *The Turkic Languages and Peoples. An Introduction to Turkic Studies*, second edition, Wiesbaden: Harrassowitz.
- Miller, R. A. (1971) *Japanese and the Other Altaic Languages*, Chicago, London: The University of Chicago Press.
- (1996) *Languages and History. Japanese, Korean, and Altaic*, Bangkok: White Orchid, Oslo: The Institute for Comparative Research in Human Culture.
- Musaev, K. M. (1984) *Leksikologija tjurkskix jazykov*, Moscow: Nauka.
- Nasilov, D. N. (1989) *Problemy tjurkskoj aspektologii. Akcional' nost'*, Leningrad: Nauka.
- Nauta, A. H. (1985) 'Lambdazismus im Tschuwassischen: Gtü. š = Tschuw. l und š', in G. Jarring and S. Rosén (eds) *Altaistic Studies*, Stockholm: Almqvist and Wiksell, 123–44.
- Poppe, N. (1960) *Vergleichende Grammatik der altaischen Sprachen*, 1, Wiesbaden: Harrassowitz.
- (1965) *Introduction to Altaic Linguistics*, Ural-Altische Bibliothek 14, Wiesbaden: Harrassowitz.
- Pritsak, O. (1963) 'Das Altürkische', in B. Spuler (ed.) *Handbuch der Orientalistik*, 1, 5:1, Leiden. Cologne: Brill, 27–52.
- Ramstedt, G. J. (1952–1957) *Einführung in die altaische Sprachwissenschaft*, 1–2, Mémoires de la Société Finno-Ougrienne 104, Helsinki.
- Räsänen, M. (1969) *Versuch eines etymologischen Wörterbuchs der Türksprachen*, Helsinki: Suomalais-Ugrilainen Seura.
- Röhrborn, K. (1977–1988) *Uigurisches Wörterbuch. Sprachmaterial der vorislamischen türkischen Texte aus Zentralasien 1–4*, Wiesbaden: Steiner.
- Róna-Tas, A. (1991a) *Language and History. Contributions to Comparative Altaistics*, *Studia Uralo-Altaica* 25, Szeged: University of Szeged.
- (1991b) *An Introduction to Turkology*, *Studia Uralo-Altaica* 33, Szeged: University of Szeged.
- Ščerbak, A. M. (1970) *Sravnitel'naja fonetika tjurkskix jazykov*, Leningrad: Nauka.
- (1977) *Očerki po sravnitel'noj morfologii tjurkskix jazykov (Imja)*, Leningrad: Nauka.
- (1981) *Očerki po sravnitel'noj morfologii tjurkskix jazykov (Glagol)*, Leningrad: Nauka.
- (1987) *Očerki po sravnitel'noj morfologii tjurkskix jazykov (Narečie, služebnye časti reči, izobrazitel'nye slova)*, Leningrad: Nauka.
- (1994) *Vvedenie v sravnitel'noe izučenie tjurkskix jazykov*, St Petersburg: Nauka.
- Scherner, B. (1977) *Arabische und neupersische Lehnwörter im Tschuwaschischen*. Wiesbaden: Franz Steiner.
- Schönig, C. (1984) *Hilfsverben im Tatarischen. Untersuchungen zur Funktionsweise einiger Hilfsverbverbindungen*. Akademie der Wissenschaften und der Literatur, Mainz, Veröffentlichungen der Orientalischen Kommission 35, Wiesbaden: Franz Steiner.
- (1997) 'A new attempt to classify the Turkic languages', *Turkic Languages* 1: 117–33.
- Schulz, P. (1978) *Verbalnomina und Konverbien als adverbiale Ergänzungen im Altürkischen*, Ph.D. thesis, University of Giessen.
- Sertkaya, O. F. (1995) *Göktürk tarihinin meseleleri*, Ankara: Türk Kültürünü Araştırma Enstitüsü. (Articles in Turkish, English and German on Old Turkic language and history.)

- Sevortjan, É. V. (1974–) *Étimologičeskij slovar' tjurkskix jazykov*, Moscow: Nauka.
- Soper, J. D. (1987) *Loan Syntax in Turkic and Iranian: The Verb Systems of Tajik, Uzbek, and Qashqay*, Ph.D. thesis, University of California, Los Angeles.
- Tekin, T. (1968) *A Grammar of Orkhon Turkic*, Indiana University Publications, Uralic and Altaic Series 69, Bloomington: Indiana University, The Hague: Mouton.
- and Ölmez, M. (1995) *Türk dilleri. Les langues turques*, Türk Dilleri Araştırmaları Dizisi 20, Ankara: Simurg.
- Tenišev, É. R. (ed.) (1984) *Sravnitel'no-istoričeskaja grammatika tjurkskix jazykov. Fonetika*, Moscow: Nauka.
- (ed.) (1988) *Sravnitel'no-istoričeskaja grammatika tjurkskix jazykov. Morfoloģija*, Moscow: Nauka.
- Wurm, S. (1954) *Turkic Peoples of the USSR: Their Historical Background, their Languages and the Development of Soviet Linguistic Policy*, Oxford: Central Asian Research Centre.
- Yüce, N. (1973) *Gerundien im Türkischen. Eine morphologische und syntaktische Untersuchung*, Ph.D. thesis, University of Mainz.

6 *Turkic Writing Systems*

András Róna-Tas

Introduction

Turkic languages have been and are written with many different writing systems. Our discussion of these systems will be based on the following assumptions. A writing system consists of a set of graphic signs and a set of rules according to which the signs are used to represent sounds or combinations of sounds. The graphic signs mostly have a normalised form. All graphemes intended to represent this form will be called allographs. In alphabetic writing systems, a graphic sign with all its allographs will be called a letter. If the system consists of graphic signs which represent a consonant plus a vowel, we shall speak about a syllabic script. A script is hieroglyphic, if words are represented by icons or complex signs of which an icon is an organic part. As a rule, scripts of Semitic origin are alphabetic, whereas scripts of Indic origin are syllabic. The Chinese writing system is hieroglyphic. All writing systems have their constraints. Thus, if an old Turkic language has been written according to several different systems, the probability of a correct reconstruction of its sound system is greater. The ways of rendering Turkic words, names and titles may vary a good deal if they are written by non-Turks according to writing systems not used by the Turks themselves. The oldest monuments of Turkic are written according to such systems, e.g. the Chinese and Middle Greek systems. Though such transcriptions are highly important for reconstruction, we shall not deal with them here.

In the rest of this chapter, angular brackets <> will be used for graphemes, slash marks // for phonemes, and square brackets [] for allophones. Small capitals in angular brackets, e.g. <A>, denote graphemes representing more than one phoneme.

Systems Used for Old Turkic

Alphabets used by Turks for writing Old Turkic texts may be divided into two major groups: those of Semitic and those of Indic origin.

Systems of Semitic Origin

The earliest documents written with Aramaic script date from the nineteenth century BC. This script developed into several branches such as the Hebrew,

Palmyran, Syrian, Nabatean, Arabic and Mandaean ones. The scripts of ultimately Semitic origin that have been used to write Turkic are the Runic, the Soghdian, the reformed Soghdian or Uyghur, the Manichaean and the Arabic script. Though the Hebrew script was probably used for writing Turkic texts in the Khazar empire, no such texts are known so far. Later on, Hebrew was used to write Karaim. Finally, there is a carved type of script called the East European runiform script. Though it is not yet deciphered, it most probably belongs to the scripts of Semitic origin.

East Turkic Runiform Script

The East Turkic runiform script is also called 'runic' because, before its decipherment by Vilhelm Thomsen in 1893, some scholars thought it to be a variant of the Germanic runic script. The similarities are only due to the technique, both kinds of signs being usually carved on rocks. Some scholars have taken the alphabet to be an invention of the Turks and to consist of *tamya* signs (marks of ownership) and pictograms. It rather appears that this writing system is of Semitic origin, developed in at least four phases, and that some signs of pictographic origin were added to the script in one of the late phases.

The Turkic documents written in East Turkic runiform script are (1) monumental and smaller memorial inscriptions on steles and rocks, (2) short inscriptions on various metal or ceramic objects, (3) graffiti on rocks, or (4) texts written on paper or wood with a brush or calamus. The number of the known texts is close to 300. About 18 have been found in the Pribaikal–Lena region, about 150 in the Yenisey region, about 35 in Mongolia, about 16 in the Altay region, about 14 in East Turkestan, about 28 in northern Kirghizstan and Kazakhstan, and about 18 in Ferghana, Alay and northern Tokharistan. The earliest inscriptions are those found in Mongolia, of which the earliest can be dated to the 720s. The most recent texts, discovered in East Turkestan, may have come to existence in the tenth century at the latest.

Originally, the script seems to have had four vowel signs: <A> = /a/, /e/, <w> = /o/, /u/, <I> = /i/, /i/, /è/, <w̄> = /ö/, /ü/. In the Yenisey inscriptions, a fifth vowel sign was added: <è> = /è/. Letters denoting consonants usually have a back form, denoted here as c¹, and a front form, c². Such letters are <b¹>, <b²>, <d¹>, <d²>, <g¹>, <g²>, <y¹>, <y²>, <l¹>, <l²>, <n¹>, <n²>, <r¹>, <r²>, <t¹>, <t²>. Beside <k¹> and <k²>, there is a <k³>, used before and after /i/, a <k⁴>, used before and after /ö/ and /ü/, and a <k⁵>, used before and after /o/ and /u/. In the earliest documents, one letter <s> denoted all sibilants. Later on, a new letter <š> denoted [š¹], [š²] and [s¹]. As a further step, a new letter was introduced for [s¹] and, in the Yenisey inscriptions, for /š/. The letters <m>, <ŋ>, <p>, <č>, <z>, <ń> are used irrespective of front vs. back distinctions. Consonant clusters are represented by the three letters <ld>, <nd>, <nč> (but cf. p. 97). As a rule, word-final vowels are denoted, but rounded vowels are not, if the preceding syllable has a denoted rounded vowel. Closed /è/ is frequently written with <I>.

Soghdian Script

The Soghdian script belongs to the Aramaic group of Semitic scripts. The earliest specimens are the so-called Soghdian Ancient Letters, dated at AD 312–313 or earlier. Among the early varieties of this script, the Samarkand type became dominant. Later, around AD 500 a calligraphic variant developed in Turkestan. Since it was mostly used for Buddhist texts, it was also called the Sūtra or Formal Script. Along with this script, in which Manichaean, Nestorian Christian and secular texts were also written, a more cursive type evolved, whose final form was fixed around the middle of the seventh century. Since in this variety the letters had no recurrent normalised form and ligatures were common, some letters became indistinguishable from others. The Soghdian script was written from right to left. Around AD 600, perhaps under Chinese influence, a vertical variety appeared in which the lines went from left to right.

Originally, only long vowels were denoted, but later on, other vowels were also written, even reduced ones. The letters *gimel* and *hēth* became indistinguishable except in word-final position. Soghdian scribes sometimes used diacritic signs. Since the original *lāmedh* denoted a fricative /ð/, a *rēsh* with a hook was used for denoting /l/. This way of denoting /l/ was taken over by the Turks and later also by the Mongols. Since the shapes of final <z> and <n> coincided, <z> was provided with an additional stroke to denote both final /z/ and /ʒ/. This stroke was later used by the Turks to represent final /ʒ/. Additional dots on *bēth* and *pē* denoted the fricative /f/.

The Soghdian language and thus also the Soghdian script were used by the Turks from very early times. Texts in Soghdian contain many Turkic names, titles and even common words. The earliest specimens, from the sixth century onwards, are coins with Soghdian inscriptions containing Turkic names and titles. The rulers of the first Turkic Kaghanate used Soghdian as an official language. The Soghdian inscriptions of Bugut (581–587) and Mongolküre are inscriptions of the Turkic rulers. Later on, the Soghdian script was used to render Buddhist texts in Turkic.

Uyghur Script

Since Buddhism first came to the Turks by the mediation of Soghdian monks, it is natural that the Soghdian script was first used almost without any changes. Later on, the texts were revised orthographically, and a standardised reformed and adapted form of the Soghdian script developed gradually, the so-called Uyghur script. In daily usage, however, many 'Soghdisms' remained unchanged.

Word-initially, 'āleph generally renders an /e/, while two 'āleph signs denote an /a/. In early, 'pre-classical' texts, back /i/ is rendered by <a> or not at all, i.e. with so-called 'defective notation'. Front /i/ is written with <y>. Later texts show the gradual neutralisation of the opposition /i/ vs. /i/. Neutralised /i/ was phonetically closer to front /i/ and thus represented by

<y>. The <a> notation of /i/ in early manuscripts is frequent in certain phonetic environments. In texts of the pre-classical period, a so-called 'hypertrophic' <a> is written before <y>, <w> or <a> in some Turkic words.

One of the orthographic peculiarities of pre-classical texts is the rendering of /ŋ/. In word-final position, mainly in second-person imperative and possessive suffixes, but also in some lexemes, <k> is found instead of the expected combination <nk>. In a few words, Old Turkic /ŋ/ is rendered by <ny> or <n>. In pre-classical texts, the Buddhist terminology of ultimately Sanskrit origin is written according to the Soghdian and not the Tokharian usage.

The stabilisation of the orthography was parallel to the growing influence of Tokharian missionary activities and the appearance of Tokharian loanwords and Sanskrit words in Tokharian shapes. Later on, the orthography began to change. The diacritic dots that distinguished graphemes for velars were inconsistently used, sometimes for *q*, sometimes for *γ*. They also served to distinguish these sounds from <a> + <n>. The letters <d> and <t> were easily interchanged. More and more, position rather than phonetic conditions determined their choice: <t> is found syllable-finally, <d> syllable-initially and between vowels. The grapheme <z> gradually disappeared, <s> also being written for /z/.

The Uyghur script with this late orthography was taken over by the Mongols in the twelfth century. At the end of the Yüan dynasty, in the fourteenth century, the Mongolian orthography reinfluenced the Uyghur one.

Manichaean Script

Manichaean texts were first written in Aramaic, Syriac, Parthian and Middle Persian. A variety of the Aramaic script was used for translations. After the death of Mani, the founder of Manichaeism, his adherents fled to the Soghdian area, where the religion began to flourish. The so-called Manichaean script used by Soghdian Manichaeans in religious texts was a variety of the original script that had been adapted slightly to the Soghdian language. It was also used by Tokharians and Turks.

Nineteen runiform letters together with the corresponding syllables in Manichaean script are listed in a fragment found in the Turfan region. Two Persian fragments written in Manichaean script in the last years of the Uyghur steppe Kaghanate (see p. 21) contain a great quantity of Turkic names and titles. The first text, *Mahrnāmag* or 'The Book of Hymns', is dated 808–821.

Word-initial /a/ is rendered with one or two 'āleph signs, /e/ with one 'āleph, /i/, /i/ and /i/ with 'ayin + *yōdh*. The use of initial <y> in a few cases may reflect a prothetic *y*- or some other special usage, e.g. <yōdug> = (y)īduq 'sacred'. Word-initial /o/ and /u/ are written with 'āleph + *wāw*, word-initial /ö/ with 'āleph + *wāw* + *yōdh*. The only Turkic word with initial *ü*- occurring in the texts, the title *üge* 'counsellor', is written with 'āleph + *wāw* or only with *wāw*. In the Persian text, long vowels are written 'plene' (full), i.e. with

'*āleph*, *wāw* or *yōdh*. Thus /e/ is rendered by '*āleph* in non-first syllables, the only exception being <swg> = *Seveg*. With a few exceptions, /o/, /ö/, /u/ and /ū/ are written with *wāw*. The vowels /i/ and /ī/ are written by *yōdh*. Double *yōdh* signs are often found word-finally, maybe in order to fill the line.

Devoicing of /b/ before voiceless consonants is marked with <p>, e.g. <xwtp'yn> = *Xut Bayan*. The consonant /t/ is mostly written with *teth*, in a few cases with *tāw*. The Aramaic *lāmedh*, transliterated <δ>, denotes an interdental fricative, while <d> generally denotes a stop. The phoneme /l/ is always written with a big diacritic hook. There are two signs for /g/, *gimel* in front environments, and its variant with a hook on its lower end in back environments. Later on, the hooked *gimel* represented a fricative. Thus, special graphotactic devices were necessary to represent a back stop, e.g. <ny>. Voiceless palatovelar stops and fricatives could be rendered by four different letters. A simple *kaph* is normally used for front [k], a *kaph* with one diacritic dot for [x]. The undotted *qōph*, which has been transliterated as <q>, simply seems to denote a stop. Notations such as <qwlycwr> = *Kōličur*, <qwylwg> = *Külüğ* or <tqyn> = *tegin* (beside <tkyn>) support this assumption. There is also a *qōph* with two superscribed dots. Two letters, *sāmekh* <s> and *šin* <š>, are used for unvoiced sibilants; the *Mahrnāmag*, however, only exhibits <s>. The letter *zayin* <z> stands for /z/, and a variant of it, <z>, renders [j], the voiced equivalent of /č/, e.g. <xwnzwy> = *Xunjuj*. The letter *šadhē* denotes /š/.

Though the Uyghurs adopted Manichaeism officially around 762, Turkic texts in Manichaean script, found in Turfan and Dunhuang, do not seem to have been written earlier than the ninth to eleventh centuries. In such texts vowels are, with one exception, rendered the same way as in the Soghdian–Uyghur writing system. Word-initially, '*ayin*, and not '*āleph*, is written before *yōdh*. Thus, word-initial /i/, /ī/ and /e/ are usually written '*ayin* + *yōdh*, though some texts also have '*āleph* + *yōdh*. In the latest texts, notations with '*āleph* are predominant. Word-initial /a/ is written with one or two '*āleph* signs. The doubling of vowel signs in a few cases does not indicate vowel length.

The front stop [k] is written with a double-dotted *kaph* or with *qōph* without dots. The back stop [q] is written with a double-dotted *kaph* or a double-dotted *qōph*. The fricative [χ] is represented by a single-dotted *kaph* or a single-dotted *qōph*. The dots are often omitted in carelessly written texts. In others, irrespective of the choice of *kaph* or *qōph*, 'no dots' generally = front stop, 'two dots' = back stop, and 'one dot' = fricative.

The nasal /ŋ/ is mostly written <ng>, in a few cases <nng>, e.g. <mnyng> = *m(e)ning* 'mine'. The phoneme /d/ is mostly written with *lāmedh* = <δ>, whereas *dāleth* is only used in a very few cases after /n/. It seems that /d/ was mostly fricative and/or unvoiced, but realised as a stop after /n/. Double *lāmedh* may have been used to mark the stop, e.g. <'ytδōyy> = *itti* (cf. the Brāhmī system).

In a few texts, Turkic /š/ is written with *sāmekh* <s>, while *šin* <š> is used

in words of Middle Persian origin. It is still unclear whether this reflects a dialectal feature. The other consonantal graphemes are used the same way as in Persian texts.

Arabic Script

With the Muslim faith, the Arabic language as the language of the Koran and the Arabic script was introduced among the Turks. From the eighth century onwards, Turkic names and titles occur in works of authors writing in Arabic. From the eleventh century onwards, the Arabic script was used for rendering Turkic texts. This usage was first confined to the Karakhanid Empire and documented in Yūsuf of Balasayun's *Qutaḍyu bilig* 'Wisdom that brings good fortune' and Maḥmūd al-Kāšyarī's *Dīwān luyāti'l-Turk* 'Compendium of the Turkic Dialects'. Kāšyarī's own account of how he used the Arabic script for Turkic texts is of great help for the evaluation of the system used in the Karakhanid Empire. Later, a generally accepted system of writing Turkic with Arabic letters with many local variants developed (see, e.g., pp. 158–60).

The Arabic writing system uses the vowel graphemes '*alif*, *wāw* and *yā* and the diacritic signs *fatha*, *kasra* and *damma*. Kāšyarī distinguishes front vowels, called *išmām*, from back vowels, called *išbā*'. This distinction is limited to '*alif* and *wāw* and does not concern *yā*, which means that the opposition /i/ vs. /ī/ had disappeared or was disappearing. While /e/ is mostly rendered by *yā* or *kasra*, <a>-notations are found in a few cases where, for example, the first vowel in *keče* 'evening' is written with *fatha*, denoting an open *e* (cf. Chuvash *kaś*). Plene (full) notation of vowels in Kāšyarī does not only indicate etymological length of Turkic vowels, but also prosodic length according to the system of '*arūd* versification. Stress also plays a role. Word-final vowels are always written plene (full), which reflects phonetic, not phonological length.

Kāšyarī also mentions secondary letters, written as the primary ones but with additional diacritic dots. Some are never used in his manuscript. Sometimes a phonetic explanation such as 'hard' = unvoiced or 'thin' = voiced is added. In Kāšyarī's material, most words with /d/ are written both with <d> and with <ḍ>, e.g. in *bodun* 'people'. The sign <ḍ> differs from <d> by an additional dot. It is uncertain whether the late copyist of AD 1266 added or omitted dots in the manuscript. Since Kāšyarī himself repeatedly tries to make the distinction, it must be concluded that, at his time, /d/ had plosive and fricative allophones.

Systems of Indic Origin

As Buddhism came to Tibet via Central Asia, the Indic script first reached the peoples living there, the Tokharians and the Sakas. In Central Asia, the original Sanskrit script changed gradually and was adapted to the needs of the local languages. It has syllabic letters, i.e. most of them have a so-called

'inherent' vowel /a/ and are, unless otherwise marked, pronounced as open syllables in *-a*.

Brāhmī Script

Northwestern or Tokharian Brāhmī Script

The Tokharians lived in the northwestern part of Turkestan and spoke a non-Iranian Indo-European language. Tokharian A or East Tokharian was spoken in the oases of Turfan and Karašahr, while Tokharian B or West Tokharian was spoken around Kucha. Turkic texts written in Brāhmī script date from between the ninth and thirteenth centuries.

The basic unit of the Sanskrit script was the akṣara, a complex consisting of a consonant and a monophthong or diphthong vowel. In Tokharian Brāhmī, no vowel sign was added if the vowel was /a/ or /e/. If not, the vowel was marked with the diacritic signs <i>, <e>, <ai>, <o>, <ä> above the akṣara or by diacritics for <u>, <r>, <l> below the akṣara. If the vowel was long, a diacritic sign was added to the right of the consonant sign. Through fusion of <l> with the length sign, a new sign <ī> for /ī/ emerged. The syllable type vowel + consonant was expressed by independent vowel signs followed by a consonant. The independent vowel signs were <a>, <i>, <u>, <e>, <o>, <ai>, <au>, <ṛ>, <ḷ>.

To express word-initial /e/, the <a> sign was provided with two diacritic dots. This usage led to the practice of equipping the independent <a> with other diacritic vowel signs. Thus, initial /o/ could be rendered by the independent letter <o> or by the independent letter <a> with an added diacritic <o>. This graphotactic device gradually spread in Turkestan, /u/ being the last vowel for which the independent sign was used. Front vowels were rendered according to the Soghdian–Uyghur model. To represent the rounded front vowels, the independent letters <o> and <u> were provided with a subscribed letter <ya> with the value of <y>. The combinations <oya> and <uya> were thus read as /ö/ and /ü/, respectively. If the front vowel occurred after a consonant, the consonant letter or cluster got the diacritics in question.

In Brāhmī script, /e/ could be written with the independent <e>, with the diacritic <e> signs or with an independent or 'inherent' /a/ sign modified by the subscript <ya>. The <e> and <aya> notations were later contaminated.

All I-sounds are written with an independent or diacritic <i> sign, e.g. <i-lim-tyu-rtyā-či-lyā-r> = *ilintürtečiler* 'those who let themselves be caught'. Front /i/ is never written with <iya>. This may mean that the neutralisation of the opposition /i/ vs. /i/ was almost finished (cf. pp. 89–90). The last remnants of /i/ are found after /r/, /l/ and before the velar stop, always in back environments.

Though long vowels can be represented by a diacritic sign in the Brāhmī script, the etymological length of Turkic vowels is not marked. There is a stronger tendency to mark length in open than in closed syllables. The length

marker was probably used to mark prosodic length.

The Indic scripts are very rich in consonant signs. For example, <p>, <ph>, , <bh> are used for labial stops. In the Tokharian variant, additional signs render word-final consonants lacking an 'inherent' vowel – in the so-called virāma position – and variants with an 'inherent' /e/ instead of /a/. The virāma position can also be rendered by special letters.

The use of the graphic signs of these series to render Turkic consonants was by no means random. For example, the spirantisation of /k/ is expressed in a regular way. The spirantisation of /d/ is more problematic and deserves further investigation. One special Brāhmī letter, transcribed as <δ>, is only used for Turkic. While /z/ is written with the new letter <z> and sometimes with <s>, it is never rendered by <ys> according to Khotanese usage.

Southern or Khotanese Brāhmī

After the collapse of the Uyghur Kaghanate in AD 840, Turks came into closer contact with the Khotan region, its Saka inhabitants and the Southern Brāhmī script used by them. The Muslim Turk conquest shortly after 1000 put an end to the Buddhist culture in the Khotan oasis. No Turkic texts written in the Southern Brāhmī script are known so far, but there are official documents containing Turkic personal names, titles and ethnonyms. There is also a Turkic-Khotanese word list with Turkic words written in the Southern Brāhmī script.

The language reflected by these Khotanese glosses and transcriptions shows almost the same dialect as the one described by Kāšgarī a hundred years later. Turkic /a/ is rendered by the independent <a> or by a sign with an 'inherent' [a]. Since E-sounds are written with <e> in first syllables and <a> in non-first syllables, there is no graphic distinction between /è/ and /e/. Turkic /o/ is written with <ā>, because the Khotanese /ā/ had developed into [o]. Of the other Turkic vowels, /ö/ is mostly written with <au>, /u/ with <ū>, /i/ always with <ī>, and /i/ with <i> or <ai>. Vowel length is not marked.

The letters , <g> and <d> denoted the Khotanese fricatives developed from the original stops /b/, /g/ and /d/. Though the stops disappeared between vowels in Late Khotanese, they were still rendered by , <g> and <d> in the script, which caused great confusion. Since /t/ had become voiced, an unvoiced [t] was written <tt>. The stop [g] was written <gg>. The cerebral sign <d> denoted a Khotanese retroflex /d/, but was used for Turkic /l/ before /i/. Turkic /z/ was rendered by <ys>, whereas /s/ and /š/ were written <ss> and <śś>, respectively, in late texts also <s> and <ś>. The letter <h> with a colon-like diacritic sign and mostly a hook-like subscript diacritic denoted front and back /g/, which may point to a voiced fricative pronunciation. Turkic /k/ is written <k> in both front and back environments.

Tibetan Script

The Old Tibetan kingdom was founded in the middle of the seventh century, and the Tibetan writing system dates from that period. Old Tibetan sources

contain many Turkic proper names. One known Uyghur text, translated very early into Tibetan, contains valuable material for the transcription of Turkic ethnic names. Besides, seven text fragments written in Turkic in Tibetan script are known, the most important one being a Buddhist catechism.

The Tibetan alphabet is of Indic origin, and the models for it reached the Tibetans via Central Asia. Since Buddhist centres of Turkestan played an important role in the development of the Buddhist literature in Tibet, it is not surprising that some of their graphotactic devices were also applied. However, the orthographic rules had to be adapted to the structure of the Tibetan language, which has a special monosyllabic character.

The Tibetan script used one independent sign for vowels, the so-called 'great a' (*a-chen*) <">. Modifications of the 'inherent' vowel /a/ in consonant letters were possible with the four diacritic letters <e>, <i>, <o> and <u>. Word-initial vowels were written with <"> and modified with <i>, <o> and <u>. The open /e/ was rendered by adding a subscript <ya> to <"> or to the consonant sign, whereas the sign <e> was only used occasionally. The two *i*-signs, originally used for short and long [i] in the Indic alphabet, were also used to transcribe Turkic, but, as in Old Tibetan, the difference between them is only graphic and does not have any phonetic value. The phonemes /ö/ and /ü/ were rendered by <o> and <u> plus a subscribed <ya>. The high unrounded back vowel was rendered by <e>, <a>, <y>, with or without a subscribed <ya>, and /i/. The varying transcriptions seem to reflect a situation where the distinction between [i] and [i̥] sounds begins to disappear, though [i̥] is still preserved in certain phonetic environments. The Tibetan script used a subscribed <"> sign to transcribe long Sanskrit vowels. While this sign is also used in transcriptions of Turkic, it does not indicate etymological length.

The voiced nature of a Tibetan syllable-initial consonant was marked with a 'small a' (*a-chung*) <'> in front of the consonant sign. Turkic word-initial /t/ and /k/ were normally rendered by <d> and <g>, e.g. <dag> = *teg* 'like'. They were, however, represented by signs for aspirated variants if the following consonant was a tense unvoiced stop or a sibilant, e.g. <thu-kāl> = *tükel* 'complete'. This seems to be due to a kind of dissimilation. The grapheme <z> denoted a voiceless sound. A voiced pronunciation was marked with a preposed <'> or another preradical letter.

According to Tibetan graphotactic rules, only <g>, <d>, , <l>, <s>, <r>, <m>, <n>, <ŋ> and <'> were allowed to appear in syllable-final positions. Thus, a syllable-final /z/ could not be represented adequately.

Systems Used for Middle Turkic

Nestorian Script

The Christian faith reached the Turks in its Nestorian form. The followers of Nestorius (d. c. AD 451) used a variant of the Syriac Estrangelo script. Most

Turkic texts are tomb inscriptions dating from the thirteenth and fourteenth centuries. Paper fragments from Karakhoto are also known.

The texts display a few peculiarities. Thus, /ö/ is written 'āleph + wāw + yōdh, only *tēth* is used for /t/, [q] is rendered by *qōph* with a diacritic dot, [k] by 'ayin with a diacritic dot, /č/ by <s>, and /d/ by *lāmedh*, not *dāleth*.

Phags-pa Script

The Phags-pa script was composed by the lama Phags-pa under the Yüan Emperor Kubilai to render all languages of the empire. It is a variant of the Tibetan script, used for seals and with the lines running vertically. The Phags-pa orthography is influenced by the Uyghur script. Only a few Turkic words written in this script are known.

Armenian Script

Members of the Armenian diaspora in the Crimea came into contact with Kipchak Turks in the fourteenth century. Some Kipchak groups became Christians and used the Armenian script for writing ecclesiastical and secular texts. After the great migration of the Armenians to Asia Minor, some Ottoman groups also adopted the Armenian script. Orthography and graphotactics differ in many details, e.g. front and back variants of /k/ are, as a rule, distinguished in Armeno-Ottoman but not in Armeno-Kipchak (see pp. 158–9).

Arabic Script

The Arabic script was used in all the various centres of the Islamic Turks. After the Persians had joined the Islamic world, a special variant of the Arabic script, adapted to the needs of the Persian language, was designed. The most important changes include the use of *jīm* with three dots for /č/ and *bēth* with three dots for /p/.

Eventually, variants of the Arabic orthography used for Turkic developed. The most important regional variants were the Khwarezmian, Chaghatay, Volga Turkic, Mamlūk and Ottoman ones. Some orthographic conventions became widespread and were used in most of the regions. There was thus a tendency to use emphatic Arabic letters to render consonants in back environments. Even in texts which otherwise use the three-dotted *bēth*, word-final /p/ is rendered by *bēth* with one dot, mainly in suffixes. Plene (full) notations of vowels with *wāw*, *yā* or 'alif do not reflect etymological length (cf. p. 90). The so-called *siyāqat* script was a special script used in the Ottoman administration for treasury accounts etc.

Systems Used for Modern Turkic Languages

Arabic Script

In modern times, the use and orthography of the Arabic script have undergone several reforms.

contain many Turkic proper names. One known Uyghur text, translated very early into Tibetan, contains valuable material for the transcription of Turkic ethnic names. Besides, seven text fragments written in Turkic in Tibetan script are known, the most important one being a Buddhist catechism.

The Tibetan alphabet is of Indic origin, and the models for it reached the Tibetans via Central Asia. Since Buddhist centres of Turkestan played an important role in the development of the Buddhist literature in Tibet, it is not surprising that some of their graphotactic devices were also applied. However, the orthographic rules had to be adapted to the structure of the Tibetan language, which has a special monosyllabic character.

The Tibetan script used one independent sign for vowels, the so-called 'great a' (*a-chen*) <ʾ>. Modifications of the 'inherent' vowel /a/ in consonant letters were possible with the four diacritic letters <e>, <i>, <o> and <u>. Word-initial vowels were written with <ʾ> and modified with <i>, <o> and <u>. The open /e/ was rendered by adding a subscript <ya> to <ʾ> or to the consonant sign, whereas the sign <e> was only used occasionally. The two *i*-signs, originally used for short and long [i] in the Indic alphabet, were also used to transcribe Turkic, but, as in Old Tibetan, the difference between them is only graphic and does not have any phonetic value. The phonemes /ō/ and /ü/ were rendered by <o> and <u> plus a subscribed <ya>. The high unrounded back vowel was rendered by <e>, <a>, <y>, with or without a subscribed <ya>, and /i/. The varying transcriptions seem to reflect a situation where the distinction between [i] and [ī] sounds begins to disappear, though [ī] is still preserved in certain phonetic environments. The Tibetan script used a subscribed <ʾ> sign to transcribe long Sanskrit vowels. While this sign is also used in transcriptions of Turkic, it does not indicate etymological length.

The voiced nature of a Tibetan syllable-initial consonant was marked with a 'small a' (*a-chung*) <ʿ> in front of the consonant sign. Turkic word-initial /t/ and /k/ were normally rendered by <d> and <g>, e.g. <dag> = *teg* 'like'. They were, however, represented by signs for aspirated variants if the following consonant was a tense unvoiced stop or a sibilant, e.g. <thu-kāl> = *tükel* 'complete'. This seems to be due to a kind of dissimilation. The grapheme <z> denoted a voiceless sound. A voiced pronunciation was marked with a preposed <ʿ> or another preradical letter.

According to Tibetan graphotactic rules, only <g>, <d>, , <l>, <s>, <r>, <m>, <n>, <ŋ> and <ʿ> were allowed to appear in syllable-final positions. Thus, a syllable-final /z/ could not be represented adequately.

Systems Used for Middle Turkic

Nestorian Script

The Christian faith reached the Turks in its Nestorian form. The followers of Nestorius (d. c. AD 451) used a variant of the Syriac Estrangelo script. Most

Turkic texts are tomb inscriptions dating from the thirteenth and fourteenth centuries. Paper fragments from Karakhoto are also known.

The texts display a few peculiarities. Thus, /ō/ is written 'āleph + wāw + yōdh, only *tēth* is used for /t/, [q] is rendered by *qōph* with a diacritic dot, [k] by 'ayin with a diacritic dot, /č/ by <s>, and /d/ by *lāmedh*, not *dāleth*.

Phags-pa Script

The Phags-pa script was composed by the lama Phags-pa under the Yuan Emperor Kubilai to render all languages of the empire. It is a variant of the Tibetan script, used for seals and with the lines running vertically. The Phags-pa orthography is influenced by the Uyghur script. Only a few Turkic words written in this script are known.

Armenian Script

Members of the Armenian diaspora in the Crimea came into contact with Kipchak Turks in the fourteenth century. Some Kipchak groups became Christians and used the Armenian script for writing ecclesiastical and secular texts. After the great migration of the Armenians to Asia Minor, some Ottoman groups also adopted the Armenian script. Orthography and graphotactics differ in many details, e.g. front and back variants of /k/ are, as a rule, distinguished in Armeno-Ottoman but not in Armeno-Kipchak (see pp. 158–9).

Arabic Script

The Arabic script was used in all the various centres of the Islamic Turks. After the Persians had joined the Islamic world, a special variant of the Arabic script, adapted to the needs of the Persian language, was designed. The most important changes include the use of *jīm* with three dots for /č/ and *bēth* with three dots for /p/.

Eventually, variants of the Arabic orthography used for Turkic developed. The most important regional variants were the Khwarezmian, Chaghatay, Volga Turkic, Mamlūk and Ottoman ones. Some orthographic conventions became widespread and were used in most of the regions. There was thus a tendency to use emphatic Arabic letters to render consonants in back environments. Even in texts which otherwise use the three-dotted *bēth*, word-final /p/ is rendered by *bēth* with one dot, mainly in suffixes. Plene (full) notations of vowels with *wāw*, *yā* or 'alif do not reflect etymological length (cf. p. 90). The so-called *siyāqat* script was a special script used in the Ottoman administration for treasury accounts etc.

Systems Used for Modern Turkic Languages

Arabic Script

In modern times, the use and orthography of the Arabic script have undergone several reforms.

The new system used for Modern Uyghur in China has some special features (see Appendix 3). All vowels are written. Thus, the letter *hē* is used for /e/, *yā* with two dots for /è/, *wāw* with three dots for /w/. Simple *wāw* is used for /o/, *wāw* with a *háček*-like diacritic sign for /ö/, a *wāw* plus a *damma* /u/, and *wāw* with a superscribed 'alif for /ü/. Several Turkic peoples of the Soviet Union had to abandon use of Arabic script in the 1930s. Some of the newly independent Turkic republics have been considering adoption of the Arabic script again.

Hebrew Script

Kipchak Turkic groups in the Crimea became believers in the Old Testament around the twelfth century. Some of the so-called Karaims later migrated to the regions in modern Lithuania and Ukraine where they still live. They used the Hebrew script for writing both Hebrew and Karaim texts. The graphotactics reflect the local dialects, e.g. the use of *yōdh* in the Halič dialect to denote delabialised vowels (*e* < *ö*, *i* < *ü*).

Latin Scripts

By a decree of 3 November 1928, Kemal Atatürk introduced the modified Latin alphabet in Turkey. Letters of special interest are <c> for /j/, <ç> for /č/, <j> for /ž/, undotted <ı> for /i/, <ş> for /š/, and the so-called 'yumuşak ğ' (see pp. 205, 207). One year later, Soviet Azerbaijan adopted a similar Latin script which had been in preparation for several years.

In the 1920s and early 1930s, several Turkic peoples of the Soviet Union elaborated Latin-based alphabets in order to change their earlier writing systems and published books, newspapers etc. in these scripts. After a governmental decision, however, all had to replace their systems by new Cyrillic ones, and most of the texts written in Latin scripts were obliterated. In the 1990s, several Turkic republics, Azerbaijan, Kazakhstan, Kirghizstan, Turkmenistan and Uzbekistan, have developed plans to adopt the Roman script again and have begun to implement these plans.

Cyrillic Script

In pre-revolutionary Russia, certain Turkic groups tried to use the Cyrillic script to write their languages. Chuvash intellectuals were most successful in these efforts. After long experiments, a Cyrillic alphabet and an orthography based on phonological principles was designed in 1871. The present version of the Chuvash script is based on a reform carried out in 1938.

At the end of the 1930s, all Turkic peoples of the Soviet Union acquired their own Cyrillic alphabets by political decision. The design of the new alphabets was, however, not coordinated. One and the same phoneme was rendered by different letters or combinations, and one and the same letter represented different sounds in different languages. Thus, Tatar <κ> is used for front and back variants of /k/. The back variant [q] is rendered by *ķ*, i.e. <κ> plus a diacritic sign in Bashkir, and by *къ*, i.e. <κ> plus a so-called 'hard

sign', in Kumyk. In Noghay, a front syllable such as *kü-* is marked by a 'soft sign' after each letter: *кьуь-*. In Uzbek, the letter <o> represents a labialised /â/. For Uzbek and Kazakh, see Appendices 1 and 2.

Not only were different orthographic solutions offered for individual Turkic languages. Inconsistencies were often found within the individual systems themselves. In the committee responsible for the Tatar orthographic reform, the problem of how to render the word-initial sound corresponding to Old Turkic *y-* was solved by majority decisions for each single word. Thus, the present written norm of Tatar reflects a pronunciation which does not exist anywhere as such.

Further Reading

- Clauson, G. (1962) *Turkish and Mongolian Studies*, London: The Royal Asiatic Society.
 Daniels, P. T. and Bright, W. (eds) (1996) *The World's Writing Systems*, Oxford: Oxford University Press.
 Gabain, A. von (1974) *Alttürkische Grammatik*, 3rd edition, Wiesbaden: Harrassowitz.
 Róna-Tas, A. (1991) *An Introduction to Turkology*, Szeged: University of Szeged.
 Scharlipp, W.-E. (1995) *Türkische Sprache, arabische Schrift. Ein Beispiel schrifthistorischer Akkulturation*, Bibliotheca Orientalis Hungarica 44, Budapest: Akadémiai Kiadó.

7 Old Turkic

Marcel Erdal

Introduction

In the present chapter, Old Turkic is taken to be the language underlying three corpora. The first one consists of official or private inscriptions in the runiform script, dating from the seventh to tenth centuries, in the territory of the second Türk empire and the Uyghur steppe empire – present-day Mongolia – and the Yenisey basin. The second and most extensive corpus consists of ninth to thirteenth century Old Uyghur manuscripts from northwest China in Uyghur, runiform and other scripts. It includes religious (mostly Buddhist), legal, literary, medical, folkloric, astrological and personal material. The third corpus consists of eleventh-century texts from the Karakhanid state, mostly in Arabic script, including a 6,500-couplet poem, Yūsuf of Balasayun's *Qutaḍyu bilig* 'Wisdom that brings good fortune', and a Turkic–Arabic lexicon and encyclopedia featuring grammatical and dialectological notes, Maḥmūd al-Kāšġarī's *Dīwān luyāti't-Turk* 'Compendium of the Turkic Dialects' (see p. 29).

Sound System

The scripts of most texts are ambiguous even on the phonemic level, but the Brāhmī and Tibetan scripts, put to use for Old Turkic, are quite explicit. The most common was the Uyghur script, used by adherents of all religions. Manichaeic and Syriac writing was used by Manichaeans and Christians respectively.

Phonetics and Phonology

Vowels

These three scripts and the Orkhon Turkic runiform script do not distinguish */i/*, */e/* and */i/*, */o/* and */u/*, */ö/* and */ü/*. They do distinguish */a/* from */e/* and have a digraph for */ö/* and */ü/*. Some runic inscriptions of the Yenisey area are more explicit than those of the Mongolian area in having special characters for */è/* or */e/*. Most runic texts write */e/* with the character used for */a/*.

There are nine vowel phonemes, */a/*, */e/*, */è/*, */i/*, */i/*, */o/*, */ö/*, */u/* and */ü/*. In Turkic words, these are distinct only in the first syllable. Elsewhere there are

oppositions between */A/*, */U/*, */I/* and */X/*.

As for */A/*, */a/* alternates with */e/* according to the front vs. back harmony. As for */I/*, */i/* and */i/* alternate, tending to generalise *i* in the last syllable, particularly when adjacent to *y*, *č* or *š*. */U/* is realised as */u/* and */ü/*, also as */o/* and */ö/* before */k/*. */X/* is realised as */i/*, */i/*, */ü/*, */u/*, depending both on fronting and rounding, and partly also as */a/*, */e/*, */o/*, */ö/* after these same low vowels, i.e. with full assimilation (cf. pp. 107–8). Quite often in contiguity with */g/*, */t/*, */l/*, and rarely elsewhere, */i/* is spelled like *a*, e.g. *amal*, *tatay*. The front vowel */i/* is lowered very rarely. In Karakhanid, the opposition */i/* : */i/* is weak in single-syllable stems, with fluctuation in suffix harmony, according to the weak distinction between the realisations of runic */I/*. However, there is enough to indicate that the fronting of */i/* was subphonemic in all stages of Old Turkic (cf. pp. 89–90).

In non-first syllables, */o/* and */ö/* seem to occur as follows. Suffixes with */o/*, */ö/* contain a */k/*, e.g. *-OK*, *-yOK*, *-sOK*, *-dOK*, cf. also the enclitic *(O)K*. Suffixes with */U/* do not contain a */k/*, e.g. *-Ur*, *-(X)nčU*, *-GARU*, *-dUrXK*; cf. also the enclitic *mU*. This gives complementary distribution, */o/*, */ö/* here being realisations of */U/*. Second syllable */o/* and */ö/* also appear in opaque words (primary stems) with */o/* or */ö/* in the first syllable, e.g. *orto* 'middle', *toqo* 'belt buckle', *yoyto* 'mane', *kövdög* 'body', *olor* 'sit', *bodo* 'paint'. */U/* is found elsewhere in second or third syllables, e.g. *altun* 'gold', *küdegü* 'bridegroom'. This distribution also being complementary, there is no phonemic */O/* in non-first syllables.

The enclitic particles *OK* and *mU* follow front vs. back harmony. In Orkhon Turkic, *GU* and some postpositions do so as well.

Early evidence for distinctive vowel length differences is limited to an opposition between */a/* and */a:/* in first syllables of one runic source. The ability of the Brāhmī and Tibetan scripts to distinguish length is not put to use. However, Karakhanid apparently did have length distinctions also in Turkic stems.

Consonants

The voiced oral consonants probably had fricative main variants, β (or ν), δ and γ , but were realised as stops, *b*, *d*, *g*, after *r*, *l*, *n* and partly *z*. Characters

Table 7.1 The consonant system

	Unvoiced oral	Voiced oral	Nasal
labial	p	v	m
dental	t	d	n
palatal	č	y	ń
velar	k	g	ŋ
sibilant	s š	z	
liquid		r l	

for *n̄* exist only in the runiform script. While some Manichaean fragments have the digraph <ny>, a few manuscripts have <n>, which, according to one assumption, in fact represents *n̄*. Most sources merge *n̄* with *y*.

The runiform script distinguishes between back and front use for most consonants, e.g. *b*¹, *b*². Most other writing systems only distinguish between front and back /k/ and /g/. The variants of /k/ have mostly been transcribed as *k* and *q*. The back variant is sometimes a fricative *χ*, notably beside *š*. The variants of /g/ are often transcribed as *g* and *γ*. The voiced consonant *b*, found in initial position, is here assigned to /p/. It is spelled with *p* in the Uyghur and Brāhmī scripts, but not in the Manichaean and runic scripts.

Morphophonology

Phonotactic rules allow no vowel sequences, and no consonant clusters at the end of syllables unless the first consonant is /r/, /l/ or /n/ and the second one voiceless. Foreign syllable-initial clusters are broken apart by /l/, e.g. *may-tri* > *may-ti-ri* 'Maitreya', in late texts also elsewhere beside /r/, e.g. *ördek* > *ödrek* > *ödirek* 'duck'. Suffixes and enclitics get phonotactically compatible with stems through allomorphs, one set beginning with a consonant, another with a vowel, to suit stems ending in vowels or consonants.

Nasals appear word-initially only in interrogative elements. Initial *b*-changes to *m*- when the following consonant is a nasal. This change is already witnessed in runic Turkic. Turkic words do not have initial /s/. Certain words beginning with vowels show a sporadic initial /y/, e.g. *yün*- 'come up' beside *ün*-.

Morphology

Nouns and Adjectives

Word Formation

Deverbal nominals in *-(X)G*, *-(O)K*, *-Xš*, *-(X)nč*, *-(X)nčU*, *-(X)m*, *-(X)n*, *-(U)t*, *-mA*, *-(X)GlXG*, *-(X)nčsXz* etc. designate actions or actants. They form first-actant nominals with intransitive verbs, e.g. *bašlay* 'leader' ← *bašla*- 'do well, be a leader', and second-actant nominals with transitive verbs, e.g. *qusuy* 'vomit, what is vomited' ← *qus*- 'vomit'. The suffixes *-(X)GčI*, *-GUčI*, *-(X)mIXG* and *-mAKsXz* derive first-actant nominals, e.g. *satiyčĭ* 'merchant' ← *sat*- 'sell'. The suffix *-(X)nčIG* forms potential second actant nominals, e.g. *amrančĭy* 'lovely', *qorqinčĭy* 'frightful'.

The most common formatives for deriving nominals from nominals are the following:

-(X)č and *-KIñA* ~ *-KIyA* express diminution and endearment, *-sIG* forms adjectives signifying 'similar to or trying to be like the base nominal', as *öjisig* 'distinct', *barsiy* 'as if existing', *uluysiy* 'vain'.

-AGUt forms designations for members of social classes, as *bayayut* 'notable'.

-GXl, *-sIl* and *-Xš* form colour names.

-dUrXK forms names of implements connected with body parts, as *beldürük* 'belt'.

-čI, *-dAš* and *-lXK* have similar functions as in later Turkic languages (see pp. 160, 187–8, 209, 250, 263, 286, 305, 322, 347, 407).

-sXz denotes lack when added to nouns, but negation when added to adjectives, e.g. *beksiz* 'infirm', *čĭnsiz* 'untrue', *adĭnsiysiz* 'immutable'.

The suffix *-lXG* signifies, in Orkhon Turkic, 'possessing the denotee of the base', as in *küčlüg* 'powerful', *küñlüg* 'having female slaves'. In Uyghur, it has a wide variety of uses, indicating origin, material etc. and forming appositions and metaphors, e.g. *kümüšlüg tirgükler* 'silver pillars'.

The orientational formative *-dXn*, which differs in shape from *-dIn* except after unrounded stems, is added to a set of deictic bases such as 'east', 'inside', 'left' or 'behind', e.g. *Selege kedin* 'west of the Selenga', *qanyudun* 'in which direction'. It never has an ablative meaning, and it may bear possessive and case suffixes, including the ablative, e.g. *Oyuzduntan* 'from the direction of the Oguz'.

The suffix *-KI* has a special status with respect to its use as a syntactic converter of phrases. It incorporates noun phrases of place and time as well as oblique forms of demonstrative pronouns into a new nominal base, e.g. *soltinqi* 'being on the left', *quriyaqi yiryaqi bodun* 'the western and northern nations', *suv üzeki* 'being on water', *ilkisizdin berüki* 'existing from the beginning of all time', *qišqi* 'winterly'.

Comparatives, Elatives and Intensifiers

The suffix *-rAK* forms elatives and comparatives from adjectives and adverbs, e.g. *yarpraq sav* 'quite difficult phrases', *ašnuraq* 'before, earlier'. Adjectives, in particular colour terms, and adverbs may also be intensified by means of reduplication, e.g. *qapqara* 'quite black', *tüptüz* 'quite level', *apariy* 'quite clean', *epesen* 'quite healthy', *tuptutčĭ* 'quite uninterruptedly'.

Declension

Nominal declension includes the categories number, possession, antonymity and case. Their use is often optional, depending on context and lexical meaning.

Plural is the marked member of the number category. In runic inscriptions, plurality is expressed only with humans, *-lAr* competing with other suffixes. Uyghur and Karakhanid use *-lAr* with any entities, e.g. *üdler* 'periods of time', *teğri meñileri* 'divine pleasures'. Noun phrases can sometimes be understood as plural without formal plural marking, e.g. *adĭn teğri quvrayĭ* 'the assembly of the other gods'.

There are the following possessive markers:

Table 7.2 Possessive suffixes

	Singular	Plural
1p.	-(X)m	-(X)mXz
2p.	-(X)ŋ ~ -(X)g	-(X)ŋXz ~ -(X)gXz
3p.	-(s)l(n)	(s)l(n), -lArI(n)

The 'pronominal *n*' of the third person is absent in the nominative, but appears before postpositions. In early texts, the third-person singular possessive suffix is often used for the plural, e.g. *Üç küreg kiši kelti, savı bir* 'There came three enemy deserters all submitting the same report'.

Table 7.3 Case suffixes

nom.	∅
gen.	-(n)Xŋ, etc.
acc.	-(X)G
dat.	-KA
loc.–abl.	-dA
abl.	-dIn (in dialects where -dA serves as locative only)
directive	-GARU 'towards'
directive–locative	-rA (rare)
instrumental	-(X)n or -(l)n (possibly alternating by dialect)
equative	-čA
similative	-lAyU
comitative	-lXGU or -lUGUn (rare and early)

There are twelve case morphemes in the paradigm for bare nominal stems. Genitive variants are *-(n)Xŋ* in inscriptional and early texts, *-XG* in the inscriptions as a dissimilative variant of *-Xŋ* after /n/, *-nXŋ* in most Uyghur texts, sometimes *-nXG* in Karakhanid. A rare variant of the Uyghur and Karakhanid ablative suffix *-dIn* is *-dAn*, e.g. *süt-den*, attested in certain manuscripts and perhaps in runic inscriptions. The directive–locative suffix *-rA*, e.g. *öŋre* 'ahead', is dissimilated to *-yA* after stems containing *r*, e.g. *quriya* < **qurira* 'to the west' (cf. *quriyaru*).

When case suffixes follow possessive suffixes, some irregular forms emerge. Before equative, locative–ablative, ablative and comitative endings the third-person singular suffix appears as *-(s)In*, i.e. with 'pronominal *n*'. The dative suffix fuses with *-(X)ŋ* and *-(s)In* to produce *-(X)ŋA* and *-(s)IŋA* respectively. The first person singular has *-(X)mA* in analogy with the second person. Similarly, the directive forms are *-(X)mArU*, e.g. *evimerü*, *-(l)ŋArU* and *-(s)IŋArU*. The accusative has the pronominal form *-(X)n*, or in some runic texts *-(l)n*, in the first- and second-person singular, e.g. *bodunumun tériu*

'organising my tribe'. The plural possessive forms show *-nI*, like the pronouns.

Pronouns

Table 7.4 Personal pronouns

	1p.sg.	1p.pl.	2p.sg.	2p.pl.
nom.	ben ~ men	biz	sen	siz
gen.	beniŋ ~ meniŋ	biz(n)iŋ	seniŋ	siz(n)iŋ
acc.	bini ~ mini	bizni	sini	sizni
dat.	baŋa ~ maŋa	biziŋe	saŋa	siziŋe
loc.–abl.	minte	bizinte ~ biznite	sin(i)de	sizinte ~ siznide
abl.	minidin	biznidin	sinidin	siznidin
similative		bizinčüleyü	siničüleyü	sizničüleyü
directive	baŋaru	biziŋerü	saŋaru	

First- and second-person personal pronouns have the plural suffix *-(X)z* instead of *-lAr*. Pronouns show vowel alternations between /e/, /i/ and /a/. They also retain obsolete case forms. They may allow two case suffixes, e.g. *nečede* 'as often as'. Some case forms are constructed on an element *-ni-* or *-in-*. Additional locative forms are *mindide* and *sintede* ~ *sintide*. The *n* of *bizniŋ* and *sizniŋ* is absent in Orkhon Turkic and in Soghdian script manuscripts. Only a few runic inscriptions have *ben*, *bini* etc., all other sources showing *m-*. One inscription has *men* beside *ben* as an enclitic personal marker after verbs. In the course of Old Turkic, pronoun declension was extended to more and more nominal domains.

Table 7.5 Demonstrative pronouns

	'this'	'these'	'that'	'those'
nom.	bo	bolar	ol	olar
gen.	munuŋ ~ monuŋ	bolarniŋ (QB)	anıŋ	olarniŋ
acc.	bunı ~ munı	bolarnı	anı	olarnı
dat.	muŋar ~ muŋa	bolarka (QB)	aŋar ~ aŋa	olarka
loc.–abl.	bunta ~ munta	bolarda (QB)	anta	olarda (QB)
abl.	mundin (QB)	bolardin (QB)	antın	olardin (QB)
directive			aŋaru	
equative	bunča ~ munča		anıča ~ anıča	
similative	munılayu		anılayu	
instrumental	munın (QB)		anıñ	

QB = *Qutađyu bilig*

The oblique stems *bun-* and *an-* differ from the nominative, and *bo* and *bun-* also differ in the vowel. The plural forms are not found in the inscriptions. The nominative *ol* is also used as third-person personal marker on predicates. Some forms exhibit intercalary elements or double cases, e.g. *munčada*, *antada*, *muntuda*, *muntadan*, *antača*, *ančaqa*, *ančan*, *muntirtin*, *antiran*, *ančolayu*.

Words signifying 'all' are *qamay* ~ *qamıy*, *alqo*, *yomyi*, *tolp*, *tüzü*, *barča* and *bari*. The numeral *bir* 'one' may be used as an indefinite article, e.g. *Bir qari öküzüg qomursya yemiş* 'An old cow was eaten up by ants'.

The reflexive pronoun *kentü* 'own, one's self' takes number, possessive and case suffixes. Interrogative–indefinite pronouns taking case suffixes only are *ne* 'what' and *kem* 'who' (in inscriptions and runic manuscripts) or *kim* 'who' (elsewhere). They have nominal declension but form, like *kentü*, the accusative with *-nl*. The interrogative *qanyu* > *qayu* 'which' is formed from a root **qa* (see Adverbs). *Ne* has a collective derivate *negü*.

Numerals

While the digits and the lower decades are opaque, the cardinals *altmıš* 'sixty' and *yetmıš* 'seventy' contain *alti* 'six' and *yeti* 'seven' respectively. Higher numbers are expressed multiplicatively, e.g. *sekiz on* 'eighty' = 'eight tens', *toquz on* 'ninety' = 'nine tens', *üč tümen yeti biñ tört yüz* '37,400'.

In the runic inscriptions and earlier Uyghur manuscripts, cardinals from the second to the ninth decade are formed with the digit from the lower decade plus the higher decade, e.g. *yeti otuz* '27' = 'seven thirty' (cf. p. 52). The order may also be decade + *artoqi* 'its supplement' + digit, e.g. *tört yüz toquz on artoqi beš* '495'. Later texts have forms without *artoqi*, e.g. *otuz tört* '34'.

Distributives are formed with *-(r)Ar*, ordinals from 'three' on with *-(X)nč*, e.g. *toquzunč* 'ninth'. The ordinal for *eki* 'two' is *ekinti* 'second', and 'first' is expressed by *ilki*, *başlayuqi* or *baštinqi* 'being at the head'. Collectives may be derived with *-(A)GU*.

Verbs

Derivation

The most common formatives for denominal verb derivation are *-IA* and *-A*, which form transitive and intransitive verbs. The suffixes *-U*, *-(A)d*, *-(X)K*, *-(A)r* and *-IA*n only form intransitives. Onomatopoeic and synesthetic intransitives have *-Klr*, *-trl* or *-rA*. Intransitive verbs in *-sIrA* denote lack or loss. The suffix *-(X)(r)KA* forms transitive verbs expressing feelings, attitudes or opinions.

Denominal verb formation also includes analytic derivation by means of lexicalised verbal phrases with auxiliary verbs + nominal elements that do not function as free objects. The nominal elements are native or copied nominals or onomatopoeias, e.g. *asiy tusu qil-* 'benefit', *kšanti qil-* 'confess', *nom nomla-* 'preach' (etymological figure), *šük tur-* 'stay quiet'.

Deverbal verb derivation uses formants such as *-(X)GsA*, which forms desideratives. Adding *-(X)msIn* to a verb makes it describe a given behaviour as pretence.

Deverbal verbs may also be derived analytically with auxiliary verbs added to converbs of lexical verbs. The combinations may express various modifications of the lexical content, e.g. with respect to actionality (pp. 113–14, 150–1). Thus, *al-* can signify 'do for oneself', while *bér-* can convey the notion of acting for somebody else.

Most deverbal verb derivation suffixes express diathesis. The category of voice consists of a set of intercombinable suffixes. Stems in *-(X)š* are cooperative–reciprocal, those in *-(X)l* or, in late Buddhist texts, *-tXl* or *-tUrXl* are passive. Stems in *-(X)n*, *-lXn*, *-(X)d* or *-(X)K* are reflexive, middle or anti-transitive. Stems in *-Ur*, *-Ar*, *-GUr*, *-tUr*, *-Xz* or *-(X)t*, later *-(I)t*, are causative. For stems in *-sXK* and *-tXz*, see p. 155.

The suffix of verbal negation is *-mA*. Special variants will be indicated below.

The verb *u-* expresses ability to act. Its positive form is rare. Most instances have the form *u-ma-*, e.g. *Idiši bolu umazlar* 'They cannot become its vessel'. The verb *u-* usually follows the *-A* converb, whose vowel is mostly standardised to *-U*. Though the two elements may already have fused phonetically into a new suffix *-UmA*, they are joined in spelling in Soghdian-script texts only. The positive counterpart of this in Uyghur is *-U bol-*, *-GAlI bol-*, the negation of which is similar in content, e.g. *Bilgeli bolmadi* 'It was impossible to recognise'.

Finite Verb Forms

Most finite predicative verb forms, including most aspecto-temporal forms, may assume pronominal personal markers. The plural suffix *-IAr* is found optionally in some third-person forms, e.g. *Öčešmişler* 'They are said to have competed'. The simple past displays possessive personal markers.

The simple past in *-dl* is usually a constative preterite. The Uyghur finite form in *-yOK* expresses a vivid past.

Table 7.6 Simple past

	Singular	Plural
1p.	-dXm	-dXmXz
2p.	-dXŋ	-dXŋXz
3p.	-dl	-dl(IAr)

The aorist is formed with *-yUr*, rarely *-r*, after vowels, *-Ur* after causative, passive, reflexive and cooperative–reciprocal stems, *-Ir* after *-(X)t* stems, and

-Ar after most simple stems and denominal and deverbal stems in -(X)K. The negation is -mAz, also followed by pronominal personal markers. This present can also be used in timeless statements or for future reference, e.g. inscriptional *Yerimin suvumun qonar köçer ben* 'I alternately settle and nomadise on my land', *Ken kelteçi burxanlar burxan qutın qanta bulırlar?* 'Where will the future Buddhas attain Buddhahood?'

The post-terminal past in -mlš, negated -mAdOK, joined by personal pronouns, in the third person often *ol*, mostly expresses inferential past, e.g. *Qarloq eşiñe kelmedök* 'The Karluk have obviously not come for service', *Qaşi xan ögi qatun oyliğa neçe aytsar neñ kiginç bérmedök* 'However often his parents asked him, he never gave an answer'. The form in -mlš also combines with the copula *erür*, negated *ermez*, e.g. *Altı yüz tümen yıl ertmiş erür* 'Six million years have passed'.

For future reference, Orkhon Turkic uses the prospective in -dAçI, negated -mAçI, e.g. *Ölmeçi yitmeçi sen* 'You will not perish'. The rest of Old Turkic uses -GAy. Uyghur and Karakhanid -GAIr expresses imminent future, e.g. *Kün teñri uyaqyalır* 'The sun is about to set'.

There are also analytical forms consisting of the aorist, -mlš, -dI, -dAçI or -GA(y) followed by forms of *er-* 'be': *erdi* for anteriority, *ergey* for prospectivity, and *ermiş* for inferential past. Thus, -mlš can be transposed into the past to form a pluperfect, e.g. *Burxan yertinçüde belgürmiş erdi* 'The Buddha had appeared on earth', or into the future by adding *ergey*. Uyghur *bol-* 'become' instead of *er-* marks the action as transformative.

The modal optative-imperative paradigm has a person-number conjugation, but can add -IAr to the second- and third-person plural. Examples: *Yoriyin* 'I'll march', *Yazmayın* 'Lest I write', *Qal* 'Remain!', *Bergil* 'Give!'. The optional -GI is incompatible with negation. A variant -(A)yI appears in Karakhanid, a variant -zUnI in Karakhanid and one Soghdian script source. In some early texts, -(X)ñ is used for polite address to singular addressees and -(X)ñIAr for plural addressees. In other texts, -(X)ñ is also used for more than one socially subordinate addressee. Orkhon Turkic uses -(X)ñ and -zUn for plural addressees.

Table 7.7 Optative-imperative

	Singular	Plural
1p.	-(A)yIn	-(A)lIm
2p.	Ø, -gIl, -(X)ñ	-(X)ñ, -(X)ñIAr
3p.	-zUn	-zUn, -zUnIAr

Non-finite Verb Forms

Non-finite verb forms are action nouns referring to non-factive actions, participles referring to participants in the action or to factive actions, or

converbs. Both action nouns and participles can be used attributively or nominally, irrespective of whether they refer to participants or to actions.

Verbal nouns (action nouns) are pure infinitives or have prospective meaning. -mAK infinitives have the first actant in the genitive or nominative, e.g. *Samtso açarınıñ xan birle bitig idişmañın uqıtmaq* 'the teaching of the correspondence between Xuanzang and the emperor', *Çinek kəginç bərmeki nom* 'the treatise of Çinek's answer'.

In Orkhon Turkic and some early Uyghur sources, the prospective -sXK is used to denote actions, e.g. *Tün udışıqım kelmedi* 'I did not feel like sleeping at night', *wusanti olorsuq törö* 'the rule of sitting in fasting'. It also occurs in lexicalised expressions, e.g. *kün tuysuq* 'east'. Most texts only use -GU, e.g. *Sıyyusi yoq erdi* 'It could not be expected to fit', *tınlıylarıñ qutıaryuda* 'when endeavouring to save living creatures', *qılmayı qılınç* 'a deed not to be done'.

As for the participles, the intraterminal (present) participle in -(X)GmA, which never takes possessive personal markers, is used to refer to, or qualify, actants, e.g. *bo aynayu yatayma* 'this person lying there writhing', *teñri yèrinte erigme teñriler* 'the gods staying in the divine country'. The suffix -dAçI forms a prospective participle in the inscriptions, and an intraterminal (present) participle in Uyghur, e.g. *bo nom erdini bititdeçi upasi Küç Oğul* 'the lay brother Küç Oğul, who commissioned the copying of this sūtra-jewel', *nomuy nomladaçi* 'he who preaches the doctrine'. The aorist is rare as a participle, e.g. *uyaqur yultuzlar* 'setting stars'. The Uyghur -(X)GI participle, e.g. *edgü öglilerim* 'my well-wishers', is practically absent in Orkhon Turkic.

Relative anteriority is expressed by the post-terminal participle in -mlš, negated -mAdOK. In the inscriptions, it mostly occurs as the dependent of a head, often with possessive personal markers, e.g. *dyanqa kirmiş teñriler* 'the gods who were in a state of meditation', *Men anta azun tutmişim kergek* 'I should be born there'. It may also refer to actions, e.g. *étmiş yaratmış aš* (runic manuscript) 'a well-prepared meal'. -dOK, which is not used in Uyghur, takes optional possessive personal markers, refers to actions and is used in circumstantials, e.g. *sançdoq yèr* 'the place where we routed them'. Both -mlš and -dOK forms may be used as nouns, e.g. *Ödökümün bitidim* 'I wrote down what I remembered', *anta qalmışi* 'those among them who stayed there'. -mlš is also used with possessive personal suffixes + *bar* or *yoq*, e.g. *Burxan qutıña köñül örütmüşim bar* 'I have been aroused towards Buddhahood'. The Uyghur -yOK participle is also post-terminal.

Verbal nominals in -GUIXK can be used as non-finite predicates in 'small clauses', e.g. *Topolyuluq alp ermiş* 'It is hard to pierce, they say', *Maña burxan qutın bulıuluq üd yayumış erür* 'The time has come near for me to attain Buddhahood'. The use of -GAI in *üzgeli uçuz* 'easy to break' is similar.

The most common converb suffixes are -(X)p and -A, the latter with the variants -A, -I, -U and -yU, e.g. *Seleñe keçe udu yoridim* 'Crossing the

Selenga, I marched after (them)'. $-(X)pAn$, rarely $-(X)pAnXn$, is used in inscriptional and early Uyghur texts only, e.g. *él örginin anta örgepen* 'setting up the national throne there'. All are negated with $-mAtI(n)$, the variant $-mAtI$ occurring in runic inscriptions and Karakhanid. The inscriptions have a rare form in $-(X)yXn$, negated $-mAyXn$, e.g. *Süggülüg qantün keliyin süre èltidi?* 'Where did armed forces come from to drive you away?'

The converb in $-GAlI$ is purposive or temporal. Temporal converbs also include $-dOKtA$, which is both inscriptional and Uyghur, and $-GInčA$, which is not attested in the inscriptions. The conditional suffix is $-sAr$ or, in Karakhanid and a few non-standard Uyghur texts, $-sA$. The Karakhanid forms are often joined by first- and second-person pronominal personal pronouns. Uyghur $-sAr$ forms can be impersonal, e.g. *teğri yèrinteki yil sanin sanasar* 'if one reckons by years of the divine country'.

Adverbs

Adverbs are formed with $-IA$, case forms, fossilised $-A$ converbs, etc., e.g. *utru* 'forth', *birle* 'together', *aşru* 'through'. Interrogative adverbs are formed from the root $*qa$, e.g. *qanta* 'where', Orkhon Turkic *qantan*, Uyghur *qantiran* 'from where', *qanča* 'how much, how far, where to', *qač* 'how much', *qani* 'where', the latter used in rhetorical questions.

Other Morphological Categories

Postpositions may be opaque like *teg* 'like', have a pronominal origin like *berü* 'hither, since', or go back to other sources, e.g. *birle* 'with' ← *bir* 'one'. Most are petrified $-A$ converbs, e.g. *keče* 'beyond', *qudi* 'down', *tapa* 'towards'. The meanings may be abstract as in *üčün* 'for, because of', *osuylyuy* 'as, like', or concrete as in *qudi* 'down', *tegre* 'around'. The postposition *üze* 'over, by' both marks instruments and, in Uyghur, the agent in passive constructions.

The postpositions *ara* 'among', *teg*, *birle*, *üčün*, *osuylyuy* and *üze* govern the accusative of demonstrative and personal (not interrogative) pronouns, and the nominative of other nominals, e.g. *tatar birle toqı* 'fight against the Mongols', *yavlaqın üčün* 'because they are wicked', *taluy ögüz osuylyuy* 'similar to the sea'. Postpositions such as *öği* 'distinct, except', *kên* 'after', *berü* 'since' govern the ablative or ablative–locative, e.g. *nomta öği* 'except by the doctrine'. A number of elements are used both as postpositions and as adverbs.

Conjunctions (free junctors) include *azu* 'or', *yene* 'moreover', *inčip* 'thus', *ötrö* 'then', *taqi* 'and', *ulati* 'and, and so on', etc. Some are fossilised $-A$ converbs, e.g. *ötrö*. While most of them precede sentences, *erken* 'while', inscriptional *erkli*, are postposed. Clauses ending in $-sAr$ or $-(X)p$ can be introduced by *qalti*, e.g. *qalti taşiy alsar* 'if one takes the stone'. Conditionals may also be introduced by *birök* 'in case'. For *kim* and *tép*, converb of *té*- 'say', see p. 154.

As for particles, the emphatic element *OK*, the topicalising connective *yme* and the interrogative *mU* are enclitics inserted after the first element of

phrases or clauses, e.g. *bodisavt ol oq oronta olorup* 'the Bodhisattva sitting down in that very place'. In later Uyghur, *yme* appears as *mA*, e.g. *bizme uzun yaşap* 'we also, living a long time'.

The particle *idi* strengthens negations, e.g. *idi oqsuz* 'with no tribal structure at all', *Temir qapıyqa tegmiş idi yoq ermiş* 'It had never been the case before that they had reached the Iron Gate', *Idi yorımazun* 'In no case must they roam around'. The particle *neñ* signifies 'not any', e.g. *Neñ kiginč bërmedök* 'He is said not to have given any answer', *Neñ buñuy yoq* 'You have no trouble at all'. The elative *eñ* and *ariti* and the comparative *taqi* are preposed to noun phrases, e.g. *eñ ögre* 'foremost', *antada taqi yègrek* 'even better than that'. Uyghur *erki* 'I wonder' and inscriptional *erinč* 'apparently' follow sentences, expressing doubt.

Interjections include $(y)A$, following vocative noun phrases, the demonstrative *mına* ~ *muna* and *ona*, etc.

Syntax

Nominal Phrases

In juxtapositions of two unmarked nominals, the attribute may be an adjective, a name, or a noun used as apposition, e.g. *yimşaq ayi* 'soft textile', *Orqon ögüz* 'the Orkhon river', *taqiyyu yil* 'the year of the hen', *qulum kügüm bodun* 'the nation consisting of my male and female slaves'. Attributes are generally not inflected, except for some cases which can be considered appositions. The demonstratives, the interrogatives *ne*, *qač*, *qayu* and Orkhon Turkic reflexive *kentü* can also serve adnominally. Plural agreement is sometimes marked on nouns after attributed numerals, e.g. *sekiz tümen tört miñ baliqlar* '84,000 towns'.

Possession may be expressed with a genitive attribute and a possessive suffix added to the head, e.g. *qurıqarnıñ toranı* 'the net of the temple'. Sometimes the head lacks a possessive suffix, e.g. inscriptional *meniñ er* 'my men'. Constructions without a genitive marker such as *qan süsi* 'the royal army', *teğriler ordolari* 'the divine palaces' are also common.

Old Turkic adjectives qualify entities as bearers of qualities, but are also used as abstracts denoting the qualities themselves. Thus, *bilge* is 'wise' and 'the quality of being wise, wisdom', *armaqçı* is both 'deceitful person' and 'somebody's deceit'. $-LXK$ forms abstracts in Karakhanid only. Uyghur uses *bilig* for mental states, e.g. *yarliqančuči bilig* 'pity'. In principle, any adjective may be used adverbially, e.g. *yavlaq saqınıp* 'having evil thoughts'.

Case Functions

Stems unmarked for case can be used in most functions, except concrete or abstract motion towards an entity or away from it. The absence of the accusative suffix is not related to non-specificity, e.g. *Qara qum aşmıñ* 'They had crossed the Kara-Kum desert'. Definiteness and specificity are not

expressed by case marking. In the inscriptions, even pronouns appear as direct objects without accusative suffix, e.g. *bo bitidökte* 'when I wrote this'. The accusative marks direct objects, e.g. *Šimnuv utup* 'defeating Mära'. The dative may signal meanings of concrete or abstract direction, time and source of processes and events, etc., e.g. *Teñriler ordoqa yüyülurlar* 'The gods assemble at the palace', *teñri quvrayıña nom nomlayu* 'preaching to the assembly of gods', *Toquz otuzqa süñüsdüm* 'I fought on the 29th', *teñriler teñri meñsige esürüp* 'the gods being intoxicated by divine bliss'. In *oqıyçıqa oqıtıp* 'having somebody called by a herald' the dative marks the causee accompanying a causative. *Tavyaçqa Bay Baliq yapıtı bértim* may thus be translated as 'I had Bay Baliq constructed for or by the Chinese'.

The locative refers to placement or states, e.g. *Kögmen tayda* 'on the Kögmen mountain', *Qırqız bodunuy uda basdımız* 'We fell upon the Qırqız people while they were asleep'. In early sources it also has concrete or abstract ablative meaning, e.g. *Qayanda bedizçi kelürtüm* 'I obtained decorators from the emperor', *Anıñ kücinte küsüşleri qanar* 'Through his powers their wishes come true'. Comparatives may govern the *-dA* form, e.g. *ay teñri tilgeninte sevigligrek* 'lovelier than the moon disc', *barçada üzerek* 'highest of all', *öginte qañinta taqı yègrek* 'even better than their father and mother', *antada taqı yègrek* 'even better than that', *qamayda öñre* 'initially'. Structures such as *avınyuta avınyu* 'most amusing' may express superlative content. The ablative signals origin etc., e.g. *Teñri yerintin tayarlar* 'They slip down from the divine land'.

The instrumental may have instrumental, local, temporal meaning as well as express manner and measure etc., e.g. *Yarliqančuçi biligin inča saqınč saqınur* 'In his commiseration, he has the following thoughts', *Çitimin yayladım* 'I spent the summer within my borders', *tüniün küniün* 'by day and by night', *Ayır qadyun yanıp bardı* 'He burned away in great grief', *teñri yerinteki yıl sanın sanasar* 'counting by divine country years'.

The equative signifies 'according to', e.g. *yalñoqlar sanınča böğser* 'if one sums it up by human reckoning', *toquzunč qay törösiniče olorup* 'sitting in the manner of the ninth stratum', *türk tilinče evir-* 'translate into Turkic'.

The simulative in *-lAyU*, common in Orkhon Turkic, Uyghur and Karakhanid, means 'like, similar to, as if'.

Verbal Phrases

A verbal phrase can contain auxiliaries, separated from the lexical verb only by particles such as *mU* and *OK*, e.g. analytical forms such as *bërmez ök erser* 'if (s)he does not give'. The auxiliary *u-* 'be able' fuses into a morpheme expressing impossibility (p. 145).

Orkhon Turkic and Karakhanid *bër-* and Uyghur *yarli(γ)qa-* express politeness when signifying 'favour with doing, to deign to do, to say'. The self-effacing counterpart of *yarliqa-* 'say' is *ötün-*, the self-effacing counterpart for 'doing' is *tegin-*.

The verbal phrase may also contain verbs expressing actionality. In the inscriptions, *id-* 'do completely' and *kör-* 'make sure to do' are added to the *-A* converb. Descriptive constructions in Uyghur include *-U alq-* 'finish doing' and *-(X)p qod-* 'do exhaustively'. Uyghur *tur-* 'stand (up), arise' expresses duration with *-(X)p* and *-A* converbs. Constructions with *-GAll* express prospectivity and imminence, e.g. *-GAll uyra-* 'intend to do', *ölgeli tur-* 'be about to die'.

Certain verbs may establish a predicative relationship between their direct object and another nominal, *Özümün öñre biña başı ıddı* 'Myself he sent forward as captain', *Eki qızın tapıy bérđi* 'He gave his two daughters as tribute'.

The Sentence

Finite clauses have a predicate, which can be verbal, nominal or adverbial. They need not have a topic or a subject, not even an implicit one. First- and second-person verb forms are sometimes accompanied by a nominal or pronominal subject, e.g. *Ilgerü bariyma bardıy, qurıyaru bariyma bardıy* 'Those of you who went east departed, and those of you who went west departed as well', *Yigirmi kün olorup bo tamqa qop Yolluy Tegin bitidim* 'I Yolluy Tegin, sat here for twenty days and wrote all of it on this wall'. There is no agreement in the sense that categories of one word must be reproduced in another, but redundant expression of categories is not excluded. Verbs often stand in the plural if they have plural first actants, e.g. *Alqo ünliylar meniñ yatim ermezler* 'No living beings are strangers to me'.

Constituent Order

The order of constituents is strongly topic-comment oriented. However, subject-object-verb seems to be the unmarked order. The sentence often starts with a connective, e.g. *anta ötrö* 'thereupon', and the finite verb is often at the end. A locative or an accusative object can precede the subject, e.g. *Tamu yer suvlarda ol yaroq yaltrıq tegir* 'That bright gleam reaches all the way to the realm of hell', *Qulum küñüm bodunuy teñri yer ayu berti* 'It was the sky and the earth who enabled me to command the nations who were (to be) my slaves'. Direct objects without accusative suffix are not confined to preverbal position, e.g. *Qaryu edgüti uryıl* 'Place the watch-towers well!'. Topicalising left-location is found in cases such as, *Kögmen, yoli bir ermiş, tumış teyin ešidip* 'As for the Kögmen range, I heard that there was only one way across it and that it was blocked'.

Negation

Verbal predicates are negated with *-mA*. The scope of the negation of a superordinate verb does not extend to subordinate ones, e.g. *Taqı qamay qamlar tər(i)lep neñ tırgürmegey* 'Even all the magicians will, assembling, definitely not bring him back to life'. Non-existence and lack are expressed by *yoq*, often combined with *er-* or *bol-*. Lack is also expressed by *-sXz* added

to nominals, e.g. *Köğülümüz birli maqali sözlegüsüz ol* 'Our feelings cannot be expressed by brush and ink'. The Uyghur aorist *bultuqmaz* 'is not found' is a common expression of non-existence. Other non-verbal predicates are negated with *emez*.

Interrogatives

Yes/no questions are expressed with the enclitic particle *mU*, in Uyghur sometimes followed by *erki* 'I wonder'. It follows the queried constituent, e.g. *Qayan mu qısayın?* 'Is it a kaghan I should make him?', but its unmarked position is after the verb. The Orkhon use of *GU* for rhetorical questions is similar. Interrogative pronouns normally start sentences, e.g. *Negülük ölüür sizler?* 'Why do you kill?', *Qanča baryalı saqınur siz?* 'Where are you planning to go?'

Expression of Possession, Existence etc.

Possession is expressed by the possessed entity followed by *bar* 'existing' or *yoq* 'non-existing' plus a possessive suffix referring to the possessor, e.g. *Qutum bar* 'I enjoy divine favour', *Ögrünçüq yoq* 'You have no joy'.

With nominal predicates, *er-* 'be' expresses 'being' and *bol-* 'becoming', *bar* 'there is' denotes existence and *yoq* 'there is not' non-existence. The predicate often includes a pronoun. There are also sentences without copula. The copulative elements *ol* and *erür* 'is' are frequently used, e.g. *Bo meniñ eñ kênki azunum erür* 'This is my last existence'. The element *bar* appears in passages such as *ol eki kişi bar erser* 'since there are those two people'. The form *erti* also serves as the past tense of *bar* 'there is'. The conditional *erser* serves as conditional junctor, e.g. *üd ertürürler erser* 'if they let time pass', *bolmadılar erser* 'if they have not become'. The aorist *tétir* of the causative *tê-t-* often signifies 'is' or 'is said to be'. For nominal predicates with modal force, e.g. *kergek* 'it is necessary', see p. 155.

Non-finite Clauses

Constituent Clauses

Verbal clauses may be embedded as constituents of other clauses by means of bound morphemes. They can be embedded as subjects, as in *Bo muntay körksüz yatayma ne türlüq kişi bo?* 'This person lying there in such an ugly way, what sort of a person is he?' or *Ötüq tilek bulyuluqı sarp* 'It is difficult to obtain what one wishes'. Here, *bulyuluq* represent an action and *yatayma* an actant. An example of object embedding is *El tutsuquñun bunta urtum* 'How you should govern people I have recorded here'. A nominal clause is made into an object by putting its predicate into the accusative, e.g. *Yer tarin uqüdi* 'He explained that the place was too small', *Bodisavtniñ burñan qutın bulyusın uqar mu siz?* 'Do you understand that the Bodhisattva will attain Buddhahood?' A dative-marked clause is found in *nomçi men tégmeke artizip* 'letting oneself get deceived by people who say "I am a preacher"'.

Clauses representing participants can also be marked with other cases and postpositions.

Converb Clauses

Clauses fulfilling adverbial functions are converb clauses. Of the converbs, the *-A* converb as well as the *-(X)p* and *-(X)pAn* ones are semantically rather unspecific. *-(X)p* has no content of itself beside its joining function. The agent (first actant) is normally coreferential with that of the superordinate verb, e.g. *Qarluq yavlaq saqinip teze bardı* 'The Karluk got bad ideas and fled', but there are exceptions such as *Ayisi barimi emgenmetin kirür* 'His wealth accrues without him working'. Examples of negation are *Tuymatin tuzaqqa ilinmiş* 'Without noticing he was caught in a snare', *Toquz tatar qalmatı keldi* 'The nine Tatar tribes did not keep away but came'.

The *-GAll* converb has temporal meaning, presenting the matrix clause event as occurring since the converb clause event occurred, e.g. *Uyghur Adrıyalı ür kéc boldı* 'It has been a long time since we parted'. It may also have purposive meaning, expressing the aim of the matrix clause event, e.g. *Süñüşgeli keldi* 'They came to fight'. In Uyghur, *-GAll* often occurs with verbs of attitude, intention etc., e.g. *kelgeli tapla-* 'be glad to come', *körkitgeli qilin-* 'set about to show', *virñar étgeli başla-* 'start building a monastery'. Its negative counterpart appears rather late, e.g. *Küseyür men kentü özüm anıtmayalı* 'I wish I would not let myself remember'. Clauses with *-GAll* usually have the same first actant as the matrix clause, e.g. *Aş aşayalı olormışlar* 'They sat down to have a meal'. Exceptions are found in some Uyghur examples with *id-* 'send', e.g. *Bizni sini alyalı iddi* 'He sent us to fetch you'.

The combination *-dOKıA* means 'when', e.g. *qayan uçdoqta* 'when the kaghan departed'. Orkhon Turkic *-dOK* + possessive + dative also has temporal meaning. The suffix *-GInčA* stresses temporal extension and can be translated as 'as long as' with non-transformatives (denoting processes without a critical point), e.g. *Küçi yétginče darnı sözlemiş kergek* 'The formula should be recited as long as one is able to'. It can be translated as 'until' with transformatives (denoting actions with a critical point), e.g. *Ançaqategi turmayay men qaltı siznidin alqış bulmayınča* 'I will not stand up as long as I do not receive blessing from you', *Men kelginče evig barqiy uz tuyıl* 'Keep house well until I come back'.

Circumstantial clauses are formed with *erken* 'while', inscriptional *erkli*, plus aorists, e.g. *tuy taşıqar erkli* 'the banner being out', *anča olorur erkli* 'while spending our time in that manner', *biz ögüz suvin keçer erken* 'while we were crossing the river'. The negative counterpart is *-mAZKAn*, e.g. *segrem taqı bütmezken* 'at a time when the monastery was not yet ready'.

Conditional clauses are formed with *-sAr* 'if', and sometimes introduced by elements such as *qalı*, *qaltı*, Uyghur *birök* or *apam* and Karakhanid *apañ*. The suffix *-sAr* may also have factive meanings such as 'seeing that', e.g. *üd*

teğri aysar 'since timing is determined by heavens'. Conditional-concessive clauses are formed with *-sAr* plus interrogative–indefinite items such as *kim*, e.g. *Kim qayu küseser qutluq tınılıylar ara etizü olorup aşayalı, ol kişi edgü qilinç qılzun* 'Whoever wishes to enjoy sitting among the blessed creatures making music, that person should carry out good deeds', *Qañı ögi oyliña neçe aysar neñ kiginç bērmedök* 'Whatever way his father and mother asked their son, he gave no answer at all'.

There are several other adverbial clause types marked with various cases or postpositions and fulfilling local, temporal, descriptive, causal and similar tasks, e.g. *bir eki atlıy yavlaqın üçün* 'because one or two knights were wicked'. The postposition *üçün* signifies 'because' with factive items such as *-mİš* and 'so as to' with non-factive ones such as *-GAlr* or *-GU*. The Uyghur and Karakhanid non-factive *-GUKA* also signifies 'so as to', whereas Uyghur *-mAK* or *-mİš* + possessive + dative has causal and temporal functions. The sequence *-mİšKA* is negated as *-mAyOKKA* (*-mAmİš* in late sources).

The converb form *tēp* 'saying' incorporates direct speech and content of thought, but also intentions, thus marking purposive clauses, e.g. *burxanqa tušalim tēp* 'in the hope of meeting Buddha'.

Relative Clauses

Participles used adnominally constitute embedded relative clauses, e.g. *yayı alqınmıš yula* 'a lamp whose oil has been consumed' (see pp. 61–3).

There are also non-embedding types of relativisation. In Uyghur, interrogative–indefinite items such as *kim* may precede postposed relative clauses, e.g. *Muntada adın aš yoq kim bo barsıy tırgürgülik* 'There is here no different food which would revive this tigris'. Postposed clauses of this type may also be asyndetic, e.g. the temporal relative clause in *Antay uyri boltı ol oq tınılıylarnıñ qanların içtim* 'There also were occasions on which I drank the blood of those same creatures'.

Coordination

Coordination may be implicit or explicit. From adjectives to paragraphs, everything can be linked by mere listing, the context serving as concatenator, e.g. *teğriler teğri qatunları* 'gods and goddesses', *oyulta qızta amraq* 'dearer than son and daughter', *nom nomlayu et'öz ürlüksüzin uqıtu* 'teaching the doctrine, explaining the body's transience'. Parallel syntactic structures can be coordinated by shared elements, e.g. *Yuyqa erkli topolyalı uçuz ermiš, yinçge erkli üzgeli uçuz* 'That which is thin is easy to pierce, they say, that which is slim (is) easy to break'.

Explicit coordination is possible by derivational means, e.g. between noun phrases as in *sansarliy tegzinç* 'the saṃsāra-whirlpool', with repeated particles as in *begleri yme bodunı yme* 'both their aristocracy and their common people', or by postposing *ulati* as in *qoy layzın ulati tınılıylarıy* 'living beings such as sheep and pigs'. In Uyghur and Karakhanid, *taqi* 'and' also joins larger units such as sentences.

The antonym marker *-Il* is attached to each of the members coordinated, e.g. *toroq buqalı semiz buqalı* 'lean bulls and fat bulls', *tersli oğlı qıltaçı* 'he who does wrong or right'. Possessive suffixes preceding it have the 'pronominal *n*', e.g. *bašinli adaqinli iki yilta* 'within two years, from beginning to end'. The members share the case ending, e.g. *teğrili yērlide* 'in the sky and on earth', *igidli kērtülig körüp* 'perceiving lie and truth'. Couplets are sometimes followed by *ikigü*, triplets by *üçegü*, e.g. *süli ašli kērtgünçli üçegü* 'army, provisions and faith'.

Diathetic Relations

Diathesis is encoded by voice suffixes, participant characteristics of the verb stems and case forms of the participants. Passive sentences usually have no explicit agent. In later texts it may be marked with the postposition *üze*. Verbs in *-(X)n* are reflexive, anti-transitive or middle. Many have metaphorical meanings distinct from those of their bases. Verbs in *-(X)K* and *-IXn* are anti-transitive, while *-(X)d* verbs are middle. Verbs in *-(X)š* express two participant groups vying or cooperating with each other or one participant vying with another.

Causative formatives added to intransitive stems have a transitivising effect. Added to transitive stems, they yield causative verbs, the argument structure of which implies a causee. The latter possibility applies to *-(X)t* verbs mostly when the preceding stem ends in a vowel. In earlier texts, *-(X)t* added to transitive bases tends towards passive meaning (see pp. 55–6), often with an explicit dative-marked agent. Verbs in *-tXz* and *-sXK* added to transitive bases represent the first actant, a creature with a will of its own, as a patient responsible for the action undergone; the agent is marked with the dative.

Modal Expressions

The speaker's volition in the form of orders and wishes is expressed with optative–imperative forms, often placed in non-final position. An additional interrogative element solicits the wish of the addressee, e.g. *Qayu baliqta tuyayın?* 'In which town should I be reborn?' The combination with *erti* can express unreal or polite wishes, e.g. *Körmeyin erti munı teg emgekig* 'I wish I would not have seen such suffering'. Necessity may be expressed by *-mİš* (sometimes with personal marker) *kergek* and *-GU ol*, e.g. *Ol uluš baliqta ažun tutmišim kerek* 'It is necessary for me to be born in that city'.

Text Syntax

The attested discourse types comprise didactic treatises, ritual texts, epitaphs, verse and private letters and much prose narrative. The narrative contains instances of direct speech which show such oral characteristics as vocatives and interjections, a freer word order and situation-bound deictics, e.g. *Körünğler körünğler qač yañlıy körtle qač yañlıy seviglig erür* 'See, see in how many ways he is pretty, in how many ways lovely!' Another speech

characteristic is the use of *-KİŃA* for endearment, covering whole passages and marking not only 'loved' entities.

The cohesion of the texts allows zero anaphora both within a sentence and across sentences. Cohesion can be underlined by pro-adverbs, e.g. the anaphoric *inča* and the cataphoric *anča*, by taking up lexemes from the co-text, or by *ančip* (a pro-verb), *anta ötrü* or *anta kën* 'thereupon' starting sentences and linking them to previous ones. Orkhon Turkic enhances cohesion by summing up preceding sentences in *-(X)p* clauses, e.g. *Ēlig anča tutmıř erinč. Ēlig tutup ...* 'This is how they appear to have governed the country. Governing the country, they ...', *Anča tēp qayanqa yayı bolmıř. Yayı bolup ...* 'With such words they opposed the emperor. Even though opposing him, ...'.

Topicalisation, which has been touched on above, is also possible with *erser*, e.g. *Tavyač eli erser, uč qıdıy yēr ol* 'As for China, that is a distant country'. Pronouns are used for contrast, e.g. *Biz az ertimiz yavüz ertimiz* 'As for us, we were few and in a bad state'. In *Ikinti ažuanta oq mu teginür azu bo ažuanta yme mu teginmeki bar?* 'Is it in the second birth that one attains it, or does attainment take place in this same birth?', several means of text organisation are joined: the double *mU* after the queried elements, the particles *OK* and *yme*, and the de-finitisation of the second verb to ensure that it is not taken to be the predicate. Demonstratives often point at previous segments within sentences. In *Men sini arxant qutin bulturyalı anı üčün sürüp üntürdüm* 'I had you banished to make you find arhathood', the final meaning of *-GAI* is taken up by *anı üčün*. In manuscript Uyghur the contents of a stretch of direct speech incorporated in the sentence are often again indicated with *munilayu* 'thus', e.g. '...' *tēp munilayu tutuzdı* 'He admonished him saying "...".

Lexicon

The lexicon reflects the occupations of speakers. While the inscriptions of the steppe empires feature numerous horse colour terms, for instance, Uyghur texts written by monks abound in religious terminology. Binomes and biverbs are a conspicuous feature. They are either near-synonyms, as *kü soruy* 'fame' or, more rarely, determinative sequences such as *isig öz* 'life', literally 'warm core'. The combinations *tütrüm terig* 'profound' and *bay baramlıy* 'wealthy' are adjectives, while *yaroq yaltrıq* 'gleam(y), bright-(ness)' is both nominal and adjectival. Note the strong tendency for alliteration. Biverbs such as *sev-amra-* 'like' and *tēr-quvrat-* 'assemble' place the longer element second.

Lexical copying follows Chinese, Soghdian, Sanskrit and Tokharian, in Karakhanid Arabic and Persian, and is mainly nominal. As far as can be judged from the written forms, the copies partly adapt to Turkic phonology and phonotactics.

Dialects and Registers

The three corpora mentioned on p. 138 represent a coherent group of fuzzy dialects differing mainly in lexicon. Syntactic differences may in part be due to the fact that they contain different textual types. Phonic and grammatical differences between the corpora are probably not much greater than those found within them. Some of the traits thought to be phonic may be due to graphic fluctuations preceding standard spelling. Geographical dialects can hardly be worked out. Texts range from imperial inscriptions to letters sent to family members and graffiti scribbled by travellers on rocks. There is no doubt that the texts represent the real language, the vehicle of an expanding society replacing the vestiges of local Iranian and Tokharian. Stylistic differences and registers are discernible: personal letters, medical texts and scribblings show progressive traits in consonantism and morphology.

Further Reading

- Clauson, G. (1972) *A Dictionary of Pre-Thirteenth-Century Turkish*, Oxford: Clarendon Press.
- Eraslan, K. (1980) *Eski Türkçede isim-fiiller*, Istanbul: İstanbul Üniversitesi Edebiyat Fakültesi.
- Erdal, M. (1991) *Old Turkic Word Formation I-II*, *Turcologica* 7, Wiesbaden: Harrassowitz.
- Gabain, A. von (1974) *Altürkische Grammatik*, 3rd edition, Wiesbaden: Harrassowitz.
- Johanson, L. (1979) *Altürkisch als 'dissimilierende Sprache'*, *Abhandlungen der Akademie der Wissenschaften und der Literatur*, Mainz, 1979: 2, Wiesbaden: Steiner.
- Radloff, W. (1892) *Atlas' drevnostej Mongolii*, St Petersburg. (Partly reprinted in *Orhun. The Atlas of Historical Works in Mongolia. Selected Pages*, Ankara: TİKA, 1995.)
- Röhrborn, K. (1977–88) *Uigurisches Wörterbuch. Sprachmaterial der vorislamischen türkischen Texte aus Zentralasien*, 1–4, Wiesbaden: Steiner.
- Schulz, P. (1978) *Verbalnomina und Konverbien als adverbiale Ergänzungen im Altürkischen*, Ph.D. thesis, Giessen University.
- Tekin, T. (1968) *A Grammar of Orkhon Turkic*, The Hague: Mouton.

8 Middle Kipchak

Árpád Berta

Introduction

The present chapter will summarise the major features of the Kipchak dialects spoken between the thirteenth and sixteenth centuries in the south Russian steppe and in the Near East.

The sources identified as Middle Kipchak are of heterogeneous origin. They are written in different scripts, Roman, Arabic, Cyrillic and Armenian, and they were created with various objectives over more than three centuries in different linguistic environments far from each other. The collection of 'Middle Kipchak' texts does not only reflect various Kipchak dialects. It also contains Oghuz elements, which sometimes differ significantly from the Kipchak material, and are sometimes difficult to distinguish from it.

One source written in Roman script has become known as the *Codex Cumanicus*. Its first part seems to be a language guide compiled for Italian traders, while its second part was made for German missionaries. The Codex proves that the Coman language played a significant role in the south Russian steppe at the time it was written, from the late thirteenth century to the first third of the fourteenth century. Its two parts reflect similar dialects.

Several Mamlūk sources in Arabic script are descriptions of the Kipchak language. The fourteenth century treatise *Kitāb bulyat al-muštāq fī luyat at-turk wa'l-qifjāq* 'Book of sufficient introduction to the languages of the Turks and Kipchaks', written in what is now Syria, deals with nominals and verbs. *Tarjumān turkī wa-'arabī* 'Turkic and Arabic interpreter', written in 1343 in Egypt, contains a description of Kipchak phonetics, conjugations, pronouns, nominal declension, particles, suffixes etc. The fourteenth-century source *Kitāb al-idrāk li-lisān al-atrāk* 'Book of the understanding of the Turkic language' contains a longer Kipchak–Arabic wordlist and a description of phonetics and morphology. *Al-qawānīn al-kullīya li-ḍabt al-luyat at-turkīya* 'The collection of the correct rules of the Turkic language', written in Egypt in the early fifteenth century, consists of a long section on grammar and a shorter glossary.

Kipchak texts in Armenian script mostly date from the sixteenth and seventeenth centuries and belong to descendants of a Crimean Armenian community that had adopted the Kipchak language and migrated to the territories of Ukraine and Poland (see p. 135).

Sound System

Phonetics and Phonology

The theoretical introductory sections in Mamlūk sources frequently use Arabic linguistic terms to describe the quality of Turkic sounds which differ from Arabic ones.

Nine vowel phonemes can be distinguished: /a/, /e/, /è/, /i/, /i/, /o/, /ö/, /u/ and /ü/. The texts do not provide sufficient ground for assuming long vs. short vowel distinctions. It is doubtful whether sporadic notations such as *aai* 'moon' and *toodac* 'lip' in the *Codex Cumanicus* indicate that old Turkic long vowels were preserved. Such examples are not found in texts written in Arabic and Armenian script.

Omission of vowel signs in texts in Roman and Armenian scripts indicate that unstressed vowels were very short or reduced in certain cases, e.g. *kši* 'man', *qš* 'winter', *χzil* 'golden'. The letter *e* for an expected *i* in the *Codex Cumanicus* does not necessarily indicate lack of front vs. back harmony, but may denote reduced vowels, e.g. *ačel-* 'open', *agre-* and *ager-* 'hurt', *tatle* 'sweet', *cheχ* 'virgin' and *ches* 'winter'.

The basic consonant phonemes are /k/, /g/, /χ/, /t/, /d/, /p/, /b/, /ŋ/, /n/, /m/, /t/, /l/, /y/, /w/, /s/, /z/, /š/, /ž/. The phonemes /h/, /ž/, /j/, /f/ mainly occur in loanwords.

Orthography

The Arabic script can clearly represent the distinctions between rounded and unrounded vowels and between high and low unrounded vowels. The texts differ from each other in their preference for using vowel letters and diacritic signs (*ḥaraka*) to denote vowels. The letter *wāw* and the diacritic sign *ḍamma* indicate *o*, *u*, *ö* and *ü*, whereas the letter *ālif*, the diacritic sign *fatha*, and, in final position, the letter *hā*, indicate *a* and *e*. The vowels *è*, *i* and *ī* are not distinguished but may all be denoted by the letter *yā* and the sign *kasra*.

Certain features of the Arabic script enable us to decide on vowel qualities. Letters for emphatic consonants, *ṭ*, *ṣ*, *ḍ*, may be helpful, since their use is mostly limited to back syllables. Some sources use auxiliary signs: *jīm* indicates that the word contains *a*, *o*, *ö* or *i*, while *qāf* denotes that the vowel is *e*, *u*, *ü* or *ī*. Thus, *jal-* 'play a stringed instrument' is written with *fatha* and *jīm* to indicate *a*, while *eri-* 'melt' is written with *qāf* to indicate *e*.

The Armenian script does not have separate letters for each vowel. It uses the sign <a> (*aib*) for *a* and *e*, the sign <ō> for *o* and *ö*, and <ō> + <w> for *u* and *ü*, but it differentiates between *i* and *ī* by the letters <i> (*ini*) and <ə> (*et*). It may also denote *e* and *è* with <ē> (*ē*) and <e> (*eč*).

The graphic representation in the *Codex Cumanicus* is inconsistent. In both parts it is difficult to distinguish *ī* from *i*. The Italian part does not distinguish *o* from *ö* or *u* from *ü*, and uses, like Armeno-Kipchak texts, the letter *a* for both *a* and *e*. The German part is more reliable in specifying vowels, but

inconsistent in representing the distinctions *u* : *ü* and *a* : *e* : *é*.

The Arabic script is not suited to denote all Turkic consonants. Thus, *p* may be represented by *bā* or *fā*, and *č* by *jīm* or *šin*. The problems may be solved by introducing auxiliary signs normally not used: *p* can be denoted by *bā* with a *šin* over it, *č* by *jīm* with three dots over it, etc. The sources are often inconsistent in the use of their own signs. In one and the same text, *ŋ* may be written with *nūn*, with *nūn* plus *kāf*, and with *γain* or *kāf* with three dots. In Armenian texts, one consonant is also often represented by more than one letter, *k* by *kim* and *ke*, *t* by *ta* and *to*, etc.

Morphophonology: Sound Harmony

Due to the limits of the notational systems, syllabic front vs. back sound harmony cannot be examined thoroughly. However, a number of clear cases suggest that it functions, though not consistently. Disruptions occur rather frequently due to a shift **a > e* in the environment of *y*, *č* and *l*. Thus, words containing both front and back vowels emerge, e.g. *ačmáj* [ačmey] 'not opening', *mugreydir* [muŋreydir] 'is lowing'.

Rounded vs. unrounded harmony does not occur consistently. Very often contradictory data coexist, suggesting vacillations such as *könüllük* ~ *könüllik* 'fairness', *oylunuŋ* ~ *oyluniŋ* 'his son's', *sordu* ~ *sordi* '(s)he asked', *qilduq* ~ *qildiŋ* 'we did'.

Morphology

Nouns and Adjectives

Word Formation

Denominal suffixes include *-AK*, e.g. *kesek* 'piece', *qisraq* 'mare', *-ČA*, e.g. *aqča* 'money', *barča* 'all', *-ČI*, e.g. *bitikči* 'clerk', *bašči* 'leader', *yükči* 'superior', *teğirmenči* 'miller', *-DAš*, e.g. *qarındaš*, *ğardaš* 'brother', *tözdeš* 'identical in essence', *-LXK* and *-LUK*, e.g. *arqunluq* 'slowness, gentleness', *aruwliχ* 'purity', *bazliq*, *baziliχ*, *bazluχ* 'peace', *oyurluχ* 'theft', *-LX* and *-IU*, e.g. *atlu* 'horseman', *yelli* 'windy', *bayli* 'tied', *awruqli* 'sick', *borčlu* 'debtor'. The archaic form **-LXG* is also found, e.g. *qorquluq* 'frightening', *eyerliq* 'saddled'.

Deverbal suffixes include *-(V)K*, e.g. *satuχ* 'trading, selling', *artuq*, *artuχ* 'more', *yazoq*, *yazuq*, *yazuχ*, *yazuq*, *yezuq* 'sin', *areχ*, *ariχ* 'thin', *tešik* 'hole'. Representatives of the old suffix **-(*)G* are *-I*, *-U*, *-Iw*, *-Ow*, *-θ*, e.g. *biši* 'ripe, cooked', *tiri* 'living', *öli* 'dead', *sürü*, *sürüw* 'herd', *duri* 'transparent', *tartu* 'scale', *satu*, *satow* 'trade', *keziw* 'epidemic', *ači* 'bitter', *isi* 'hot'. Representatives of the old suffix *-(*)m* are *-(I)m* and *-(U)m*, e.g. *aχim* 'stream, flow', *ičim* 'drink', *yarim* 'half', *ölüm*, *ölim* 'death', *artum* 'very', *tamzum* 'drop'.

A comparative is formed with *-rAK*, e.g. *yaχširaq* 'better' ← *yaχši* 'good', *yamanraq* 'worse' ← *yaman* 'bad', *yiraqraq* 'farther' ← *yiraq* 'far'. The

syntactic structure used to express comparison is illustrated by the sentence *Uzun bolmaqliq qisqadin yaχširaq turur* 'It is better to be long than short'. A superlative is formed by the particle *eŋ* plus the positive or comparative form of the adjective, e.g. *em burun* 'first of all' and *eŋ artıqraq* 'biggest, most'.

Declension

The plural suffix is *-lAr*, e.g. *yuglar* 'feathers', *oqlar* 'arrows', *išler* 'things', *ölüler* 'dead people'. Plural forms are not required after numerals, e.g. *alti yıl* 'six years'.

Among the case markers, the genitive suffix is *-lŋ*, e.g. *teğriniŋ* 'God's', *yekniŋ* 'the devil's'. The accusative suffix is *-nl*, e.g. *qulni* 'servant', *aqčani* 'money', *teğirni* 'God', *kišini* 'man'. The suffix *-l* occurs sporadically as an accusative marker, e.g. *iši* 'thing'. The dative suffix is *-GA*, e.g. *yolya* 'for the road', *toyya* 'for the feast', *qarabusqa* 'to the saddle-bow'. The dative suffix *-A* occurs sporadically, e.g. *ata* 'onto the horse'. The locative marker is *-DA*, e.g. *künde* 'on the day', *ğabaχta* 'at the inn', *uzunluqda* 'in length', *üzegüde* 'on the stirrup'. The ablative marker is *-DAn*, e.g. *yamandan* 'from the evil', *ölümden* 'from the dead', *men tağriden qorqarmen* 'I fear God', *attan* 'from the horse'. The ablative suffix *-DIn* occurs in some Mamlük sources, e.g. *ğamdin* 'of leather', *jilddin* 'of hide', *mingendin sogra* 'after mounting'.

The possessive suffixes are 1p.sg. *-(I)m*, 2p.sg. *-(I)ŋ*, 3p.sg. *-(s)I*, 1p.pl. *-(I)mlz*, 2p.pl. *-(I)ŋlz*, 3p.pl. *-(lAr)l*.

The case markers have special forms after possessive markers. The 'pronominal *n*' precedes the third-person dative, locative and ablative endings. The 3p.sg. accusative and dative suffixes are *-nl* or *-n* and *-nA* or *-ŋA*, respectively. Examples: nominative *qoluŋ* 'your hand', *ayirliyi* 'its weight', *anası* 'his mother', *alamları* 'his flags', accusative *qonşuŋni* 'your neighbour', *atıŋni* 'your horse', *ayaqlarıŋni* 'your feet', *učiŋi* 'its head', *barçasın* 'all of it', *taniχin* 'his witness', *išlerini* 'his things', *tağların* 'his wonders', *ayrıyimizni* 'our illness', *ömekimizni* 'our bread'; dative *sol yanıŋa* 'to your left', *beliŋe* 'to your waist', *kögsine* 'to his chest'; locative *eliŋde* 'in your hand', *qilič qinında* 'in the sword sheath'; ablative *eyeriniŋdin* 'from your saddle'.

Pronouns

The normal forms of the personal pronouns are *men* ~ *mèn* 'I', *sen* ~ *sèn* 'you', *ol* '(s)he, it', *biz* 'we', *siz* 'you', *anlar* 'they'. Less frequent forms are *ben* 'I', *seŋ* 'you', *bizler* 'we', *alar*, *allar*, *onlar* 'they'. The 3p.sg. of the reflexive pronoun, *özi*, *kendözi*, also functions as a personal pronoun in a few Mamlük sources.

Case forms: accusative *meni*, *seni*, *anı*, *bizni*, *sizni*, *anlarıni* ~ *alarıni*; genitive *menim*, *seniŋ*, *anıŋ*, *biziŋ*, *siziŋ* ~ *sizniŋ*, *anlarıniŋ* ~ *alarıniŋ* (sometimes with rounded vowel, e.g. *senüŋ*, *anıuŋ*); dative *maŋa* ~ *maya* ~ *mā*, *saŋa* ~ *saya* ~ *sā*, *aŋa* ~ *ağar* ~ *ayar* ~ *ār*, *bizge*, *sizge*, *anlarya* ~ *alarya*;

locative *mende, sende, anda, bizde*; ablative *menden, senden ~ sendin, andan, bizden, sizden, anlardan*.

Demonstrative pronouns are *bu* 'this', *ol* 'that', *oşbu* 'this other thing' and *o şol* 'that other thing'. Reflexive pronouns are *öz, kendi ~ kendü, kendi öz(i)* and its contracted form *kensi*. Interrogative pronouns, which can also function as relative pronouns, include *kim* 'who', *ne* 'what', *qayu ~ qaysi* 'which', *qaç* 'how many'. Indefinite pronouns usually consist of interrogative pronouns followed by enclitics, e.g. *kimse* 'someone', *neme, nesne, neste* 'something', *barça* 'everyone, all'. The numeral 'one' may serve as an indefinite pronoun and also precede the forms cited, e.g. *biri ~ birisi* 'someone', *bir neçe* 'a few'.

Numerals

The cardinal numbers are *bir* 'one', *ek(k)i* 'two', *üç* 'three' etc. Ordinal numbers are formed with *-(I)nçI*, e.g. *birinçi* 'first', *ek(k)inçi* 'second', *üçinçi, üçünçi, üçünçü* 'third'. Other words for 'first' are *awwal, burun* and *burunyî*.

Collectives are formed with *-Aw ~ -Ow < *-AγU*, e.g. *ëkkew* 'two together', *üçew, üçöw* 'three together', distributives with *-(ş)Ar*, e.g. *altışar* 'six each'. Approximative numbers are expressed by juxtaposing two numerals, e.g. *ekki üç* 'two or three'. Multiplication is expressed with the words *qat, kez, qurla* and *qatla* 'times'.

Verbs

Verb Derivation

Denominal verbs may be formed with *-A*, e.g. *yulqa-* 'tear out', *boşa-* 'loosen', *oyna-* 'play', *sana-* 'count', *-IA*, e.g. *bayla-* 'tie', *baluqla-* 'fish', *boluşla-* 'help', and *-dA*, e.g. *alda-* 'trick', *izde-* 'look for, search for'.

As for deverbal verbs, the most frequent causative suffixes are *-DUr*, *-(I)t* and *-Ur*. Reflexes of the old suffixes *-Ar*, *-z* and *-GUr* occur sporadically, while *-Glz* is relatively frequent only in Armeno-Kipchak. Examples: *açtur-* 'open', *biştür-* 'let ripen', *keltür-* ~ *keltir-* 'bring', *sındur-* 'tear apart', *bildir-* 'make known', *arıt-* 'clean', *aqüt-* 'slide, move', *bişür-* ~ *bişir-* 'cook', *batur-* 'sink, hide', *yatur-* 'lay down'. The cooperative-reciprocal suffix is *-(I)ş ~ -(U)ş*, e.g. *alış-* 'exchange', *boluş-* 'help', *satış-* 'trade with each other'. The suffixes *-(I)l* and *-(I)n* form passive, reflexive and middle verbs, e.g. *açıl-* 'open', *beril-* 'be given', *bilil-* 'be known', *çewril-* 'be turned', *yazıl-* 'be written', *ayrın-* 'be separated', *kerin-* 'be spread out', *awurun-* 'feel pain', *sürtün-* 'scratch oneself', *körin-* 'be seen'.

The productive suffixes also include *-(X)k*, denoting physical conditions of the body, e.g. *açiq-* 'starve', *erik-* 'feel disgust', *turuq-* 'stop, be in a state'. The suffix *-A*, absent in Old Turkic, has iterative meaning, e.g. *aça-* 'often (constantly) open', *kese-* 'often (constantly) cut'.

In compound verbs, the nominal part is mostly of foreign origin, and the verbal part *et-* or *qil-* 'do, make', e.g. *aqın et-* ~ *aqın qil-* 'attack', *namâz qil-*

'pray'. Examples with other verbs are *ögünç ber-* 'compliment', *tım bol-* 'quieten down', *könül aç-* 'console', *boş tur-* 'be idle', *ayır bol-* 'become a burden', *ayır kel-* 'be heavy'. In Armeno-Kipchak texts, infinitives mostly function as the nominal part, e.g. *zaymit et-* 'grab, get hold of', *vykupit et-* 'pay ransom, redeem', *zvolitca bol-* 'break free'.

Finite Verb Forms

Finite verbs take on personal markers of two types. The pronominal type has the forms 1p.sg. *-m(en)*, pl. *-(bI)z*, 2p.sg. *-sen ~ -señ*, pl. *-slz*, 3p.sg. *-Dir*, pl. *-(Dir)lar*. The possessive type has the forms 1p.sg. *-m*, pl. *-K*, 2p.sg. *-ñ*, pl. *-slz*, 3p.sg. *-ø*, pl. *-(lar)*.

A general present which can also refer to foreseen events is formed with the aorist stem plus personal markers of the pronominal type. The form varies across texts.

One type is *-(V)r*, e.g. *Alişirmen* 'I exchange', *Basarmen* 'I press', *Qanča barursen?* 'Where are you going?', *Arqun arqun kelür* '(S)he is coming very slowly', *Tiri bulax andan ayar* 'The source of life flows from there', *Köglüm aynır* 'I am excited', *Arqası açiq qalır* 'His back will be left exposed', *Batar* 'It sinks'. While the variants *-Ar* and *-Ur* have various distributions in the texts, *-Ir* is found sporadically in the German part of the *Codex Cumanicus*. An aorist in *-V* occurs in the Codex after verbs ending in *-r*, e.g. *baru* 'goes'.

A second type, occurring in a few Mamlük sources, is *-yVr* after stems ending in vowels. The labial vowel is assumed to be *U*, though its height is actually unknown, e.g. *başlayur* 'starts', *sözleyur* 'speaks'.

Negation is formed with *-mAs*, e.g. *Bermesmen* 'I do not give'. The first-person singular form may also be contracted, e.g. *Bermen* 'I do not give'. This is the standard Armeno-Kipchak form, e.g. *klemen* 'I do not want', *bilmen* 'I do not know', *bolman* 'I will not be'.

In yes/no questions, the interrogative particle is placed between the aorist suffix and the personal marker, e.g. *Sen türkiçe bilürmisen?* 'Do you know Turkic?'

Only a few examples are found of a focal present formed by combining the *-A* converb with the contracted form of an auxiliary verb, *yür* 'is moving' or *tur* 'is standing', e.g. *Beriyirmen* 'I am giving', *Keteyürmen* 'I am going away', *Başlaydırlar* 'They are starting'. Negation is formed from the negative verb stem, e.g. *Kelmeydirseñ* 'You are not coming'.

The optative in *-GAy* expresses modality including future time reference, e.g. *Anča teğeri anı artux sövgey* 'Thus God may/will love him even more', *Hiç kimersege bildürmegey* 'One should not tell anybody'. It is sometimes close to the imperative in meaning, e.g. *Bir yaxşı ayaçdın süñü yaraşturayay sen* 'Make a lance from a suitable branch'.

The most frequent finite form is the simple past in *-DI* plus personal markers of the possessive type, e.g. *Ölümden bizni çiyardıñ* 'You have released us from death', *Su qan bile ayızdırdı* 'He mixed water with blood',

Çiqtılar İerusalemđan 'They withdrew from Jerusalem'.

A focal post-terminal with resultative functions – indicating past events and their result in the present – is formed with the *-(I)p* converb plus pronominal personal markers, e.g. *Kök barçe açılıptur* 'The skies have opened completely', *Dünyeye tirelik berilipturur* 'The world has been given life', *Ayaç iliştir* 'The tree is bent down'.

The conditional in *-sA* is usually accompanied by *eger* 'if' or *qačan* 'when', e.g. *körseñ* 'if you see', *eger qačsa* 'if (s)he flees'.

Ability is expressed by the *-A* converb followed by *al-* 'take' or *bil-* 'know, be able to'. The converb suffix may be lost through contraction. Examples: *Köktegi xanlıqta baralmazbiz teğeri qattında* 'In the heavenly kingdom we will not be able to stand before God', *Tağların aya bilipsen* 'You have been able to proclaim his wonders', *Kişi yolsuz barabilmez* 'One cannot move on without a road'. Necessity is indicated by periphrastic forms with the modal word *kerek*, e.g. *Aş bişyanda qotarmek kerekxen* 'After the meal is cooked, you must serve it', *Kerek kim biz köñül közi bile baqqaybiz* 'We must look with the eye of the heart'.

The texts contain a variety of optative–imperative forms. In the second-person singular, *-gXl* or *-gXn* is used, whereas no example of the bare verb stem is found. The first-person plural form is indicated by *-AllK*, *-Allm* or *-All*, the second-person plural by *-(y)lğ(lz)*. The use of *-lAr* in the third-person plural is optional. Examples: 1p.sg. *bereyim* 'I should give', 2p.sg. *Ayrıymizni oğaltıyl* 'Heal us of our illness', *Baryıl* 'Go!', 3p.sg. *Ayrıtmasun* 'Let it not hurt', *Keğ bolsun* 'Let it be wide', *Teğeri bersin kensi bazluxın* 'May God grant him peace', *Boluş kim baralı* 'Help us to go', 2p.pl. *Baruğiz körgüzüñiz papazlarya* 'Go and present yourselves to the priests', 3p.pl. *Kirmesünler* 'They should not attack'.

Non-finite Verb Forms

The aorist participle in *-(V)r* is used as a non-finite form. The converb in *-A* (after consonants)/*-y* (after vowels) is widely used. Some texts exhibit the archaic variant *-yU*, corresponding to the aorist variant *-yUr*. The ending *-yIn* is often added to negative stems. The converb suffix *-(V)p* is not represented in all texts. One text exhibits the postvocalic variant *-(y)Vp*. The suffix *-(y)VbAn* occurs in one source.

Adverbs

The most frequent suffixes forming adverbs are *-lAy*, e.g. *bulay*, *alay* 'this way' and *-lAyIn*, e.g. *yamyurlayın* 'like the rain'. Some adverbs are fossilised nominal forms with case endings, e.g. *qışın* 'in winter', *artqaru* 'backwards'. Others are formed with productive markers, e.g. *munda ~ bunda* 'here', *mundan* 'from here', *anda ~ onda* 'there', *anča* 'like this'. Derivatives of the interrogative pronoun *ne* are *nečik* 'how, in what manner', *neče* 'how many, how much', *neden ~ nelik* 'why' etc.

Other Morphological Categories

Postpositions include *başqa* 'other than, except' + ablative, e.g. *ulu künden başqa* 'except on holidays', *bile* 'with' + nominative, *egi köñül bile* 'with good grace', *deyri* 'up to, till' + dative, e.g. *Yerosolimya degri bardı* 'They went all the way to Jerusalem', *teğrü* 'to, towards', 'as far as', *deyin* 'till' + dative, *kibi* 'like, as' + nominative, *köre* 'with respect to, in accordance with', *ulam* 'across, through, by' + ablative, e.g. *Andan ulam bar barča bolyanturur* 'Everything has been created by him', *üčün* 'for, because of, in order to' + nominative, e.g. *yazuq üčün* 'because of sin'. Many simple postpositions are fossilised converb forms.

The most frequent conjunctions include the connective *dağı ~ tayı ~ tağı ~ dayın* 'and', *wa* 'and', the disjunctive *ne ... ne* 'neither ... nor', *yā ... yā* 'either ... or', *tileseñ ... tileseñ* 'either ... or', the adversative *walakin* 'but', *a* and *na* 'but' (of Slavic origin), the conditional *eger* 'if' and the causal *anıñ üčün (kim)* 'because', *čünki* 'because'.

The particles *dA* and *OK* emphasise the preceding word, e.g. *tez ök* 'very quickly', *tutyač oq* 'as soon as (s)he seizes it', *Baryın oq* 'Just go!' The interrogative particle has the forms *-mU*, *-mI* and *-mA*. Negated copula particles are *dewül*, *tögül*, *dögül* 'not' and *yoq* 'non-existing', the counterpart of *bar* 'existing'.

Lexicon

The vocabulary of the *Codex Cumanicus* is essentially of Turkic origin. Loanwords include a small number of Slavic, Greek and Middle Mongolian origin and a greater amount of Persian words and Arabic words through Persian mediation. Mamlük sources contain a higher proportion of loanwords, mostly from Arabic. Most loanwords in Armeno-Kipchak texts are of Slavic origin, while Persian words and Arabic words through Persian mediation only constitute a minor part. The number of Armenian words in Armeno-Kipchak is rather limited.

Further Reading

- Berta, Á. (1996) *Deverbale Wortbildung im Mittelkiptschakisch-Türkischen*, Turcologica 24, Wiesbaden: Harrassowitz.
- Drimba, V. (1973) *Syntaxe comane*, Bucureşti: Editura Academiei, Leiden: Brill.
- Drüll, D. (1979) *Der Codex Cumanicus. Entstehung und Bedeutung, Geschichte und Gesellschaft*, Bochumer Historische Studien 23, Stuttgart: Klett-Cotta.
- Gabain, A. von (1959) 'Die Sprache des Codex Cumanicus', in J. Deny et al. (eds) *Philologiae Turcicae Fundamenta* 1, Wiesbaden: Steiner, 46–73.
- Grønbech, K. (1942) *Komanisches Wörterbuch. Türkischer Wortindex zu Codex Cumanicus*, Monumenta Linguarum Asiae Maioris, Subsidia 1, Copenhagen: Munksgaard.
- Pritsak, O. (1959) 'Armenisch-Kiptschakisch', in J. Deny et al. (eds) *Philologiae Turcicae Fundamenta* 1, Wiesbaden: Steiner, 81–7.

9 Chaghatay

Hendrik Boeschoten and Marc Vandamme

Introduction

The term Chaghatay is fraught with inconsistencies. Orientalists have often applied it to any form of written Turkic used in an Islamic context in Eurasia outside the Ottoman Empire from the thirteenth century up to World War I. In contrast, as a linguistic term used by some authors from the seventeenth century onwards, Chaghatay has a more restricted meaning both historically and geographically, referring to the high literary language of the classical period of the fifteenth and sixteenth century.

Chaghatay can be defined as a succession of stages of written Turkic in Central Asia. In many respects it is also a continuation of earlier stages, notably of Karakhanid Turkic, with Khwarezmian Turkic as a transitional stage. It cannot be defined as a fixed entity in time and space. Chaghatay sources are a hybrid collection of texts written in a vast territory in Islamic Eurasia by speakers of different varieties of Turkic, who from the late fifteenth century onwards more or less tried to focus on a specific model known as Classical Chaghatay.

The Formation of Chaghatay

Chaghatay is the name of Chingis Khan's second son, who in 1227 inherited as his realm (*ulus*) the Central Asian part of the Mongol empire except Khwarezm. After an unstable period following Chaghatay's death, Timur Lenk (Tamerlane) came to power in 1405 and took a firm hold of Transoxania. As a result of his conquests, the literary language that came to be called Chaghatay spread over the entire Turkic-speaking Central Asia. In the Ottoman Empire its impact remained limited, because a local literary language was developing in Asia Minor. The Timurids also promoted the use, in their chanceries, of Khwarezmian Turkic written in the old Uyghur script. The corpus of texts ensuing is known as *yarliq* documents. This policy also resulted in the production of several works of Islamic literature in Uyghur script during the fifteenth century.

Several historical developments had important consequences for the formation of the literary language, e.g. the break-up of tribes into multi-ethnic units and the growing impact of Islamic civilisation accompanied by an

increasing orientation towards Persian urban culture. The centres of literary activity kept shifting: first from Kashgar to the valley of the Surkhan Darya and Khwarezm, then to Samarkand, Herat in the south and the Ferghana valley in the east, then to northern India in Bābur's time, etc. With every move, the language picked up local traits. Owing to the complex history of ethnic mingling of the Turkic-speaking population and the shifting geographical centres it is impossible to establish a specific 'dialectal basis' for Chaghatay. Generally speaking, the geographical shift to the west in the initial stage of Khwarezmian Turkic caused a certain Oghuz influence, which includes lexical items, vowel length in the early period, the first-person copula *-(y)Am*, the optative in *-(y)A*, use of the converb in *-bAn* and occasional loss of suffix-initial *G* in poetry. From then on, there was also an increased influx of Kipchak features including phonological properties and the replacement of the participles *-miš* and *-duq* by *-GAn*. An exact matching of dialects with tribes seems impossible and would even be incorrect.

The impact of Persian was very strong and reinforced by widespread bilingualism not only among the elite. The analysis of the actual situation is complicated by the fact that so much of Chaghatay literature consists of translations from Persian. The influence of Mongolian was restricted to a number of loanwords from the domains of warfare and administration.

In studying the language on the basis of existing documents, different sources of internal variation must therefore be taken into account, e.g. the influence of spoken varieties and of Persian elements. An important fact is that the court poetry has a central place in the Chaghatay text corpus. Cases of internal variation were always numerous, as the language was only partly standardised.

The Periods of Chaghatay

It seems that the formative period can well be extended to the Khwarezmian Turkic stage. It thus includes, the fourteenth-century Rabghuzi's *Qisasu'l-*anbiyā** 'Stories of the prophets', a work in prose with the inclusion of poems, the prose work *Nahju'l-farādis* 'The right way to the gardens of Paradise', and the 'early Chaghatay' poets from around 1400, e.g. Atā'i, Sakkākī and Lutfī. The classical period begins in the second half of the fifteenth century in Samarkand and Herat with the writings of the famous poet 'Alī Šēr Navā'i, whose language has since then been the model for high style poetry and prose. Another landmark in the development is the prose and poetry of the Timurid ruler and first Moghul emperor Bābur, notably his memoirs *Bāburnāma*, written in the early sixteenth century. Another ruler, 'Abu'l-Γāzī Bahādur Khan, wrote the treatises *Šajara-i tarākima* 'Genealogy of the Turkmens' and *Šajara-i türk* 'Genealogy of the Turks' in the middle of the seventeenth century, i.e. after the classical period.

From now on, the written language used an increasing number of elements

from the spoken language. Finally, the local varieties developed into separate languages, a process that was completed after the October Revolution in 1917. In terms of literary production, the seventeenth and eighteenth centuries are often seen as a period of decay. It is a period in which Chaghatay loses ground to Persian. The change in attitudes can be illustrated by the views expressed by two important authors. In his treatise *Muhākamatu'l-luġatayn* 'Judgement of two languages', 'Alī Šēr Navā'ī shows that 'Türki' is by no means inferior to Persian, and exhorts poets to write in it. 'Abu'l-Fāzī is motivated by functional considerations and describes his choice of language and style in the sentence 'I did not use one word of Chaghatay (!), Persian or Arabic'. As is clear from his actual language use, he aims at making himself understood to a broader readership by avoiding too ornate a style, notably *saj'*, rhymed prose.

The literary language of the focal area of Chaghatay no doubt represents previous stages of modern Uzbek and Uyghur. The continuation of the Islamic literary language in the east, the *Kāšġar tili*, and its interrelations with the language in Western Turkestan are poorly researched. The historical relationship of other modern languages with Chaghatay is even more problematic. Arguments for linking Uzbek to Chaghatay as its forerunner are specific cases of Chaghatay variation that are still found in Uzbek dialects. Some of them even date back to the Khwarezmian Turkic stage, e.g. *keyin* ~ *soġ* 'after', *-dimiz* ~ *-duq* for the first-person plural of the simple past.

Schematically, the historical situation can be represented as follows:

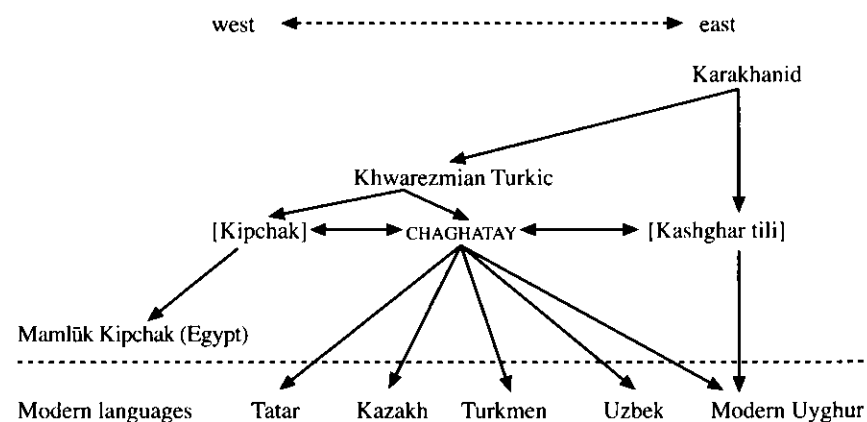


Figure 9.1 Relations of Chaghatay with other written varieties

The Study of Chaghatay

In the twentieth century, the study of Chaghatay has suffered from nationalist bias. In the former Chaghatay area, separate republics have been claiming

Chaghatay as the ancestor of their own brand of Turkic. Thus, 'Old Uzbek', 'Old Uyghur', 'Old Tatar', 'Old Turkmen' and a Chaghatay-influenced layer in sixteenth-century Azerbaijanian have been studied separately from each other. There has been a tendency to disregard certain characteristics of Chaghatay itself, e.g. its complex syntax copied from Persian.

Much more research is needed, not only with the aim to anchor modern national languages in history. The diachronic layers must be sorted out, and, in doing this, more attention should be given to texts that do not belong to the canon of high style prose and poetry. Features emerging from interaction with Persian are of special linguistic interest, particularly because most authors can be assumed to have been bilingual. Another urgent need is a dictionary of Chaghatay or of east Middle Turkic, since the existing dictionaries are all over a century old.

The survey below will focus on characteristics of the classical period of Chaghatay, with occasional references to the Khwarezmian stage. The description can only be approximate, since much of the variation found in the texts must be suppressed.

Sound System

Diachronic Phonetic Facts

The shift $\delta > y$ is connected with old dialect differences and dates back to the Old Turkic period. Khwarezmian Turkic and early Chaghatay show a fair amount of corresponding words with z . This z disappears later, but there are still traces of δ as late as in the sixteenth century. Examples: Karakhanid *aġaq* 'foot', Khwarezmian Turkic *aġaq* ~ *azaq* ~ *ayaq*, Chaghatay *ayaq*. This is, at least partially, a result of dialect levelling.

As for the shift $(*b >) w > v$, Khwarezmian Turkic has w , e.g. *aw* 'hunting', whereas Chaghatay has v , e.g. *av*. There are occasional shifts $p > f$, e.g. *tufraq* < *tupraq* 'earth'. The general confusion of f and p points to the lack of a phonemic distinction between them, at least in some spoken varieties, e.g. *fitildi* ~ *bitildi* 'was written', *fēšīn* ~ *pēšīn* 'morning prayer'. There is a widespread assimilation of $*b-$ to $m-$ before nasal consonants, e.g. *meġze* 'resemble', *munda* 'here'. In specific cases, $*t-$ develops into $d-$, e.g. *de-* 'say', *deg* ~ *teg* 'like', *dayi* 'and'.

The spelling of word-final $-q$ and $-γ$ is increasingly inconsistent, e.g. *sariy* ~ *sariq* 'yellow' < *sariy*, *artuq* ~ *artuy* 'more' < *artuq*. This points to word-final neutralisation between q and $γ$, and, by implication, between k and g . Regressive labial assimilation $e (> \acute{e}) > \acute{o}$ and $a > o$ is widespread, e.g. *oruq* ~ *aryu* 'lean', *ötük* 'boot'.

Phonotactics and Morphophonology

Front vs. back harmony is, with rare exceptions, in force. Copies of Arabic and Persian words tend to take on back suffix variants, e.g. *čamanyā* 'to the

meadow', *yānīlarya* 'to the rich'. Exceptions are found in words containing *k* or *g*, e.g. *gülge* 'to the rose', *pakrek* 'purer'.

Rounded vs. unrounded harmony is observed in some suffixes. In some cases the situation varies across diachronic stages of the language. A phonetic characteristic of Khwarezmian Turkic is that certain suffixes assimilate to stem-final labial consonants, e.g. *sewünč* 'happiness', *yemüm* 'my food', 'Arabnuğ 'Arab + genitive', *čapup* 'galloping'. This is not a transitional phenomenon, since in Chaghatay proper the labialisation is restricted to stems ending in *-v* and loanwords containing *av*, e.g. *evüñ* 'hurry up!', *qavlug* 'your word'.

Morphology

Nouns and Adjectives

Declension

The plural suffix is *-lAr*. The possessive suffixes are *-(I)m* 'my', *-(I)ñ* 'your', *-(s)I* 'its/his/her', *-(I)mlz* 'our', *-(I)ñlz* 'your' (plural and honorific), *-lArI* 'their'.

As for case suffixes, the genitive suffix *-nlğ* seems to obey the fourfold vowel harmony in earlier times, but tends to be restricted to the unrounded variants *-nlğ* from the sixteenth century on. The accusative suffix is *-nl*, after the third-person possessive suffix also *-n*. Already in Navā'ī's time, cases of coincidence of genitive and accusative suffixes are found, e.g. *oqın* 'of his arrow' instead of *oqınığ*. The dative suffix is *-GA*, i.e. *-ya* and *-ge*, after voiceless consonants mostly *-KA*, i.e. *-qa* and *-ke*. The locative suffix is *-DA*, and the ablative suffix is *-DIn*, though the Kipchak and Oghuz form *-DAn* also occurs. After voiceless consonants these suffixes are often written with an initial *t*. The use of the accusative is confined to specific direct objects, though the distribution is not quite the same as in Turkish.

The equative suffix *-čA* may behave as a word-formation suffix, e.g. *özgeče* 'differently', *miñče* 'about one thousand', but is also attached to case suffixes, e.g. *tört kündeče* 'in about four days', *qiyāmatyača* 'until the Day of Judgement', and to possessive suffixes, e.g. *imkāni bariča* 'as long as possible'.

The 'pronominal *n*', in other Turkic languages generally inserted before case markers following the third-person possessive suffix, is mostly lost, but it has, for example, been retained in 'Abu'l-Γāzī's language.

Pronouns

The personal pronouns are *mēn* 'I' with the genitive *mēniğ* and the dative *mağa*, *sēn* with the genitive *sēniğ* and the dative *saga*, *biz(ler)* 'we' with the genitive *biz(n)iğ* and *siz* 'you' (plural and honorific) with the genitive *siz(n)iğ*, *sizler* 'you' (plural).

Demonstrative pronouns are *bu* 'this' with the oblique stem *mun-*, plural *bular*, and *ol* 'that' with the oblique stem *an-* and the plural *alar*, in the pre-Navā'ī period *olar*. Both can be provided with the deictic particle *uš*, yielding *ušbu* and *ušol*. For anaphoric and cataphoric reference, *bu* is normally used. *Ol* is used as functional equivalent of a third-person personal pronoun.

Reflexive pronouns are formed with possessive suffixes added to *öz*, e.g. *Özüñni xāmūs eylegil* 'Keep silence' ('Silence yourself'). Persian *xūd* is also used, e.g. *mēn xūd* 'I myself'. *Öz* used attributively means 'own', e.g. *öz hayātida* 'in his own life'. Reciprocal pronouns are formed with *bir(i) biri*, e.g. *biri birin körüp* 'seeing each other'. The most important interrogative pronouns are *kim* 'who', *ne* 'what', *qay(u)* 'which', *qaysi bir* 'which'.

Numerals

For ordinal numerals there are two suffixes, *-(I)nčI* and *-lAnčI*, e.g. *bēšinči* ~ *bēstenči* 'fifth'. Collective numerals are formed with *-Av*, or with *-AlA(sI)*, e.g. *ikev* (in Khwarezmian Turkic still *ikegü*) ~ *ikelesi* 'they both', *törttev* ~ *törttele* 'the four of them'. Distributives are formed with *-(š)Ar*, e.g. *üčer* 'three each', *altışar* 'six each'.

Copulas

There exist various classes of copulas in Chaghatay. The personal pronouns may, in cliticised form, be used as copula suffixes with present meaning, e.g. *Čečeksēn* 'You are a flower'. No copula is required in the third person, e.g. *Köñli özge, tili özge* 'His/her heart is one thing, his/her tongue another'. The use of the demonstrative pronoun *ol* as a copula became largely obsolete in the Khwarezmian Turkic period. The copula suffixes are mostly attached to the form *-Dur*, sometimes *durur* ~ *turur*, aorist of *tur-* 'stand', e.g. *Aniğ ati Allah turur* 'His name is Allah', *Qulugdurmēn* 'I am your servant'.

A number of forms of the defective auxiliary verb *ē(r)-* are used: a past copula *ērdi*, e.g. *Yükli ērdi* 'She was pregnant', a present copula *ērür*, e.g. *Ērürsēn šaftıq* 'You are compassionate', an inferential copula *ē(r)miš*, e.g. *İnim, bahādur ēmišsiz* 'Brother, you are a hero!', and a conditional copula *ērse*, e.g. *Eger sēn "āšiq ērset, sözleme hēč* 'If you are a lover, do not speak at all'. The negative aorist *ē(r)mes* is generally employed for negation of nominal predications, e.g. *Oğuz Xān neslindin ērmesler* 'They are not from the line of Oghuz Khan', *Rūh ēmes ēdi tan birle āšnā* 'The soul was not acquainted with the body'.

The forms *ē(r)ken* ~ *ē(r)kin* express doubt and supposition, e.g. *Sēndin ēkendur bu gunāh* 'This must be your fault', *Ne hālat ērkin bu?* 'What situation might this be?'

Existential sentences and possessive constructions, are formed with *bar* 'there is', negation *yoq*, e.g. *Kişi yoq* 'Nobody is there', *Xabariğ bar mu?* 'Are you informed?'

Verbs

The verb derivation processes are similar to the ones in other Turkic languages. There are also actional constructions with converbs and auxiliary verbs. The converb in *-A*, with *-y* after vowels, may combine with auxiliary verbs, adding various actional and modal meanings to the main verb, e.g. *İnana almas* '(S)he cannot believe it', *söz ayta bilmey* 'unable to say anything', *Ala körsün* 'Let him try to get it', *Xatun qıçqıra başladı* 'The lady began to shout', *Yayı qaça berdi* 'The enemy ran away quickly', *Bu vilayatlar qadimdin Türkke bola kélgendur* 'These provinces have belonged to the Turks for ages', *Yol başında tura qaldı* '(S)he stopped on the way', *Farq bola yazdı* '(S)he almost drowned'. The converb in *-(I)p* is also used with auxiliary verbs, e.g. *Qaçıp bar-* 'Run away', *Alıp kél-* 'Bring!', *Atı harip qaldı* 'His horse became tired', *Bayırların söktürüp taşlatur érdi* '(S)he used to have their livers torn out'. Negated *bol-* may denote impossibility, e.g. *Anlap bolmas édi* 'It was impossible to understand'.

Finite Verb Forms

A focal present tense is formed with the *-A* converb followed by *-dur* and copula suffixes, e.g. *Tapadurmén* 'I find, I am finding', *Tégmeydur* 'It does not reach'. At later stages, *-dur* is often dropped in the first and second person, e.g. *Bu né kişidur ki munı életesiz?* 'Who is this person you are leading?'

The so-called aorist is formed with *-(V)r*, the vowel being *-A* or *-U*; the negation is *-mAs*. The old form *-yUr* is used up to the fifteenth century. The aorist is a more general present tense, also used with modal meanings and future time reference, e.g. *Satranjnı xüb oynar* '(S)he plays chess well', *Emdi bilmesmén* 'Just now I do not know', *Mén kélürmén* 'I will come'.

The simple past is marked with *-d*, mostly *-t* after stem-final voiceless consonants. The personal markers are identical to the possessive suffixes. However, *-UK* is used in the first-person plural, where Khwarezmian Turkic still has *-(I)mUz*. There are second-person plural variants with the nominal plural marker *-lAr*, e.g. *Dédiñiz ~ dèdiñler ~ dèdiñizler* (seldom) 'You (plural) said'.

A post-terminal with resultative and perfect functions is formed with the *-(I)p* converb, *-dur < turur* and copula suffixes, e.g. *Yürekim qan tolupdur* 'My heart is filled with blood'. The form is seldom negated, instead, *-mAy* is used, e.g. *Kélmaydur* '(S)he has not come'. Later *-dur* is often dropped in the first and second person, e.g. *Emdi bolupmén gadā* 'Now I have become a beggar'. A more focal post-terminal with resultative nuances is formed with *-GAn* plus pronominal copula, in the third person *-turur ~ -dur*, e.g. *Kitābı dèrler kim jam qilyanturur* 'They say (s)he has composed a book'.

The old perfect in *-mİš* occasionally occurs as an inferential, e.g. *Ma'lum boldı kim xabardār bolmıšlar* 'It became clear that they had been informed'.

The 2p.sg. imperative is formed with the bare stem, or, more frequently, by adding the suffix *-Gıl* or the variant *-GIn*, e.g. *Yemegil* 'Do not eat!', *Alyın* 'Take!', *Éšitgil ~ Éšitkil* 'Listen!' A plural and honorific imperative is formed

with *-(I)ñ(lAr)* or *-(I)ñIz(lAr)*, e.g. *Körüñ* 'See!', *Soruñızlar* 'Ask!' The basic third-person form is *-sUn(lAr)*, e.g. *Barsun* 'Let him/her go', *Kirsünler* 'Let them enter'. The variant *-dék* is characteristic of Navā'ī and a few other writers, e.g. *Sordékler* 'Let them ask'. The singular voluntative has the variants *-(A)y* and *-(A)yIn*, e.g. *Bitiy* 'I will write', *Barayın* 'Let me go'. The plurals are *-(A)llıñ*, e.g. *Oqulı* 'Let us read', *İçeliñ* 'Let us drink'.

An optative is formed with *-GAy* (*-gay*, *-gey*, after voiceless consonants mostly *-qay*, *-key*) plus copula suffixes. It has various modal meanings, e.g. *Sulh qilyaysiz* 'You will make peace', *Quryandın daryā bir oq atımı bolıyay* 'The river must be one bowshot from the fortress'.

A prospective is formed from the verbal noun *-GU* with possessive suffixes, to which *-dur ~ -durur* may be added, e.g. *Bar édi, bardurur u bolıyusı hem* '(God) was, is and will be', *Tüşüñge dèğümdurur ta'bār* 'I will interpret your dream'.

The conditional is expressed with *-sA* and personal markers of the possessive type, with *-K* in the first-person plural, e.g. *āsāyış tileseñ* 'if you wish tranquillity'. The protasis of counterfactuals is formed by combining the conditional with a past copula, e.g. *Eger sèn müsülmān bolsañ érdi sèni alur érdim* 'If you were a Muslim, I would take you'.

Possibility can be expressed by combining the conditional with *bol-*, e.g. *Anıñ ahdiya i'timād qılsa bolmas* 'One cannot rely on his promise'. The conditional + *kerek* may yield necessitative meaning, e.g. *Her miñ yılda qırq arqa ölse kerek* 'In every thousand years forty generations should pass by'.

The finite items combine with copulas formed from *é(r)-* and *bol-* 'become, be'. Present and aorist stems combined with *é(r)di* or other past copulas yield past items presenting an intraterminal perspective, e.g. *Kèçe tahārat qıladur édim* 'I was performing an ablution at night', *Satranjnı bisyār oynar édi* '(S)he played chess well' (habituality), *Qadım Türk tilinde içi qovuş ayaçnı Qıpçaq dèrler érgendurlar* 'In Old Turkic they seem to have called a hollow tree Kipchaq'. There are pluperfects in *-GAn érdi* and *-(I)p érdi*, e.g. *Miñ atlıq yiyiñyan érdi* 'A thousand horsemen had assembled', *Bu bir iş érdi kim andın burun hèç padšāh qılmaydur érdi* 'This was an act no king before him had performed'. Other forms are *Yibergüm érdi* 'I was going to send', *Kélgen érmiş ol Masih* 'That Messiah has reportedly arrived', *Kèldük èse, tapmaduq anı* 'When we came, we did not find him' (conditional with temporal meaning), *Yüsuf öter bolsa* 'When Joseph passes by'.

Adverbs

Some adverbs carry the instrumental suffixes *-(I)n* and *-lA*, which are unproductive in Chaghatay, e.g. *baş töpen* 'headfirst', *tağla* 'tomorrow'. Interrogative adverbs are *nège*, *nèdin*, *nayı* 'why', *nèçe* 'how, how much, many', *nèçük* 'how', *nèçün* 'why', *qay(u)* 'which', *qaysı bir* 'which', *qaçan* 'when', *qanı*, *qanda*, *qayda* 'where', *qayan(γa)* 'where (to)', *qandın*, *qaydın* 'from where'.

Other Morphological Categories

Examples of postpositions governing the nominative are *bu fursāt ara* 'at this time', *‘ālam içre* 'in the world', *sol qol sari* 'to the left', *çandaq tapa* 'up to the moat', *taht üze* 'on the throne', *mèn birle(n) ~ bile(n)* 'with me', *quyaş dæg yüzi* 'his sun-like face', *altun kibi* 'like gold', *dur mægizlig* 'pearl-like', *bu yaqlıy* 'like this, such, so', *miñ çaqliq* 'about one thousand' ('as much as'), *teñge üçün* 'for money'.

Examples of postpositions governing the dative are *tañga dëgin ~ tégi* 'until dawn', *qiyāmat bazmiya tegrü* 'till the meeting of resurrection', *javhariya köre* 'according to its substance', *mihräbya qarşu* 'opposite the Mihrab (mosque niche)', *Atasiya utru çiqti* '(S)he went out to meet his/her father'.

Examples of postpositions governing the ablative are *qadimdin bəri* 'since the old times', *baridin burun* 'before all', *nêçe kündin këyin* 'after some days', *üç kündin soğ(ra)* 'three days later', *Xudäydin özge* 'other than God', *sëndin ögün* 'besides you'.

Many postpositional phrases are possessive constructions containing nominal cores. The preceding nominal is in the nominative or the genitive. Examples: *sunıñ içige* 'into the water', *şahr taşıda* 'outside the town', *tahtnıñ al(l)ıda ~ anlida* 'before the throne' (← *alın* 'forehead'), *ileyleride* 'before them', *quryanınıñ astıdın* 'from under its fortress', *buland jar üstide* 'over a deep abyss', *ata qatıya* 'to the father', *çan qaşıya* 'to the khan', *Husaynnıñ basıya* 'towards Husain', *Samarqand bile Kêş arasıda* 'between Samarkand and Kesh', *quyaş tegréside* 'around the sun', *su ornıya yay* 'oil instead of water'.

A number of prepositions have emerged from the conventionalised use of the *izāfat* after certain nouns, e.g. *misl-i sahāb* 'like a cloud', *Fayr-i ölmek yoq êrdi tadbiri* 'There was no way out but to die', *taraf-i Türkistan* 'to Turkestan' (see p. 175).

Phrases and clauses can be linked by certain coordinative conjunctions, e.g. *taqi ~ dayi* of Turkic origin, *ve, u* 'and', *hem* 'also', repeated forms such as *hem ... hem* 'both ... and', *nê ... nê* 'neither ... nor', all copied from Persian. Disjunctive and adversative conjunctions are also mostly copied from Persian or Arabic, e.g. *veyā, yoqsa, yoq ki* 'or', *çāh(i) ... çāh(i), yā ... yā* 'either ... or', *ammā, lëkin, veli, velëkin* 'but', *belki* 'but, on the contrary'.

Frequently used particles are the deictic *muna, ana* and *uş* 'here!, now!, look!' and the corroborative particle *OK*, e.g. *Munça oq bolıyay tafāvut* 'That is precisely the difference'. The clitic *-çI* serves to form polite imperatives, e.g. *Barçi* 'Please go!'.

Syntax

Possessive constructions are of the type usual in Turkic; except for personal names and toponyms, the possessor is normally marked with the genitive, and

the possessed with the third-person possessive suffix, e.g. *quyçiniñ iti* 'the shepherd's dog', but *Şāh Melik buzuqlıyında* 'at (the time of) Shah Melik's defeat'.

As in Ottoman, the Persian *izāfat* construction is copied to attach attributes to nouns (see pp. 118, 200–1). In Chaghatay, the attribute can also be an etymologically Turkic word, e.g. *kisver-i körk* 'land of beauty'. The Persian *-i* for singling out referents, the so-called *yā-i vahdat*, is also occasionally copied and attached to Turkic stems, e.g. *her oqi kim alar sari atti* 'every single arrow he shot at them'.

The passive may be constructed with the accusative, implying that the speaker is the agent, e.g. *Ani üç qisim qilildi* 'It was divided into three parts (by me)'.

For verbal negation, the stem is extended with *-mA*. Constructions with *yoq* are used for categorical negation, e.g. *Qilyan ê(r)mes-mèn, Qilmayan-mèn* 'I have not done it', *Almayumdur, Alyum yoq* 'I will not take it', *Hêç köz körmiş i yoq* 'No eye has seen it'.

The enclitic particle for yes/no questions is *mU*.

Non-Finite clauses

Relative Clauses

For left-branching relative clauses, the participle in *-GAn* is most widely used, e.g. *kök qapuyın saqlayan ferişte* 'the angel guarding the gate of heaven', *Yaxşı aqlamayın sözni aytmayıl* 'Do not utter anything [you] do not understand well'. The participles *-miş* and *-dUK* are still used in Khwarzmiyan Turkic, but later increasingly replaced by *-GAn*. The extension *-duryan* has intraterminal meaning, e.g. *baraduryan Özbek bile* 'together with the Uzbeks arriving'. The participle in *-(V)r*, negative *-mAs*, has mostly modal meaning, e.g. *Uruş eyler kişibiz* 'We are people that would fight', *Türki bilmes kişi* 'a person who does not know Turkic'.

There are also right-branching relative clauses of the Iranian type with finite verb forms. These constructions, which are often non-restrictive, revolve around the interrogative pronoun *kim* 'who', used as a generalised subordination marker. It is later increasingly replaced by Persian *ki*, e.g. *aya ini ki toqqan êrür* 'older and younger brothers who have been born', *bu sözler ki dediniz* 'the(se) words that you spoke'.

Occasionally the head itself is marked with the syntactic function of its coreferential counterpart within the relative clause, e.g. by means of the accusative suffix *-ni* in the sentence *Yay ve oqni ki tapib keltürdigizler ol kişidin bolmadi* 'The bow and arrow that you found and brought was not from that person'. This syntactic function, except that of the first actant, is marked on a resumptive pronoun, e.g. *Mèn ol quşturmën kim ‘āsiqlar mēni havadın tutarlar* 'I am the bird that the lovers catch from the air', *Kümişkent kentide, kim mēzāri ‘Alī Ata anda durur* 'in Kemishkent, where the tomb of Ali Ata

is located'. This marking is, however, optional with second actants of transitive predicates (direct objects).

Particularly frequent are relatives introduced with a demonstrative pronoun, e.g. *Bu kim mèn eylegüm durur durur mastūr* 'What I am going to do has been recorded', *anıñ kim al yüzinde xāli bolıyay* 'He/she who will have a mole on his/her ruddy face' (with genitive marking on the pronoun), *Yetti uşanda ki Zulayxā erdi* '(S)he arrived where Zulaikha was'.

Relative clauses with generalised meaning are a variant of the former, with an interrogative or a quantifier pronoun instead of a demonstrative, e.g. *kimge kim bērdim kögül* 'whoever I give my heart' (with dative marking on the pronominal head), *her nē kim aytsam andın artuq* 'whatever I say in addition to that', *qaysiniñ allıya kim qoydum baş* 'before whoever I put down my head', *Her nē ma'lūmdur i'lām etgil* 'Announce all that is known'.

Constituent Clauses

Constituent clauses of the Turkic type are constructed with verbal nouns, mostly those in-(V)r, negative -mAs, -GAn, and -mAK. Of these, -(V)r is mostly used for potential states of affairs, and -GAn for states of affairs presented as real. If the first actant of the constituent clause is not coreferential with that of the main sentence, it is usually, but not always, indicated by possessive suffixes on the verb form and marked with the genitive. Examples: *Bistar-i rāhat tilerni qoy* 'Stop asking for a pillow to rest on', *Sensiz bu dunyada yürümek maña harām turur* 'For me to walk around in this world without you is out of the question', *mēniñ ta'riḫni yaxşı bilürümni eşitib* 'hearing that I am well versed in history', *düstlarniñ külgenin düşmenlerniñ yıylayanın körüb* 'witnessing the laughter of the friends and the weeping of the enemies'. Nominal predications are subordinated with *ē(r)ken*, e.g. *ve tamyalariniñ nečük erkenini . . . mundaq aytib tururlar kim* 'and they explain what their tamghas are like as follows'. Constituent clauses may also be based on finite forms, e.g. *Bar murādım kim sēni öltürgümēn* 'It is my full intention to kill you'. In the following example, the subject of such a constituent clause is topicalised: *Qabri ma'lūm emes kim qaydadur* 'It is not known where his/her grave is'.

Adverbial clauses

Temporal clauses and clauses of purpose and reason are constructed with the same participles, but without genitive marking on the subject of the subordinate clause, e.g. *Yār ayiz ačmasya dardim sorıyalı taptım sabab* 'I found the reason why my beloved did not open his/her mouth to ask what my sorrow was', *namāz ayturda* 'while praying', *May iç peymāna tolmasdın burunraq* 'Drink wine before the cup has been filled', *hālın körgen zamān* 'when (s)he saw their situation', *haqıqatin bilmek ücün* 'in order to know the truth about it'.

Functionally equivalent clauses are constructed with conjunctions such as *kim ~ ki*, *qačan (kim)* 'when', *čün (kim)* 'when, as', *avval ki* 'before' and *tā*

(*kim*) 'until', e.g. *avval ki qadam qoydı* 'before (s)he took a step'.

Among the converbs, the one in -A (after vowels -y) may express manner, e.g. *Yıylay kirdi* '(S)he entered crying', *Andın toya içti* '(S)he drank from it till satiation'. It sometimes has temporal meaning, e.g. *Āftab oltura yēttim* 'I arrived at sunset', and may also be used with a purposive meaning, e.g. *Taşkandda qışlay bardı* '(S)he went to spend the winter in Tashkent'. If doubled, it has iterative meaning, e.g. *Qayyuḡni čēke čēke qarıptur Bābur* 'Bābur has grown old out of constant sorrow for you'.

The converb in -(I)p may locate the event of the predication formed with it before the event expressed with the main predication. The two events can be more or less intrinsically coordinated, e.g. *İlgeri barıp turdı* '(S)he went ahead and stopped', *Altun yayni tapıp keltürdigiñiz, taqı anı buzup ülestigiñiz* 'You found the golden bow and brought it, then you broke it and took a share each'.

The form -mAy serves as a negation for both -A and -(I)p, e.g. *Bir sā'at sözlemey olturdı* '(S)he sat for a moment without speaking', *keltüre bilmey* 'not being able to bring'. It has also largely taken over functions of -mAyIn < *-mAdIn, e.g. *on kün ötmey* 'before ten days had passed'. The latter converb was common in Khwarezmian Turkic, but became confined to poetry in Chaghatay.

The post-terminal converb in -GAč mostly indicates an event preceding the event of the main clause, e.g. *Kirgeč üç yükündüm* 'Upon entering, I prostrated myself three times'. The form *ē(r)geč* can be used for nominal predications, e.g. *čün oqiniñ hayālī başında egeč* 'as the image of his arrow was in my head'.

The converb in -GAlı mostly indicates purpose or intention, e.g. *Aş qıyalı bavurēi tapılmadı* 'No cook was found for making a meal', *Yētti sēkkiz kişini ilgeri xabar alyalı yiberildi* 'Seven or eight men were sent ahead to gather news'. It may also have abtemporal meaning ('since'), e.g. *Mēn ölgele köp boldı* 'It has been a long time since I died'.

The limitative-terminative converb -GUNča indicates an event during which the event of the main clause takes place, e.g. *Ālam bolıunča bu söz bolıusü turur* 'As long as the world exists, this saying will be valid'. It often focuses on the crucial limit of a transformative action (p. 42), e.g. *tañ atqunča* 'till dawn'. An example with negation is *Ol işni qılmaıunča köñli tinmas erdi* '(S)he was not at ease until/unless (s)he finished that job'.

References

- Budagov, L. (1869–71) *Sravnitel'nyj slovar' turecko-tatarskix narečij*, 2 volumes, St Petersburg: Imperatorskaja Akademija Nauk.
 Eckmann, J. (1959) 'Das Tschagataische', in J. Deny et al. (eds) *Philologiae Turcicae Fundamenta* 1, Wiesbaden: Steiner, 138–60.
 — (1964) 'Die tschagataische Literatur', in L. Bazin et al. (eds) *Philologiae Turcicae Fundamenta* 2, Wiesbaden: Steiner, 304–402.

- (1966) *Chagatay Manual*, The Hague: Mouton.
 Pavet de Courteille, A. J. B. (1870) *Dictionnaire turc-oriental*, Paris. Reprinted
 Amsterdam: Philo Press, 1972.
 Ščerbak, A. M. (1962) *Grammatika starouzbekskogo jazyka*, Moscow, Leningrad:
 Nauka.
 Šeyx Süleyman Efendi (1298) *Luyat-ı Čayatay ve Türki-yi ‘Osmānī*, Istanbul, 1882.
 Zenker, J. T. (1866–76) *Türkisch-arabisch-persisches Handwörterbuch*, Leipzig.

10 Ottoman Turkish

Celia Kerlake

Introduction

As a linguistic term, ‘Ottoman’ denotes the form of Turkic which became the official and literary language of the Ottoman Empire. This was, essentially, the variety of West Oghuz Turkic (p. 82) which developed in Anatolia after that region was settled by Oghuz Turks in the eleventh to thirteenth centuries.

Historical Development

The earliest surviving firmly datable Turkish works to which an Anatolian provenance can be confidently attributed belong to the thirteenth century, and clearly predate the emergence of an Ottoman principality. The Seljuk dynasty which had ruled most of Turkish Anatolia since the late eleventh century did not either use Turkish for official purposes or encourage its use in literature. It was around the end of the thirteenth century, with the replacement of Seljuk power by a multiplicity of small principalities whose rulers had not been connected with the Persianised Seljuk court, that Turkish began to be used as the language of administration. There was then a concomitant upsurge of Turkish literary activity, much of it religio-didactic in character, and often consisting of translation or adaptation from Persian.

There is evidence that a written form of Oghuz Turkic, distinct from the established written Turkic of the Karakhanid period (p. 85), had already begun to develop in Central Asia from the twelfth century at the latest. However, the earliest surviving Anatolian works still show strong linguistic influences from the Karakhanid literary tradition. It is not until about 1300 – coincidentally the period of the foundation of the Ottoman principality – that these eastern Turkic residues completely disappear. From that time on the Turkish texts produced in Anatolia are written in a language which is fully Oghuz in character.

From the latter part of the fifteenth century onwards, however, when the Ottoman court and ruling class had acquired a consciousness of imperial power, certain stylistic registers emerged – particularly for official correspondence and elite literature – in which the Turkic base was all but submerged beneath a heavy overlay of Arabic and Persian elements. These reflected the dominance of Arabic and Persian in the Islamic high culture and

learning which the Ottoman elite sought to emulate and advance. The borrowings were not confined to the lexicon, but included also some morphological elements such as plurals- and gender-marked forms, and a number of highly prevalent syntactic structures: right-branching nominal phrases, prepositional phrases, adverbial clauses of the Indo-European type. With the exception of a small lexical component, these imported elements were absent from the speech of the mass of the Turcophone population. The texts which survive from the Middle Ottoman period thus display an immense stylistic diversity, ranging from simple, unsophisticated registers which must be assumed to approximate to the spoken idiom of the day, to highly elaborated styles which were comprehensible only to the effectively trilingual elite.

The application of the essentially dynastic and political term 'Ottoman' to the official language of the state was an innovation of the period of reform known as the *Tanzīmāt*, beginning in the mid-nineteenth century, when deliberate attempts were being made to foster a sense of Ottoman identity as the basis for a modernised Ottoman state. Up to that time the term used to distinguish this language from others had been *türkī* or *türkçe* 'Turkish'. Turkish had had no place in the traditional Islamic schools (*medrese*), which concentrated exclusively on Arabic. Similarly, no dictionaries of Ottoman Turkish had been produced except by foreigners for foreign use. The same was true of grammars, with the exception of two works from the sixteenth and seventeenth centuries which appear to have aroused little interest, and to have been quite forgotten by the nineteenth century. Now, however, the study of the Ottoman language became an important part of the curriculum of the new state education system, and for the first time efforts were made to define and regularise this language, with the aim of maximising its effectiveness as a means of public communication.

Particular emphasis was placed upon (a) the orthography of the Turkish elements in the lexicon, whose representation in the Arabic script had remained erratic, and (b) the delimitation of the Arabic and Persian component of the lexicon, that is to say the determination of which lexical items from those languages – including some plural forms and fixed phrases – and which specific meanings of individual items had acquired sufficient currency in Ottoman to warrant their inclusion in a dictionary of that language.

Problem of Terminology

During the closing decades of the Ottoman Empire the designation of the language as Turkish rather than Ottoman became increasingly common. With the establishment of the Republic of Turkey in 1923 the term 'Ottoman' was proscribed as a descriptor of anything pertaining to the new nation state, including its language. After the replacement of the Arabic by the Latin alphabet (1928) and the launching of the language reform movement in 1932 (see Chapter 13) the term *Osmanlıca* 'Ottoman (language)' increasingly

acquired, within Turkey itself, overwhelmingly negative, alien connotations. It is still the belief of many educated Turks today that Ottoman was an entirely artificial hybrid of Arabic and Persian, spoken and written by a privileged elite, and bearing virtually no connection to the 'Turkish' which was kept alive by the illiterate masses.

On scientific grounds, also, many scholars have questioned the appropriateness of Ottoman as a linguistic denomination. During the first two centuries of its political existence (1300–1500), the Ottoman state encompassed within its frontiers only a section – albeit a growing one – of the Turcophone population of Anatolia. In view of the fact that many of the texts surviving from this era, especially from the period before 1450, were produced in areas that were under the control of other Turkish dynasties, modern Turkish scholars tend to use the geographical designation 'old Anatolian Turkish' or the anachronistic 'old Turkey-Turkish' for this diachronic phase of the language. Even in the late fourteenth century, however, there were already communities of Turkish-speaking people who had migrated from Anatolia to the Balkans with the expansion of Ottoman rule into that region. And from the fifteenth or sixteenth centuries there were Turkish-speaking officials and soldiers in all the urban centres of the far-flung empire, from Sarajevo to Baghdad and from Belgrade to Cairo. A considerable proportion of the vast heritage of documents and books in Turkish surviving from the Ottoman period is of extra-Anatolian provenance. Given the need for a term to encompass both the early Anatolian phase and the later imperial phase of this single language, Ottoman still appears to be the most reasonable choice.

Periodisation

It is generally accepted that Ottoman can be divided diachronically into three phases, and that these can be dated, very roughly, as follows:

Old Ottoman: thirteenth to fifteenth centuries

Middle Ottoman: sixteenth to eighteenth centuries

New Ottoman: nineteenth century to 1928.

Like all attempts to introduce chronological demarcation lines into the continuum of language change, there is an element of artificiality in this periodisation. Whether one is looking at phonological, lexical, morphological or syntactic change, it does not occur in tidy diachronic packages, but tends to operate at different speeds, and even sometimes in different directions, with respect to specific items affected. Nevertheless, there is certainly some value in the above periodisation, for the following reasons.

The phonology and morphophonology of Old Ottoman show a number of clear differences from modern Turkish. The evidence of the 'transcription texts' (pp. 183, 184, 187) suggests that the transition to the modern state of phonology took place during the Middle Ottoman period.

As far as morphology is concerned, there are a number of suffixes which, while common in Old Ottoman, are rarely encountered after the fifteenth century, except in poetry. Morphologically speaking, New Ottoman is almost indistinguishable from modern Turkish.

Middle Ottoman coincides with the period when the highly elaborated, Arabic-Persian-dominated stylistic register known as *inşā* ('construction') enjoyed its full flowering. The overwhelming preponderance of Arabic and Persian lexical items in this register had the effect of 'downgrading' a considerable proportion of the native vocabulary of Old Ottoman, which gradually disappeared from the written language altogether.

Middle Ottoman is also the period when, under the influence of the *inşā* aesthetic, right-branching nominal phrases constructed on the Persian model acquired wide currency even in relatively unpretentious texts, and became overwhelmingly dominant in the elevated style. A large number of collocations of this type became lexicalised and passed into the speech of the urban classes. This process accelerated in the period of New Ottoman, when the coining of terms to designate entities belonging to the modern world – from political institutions to scientific terminology – was almost invariably done using Arabic vocabulary and Persian syntax.

The three periods are also differentiated from each other in terms of their sentence structure. In contrast to the developments in the areas of lexicon and sub-clausal syntax, the transition from Old to Middle Ottoman is marked by an increase in the use of native Turkish subordinative structures, at the expense of the Indo-European type. In Middle Ottoman there is also a marked increase in the complexity and length of sentences, often at the expense of clarity. One of the most conspicuous characteristics of New Ottoman is the pragmatic reform of sentence structure through the rationalisation of the use of converbial forms.

Presentational Basis of this Chapter

The present chapter will be devoted to identifying the specific features which differentiate Ottoman or some diachronic sub-section of Ottoman from modern Turkish. No attempt is made here to give an account of dialectal variation. However, a collation of the data given for Old and Middle Ottoman with that presented for modern Turkish dialects (see Chapter 12) will give a fuller picture of the continuities between the earliest, relatively heterogeneous profile of written Anatolian Turkish and the regional divergences which persisted at the popular level after the standardisation of the official and literary language.

Sound System

Orthography

Attempts to describe the phonological state of a language at any period before the invention of audio-recording technology can only be conjectural. In the

case of Ottoman there is the added difficulty that it was written in a script which primarily represented the phonology of a highly different language, Arabic. Because the use of this script continued into modern times – up to 1928 – we have a clear picture of the 'fit', or lack of it, between the Arabic-script orthography and the phonology of Ottoman in its final phase. By the end of the Ottoman period there were many discrepancies between the phonological situation implied by the orthography and actual pronunciation, e.g. the consistent use of the letter indicating *i* or *ī* in a suffix which in certain phonological contexts was pronounced with *ū* or *u* or vice versa, or the use of one of the letters corresponding to *t* for a consonant which was pronounced *d*. In such cases it is impossible to determine with any certainty at what historical period the phonology of the specific lexeme or suffix began to diverge from its graphic representation. Some conventionalised spellings may simply have been continued from the earlier, Central Asian orthographic tradition and may not correspond to the actual phonological facts of even fourteenth-century Anatolian Turkish. Another problem is that some of the phonological distinctions (e.g. *g:k*, *ö:o:ü:u*, *i:i*) were not visible at all in the orthography.

It should be noted that most Arabic and Persian borrowings, including all those confined to the sophisticated written registers, were in any case always represented in their original orthography regardless of how they were pronounced by Turkish speakers.

Fortunately, from the sixteenth century onwards, the interest shown by Europeans in the Turkish language, and the need of some of them to learn it for practical purposes, resulted in the publication of a number of 'transcription texts' in which passages of Turkish are presented in Latin characters, and also of various dictionaries and grammars in which, again, the authors attempted to show the pronunciation of Turkish words using a Latin alphabet. Although their reliability obviously varies according to the skill and accuracy of the individual authors, these works are collectively of great value in providing some of the kinds of phonological information which are obscured by the limitations of the Arabic script or by the conservatism of Ottoman orthographic practice.

Phonetics and Phonology

Old Ottoman appears to have possessed a number of phonemes which are absent from standard modern Turkish, and which gradually disappeared during the Middle Ottoman period.

A number of lexical items which today are pronounced with *e* in the first syllable were written in Ottoman in a way which may be interpreted as *é*, e.g. *dé*- 'say', *gêje* 'night', *gêrü* 'back', *ér*- 'reach', *ét*- 'do', *vêr*- 'give', *yê*- 'eat'; cf. modern Turkish *de-*, *gece*, *geri*, *er-*, *et-*, *ver-*, *ye-*. This 'closed *é*' (see pp. 89, 94–5) is found today in many Anatolian dialects.

The consonant *ɣ*, which occurred only in back environments, seems to

have largely disappeared in syllable-initial position, and in word-final position in polysyllables, in Oghuz Turkic before 1300. The orthography even of the earliest Anatolian texts reflects this, often showing a labialised vowel where there had earlier been a final γ . However, the fact that the grapheme which represented γ continued to be used in other positions right down to the end of the Ottoman period makes it very difficult to date the process of the total demise of this phoneme with any precision. This problem is carried over into the 'transcription texts', whose authors seem generally to have felt the need to transliterate this grapheme with g or gh .

The consonant χ , which survives today in Anatolian dialects, seems to have been in free variation with q in many lexemes, in word-final and medial position. In final position it seems to have disappeared rather earlier than in medial position. Examples: $\check{c}o\chi$ 'much', 'many', $yo\chi$ 'non-existent', $a\chi\check{s}am$ 'evening', $yo\chi\check{s}ul$ 'poor', $qo\chi u$ 'smell', $qor\chi u$ 'fear'.

The consonant η disappeared from the standard language, but was retained in many dialects.

Phonological and Morphophonological Changes

The main phonological and morphophonological changes which seem to have taken place during the Old and Middle phases of Ottoman are the following.

- 1 Disappearance of intervocalic and syllable-final G . Between vowels γ dropped out altogether, or – in the vicinity of rounded vowels – was replaced by \emptyset in free variation with v , e.g. $aya > aa$ 'lord', $aya\check{c} > aa\check{c}$ 'tree', $soyuq > so(v)uq$ 'cold'. Front g in this position was transformed into y in unrounded contexts, v between a rounded and an unrounded vowel, and \emptyset or v between two rounded sounds: $begen- > beyen-$ 'like', $gügerjin > güverjin$ 'pigeon', $dügün > düün$ 'wedding', $dögin- > dö(v)ün-$ 'beat one's breast'. In both pre-consonantal medial position and word-final position, γ was replaced by vowel length: $ayri > āri$ 'pain', $day > dā$ 'mountain'. In similar positions, g was replaced by y after e , and transformed into vowel length after i or $ü$. After $ö$, it was transformed into vowel length v or y , depending on the nature of the following consonant. Examples: $beg > bey$ 'lord', $igne > ine$ 'needle', $dügme > düme$ 'button', $ögren- > ören-$ 'learn', $dögmek > dövme$ 'beat', $ögle > öyle$ 'noon'.
- 2 Replacement of η by m or n , e.g. $to\eta uz > domuz$ 'pig', $de\eta iz > deniz$ 'sea', genitive suffix $-U\eta > -In$, second-person singular possessive suffix $-(U)\eta > -(I)n$.
- 3 Voicing of initial t has taken place in a large number of lexemes which preserved a spelling indicating t - but are known to have acquired a pronunciation with d - by the New Ottoman period. Examples: $tā > dā$ 'mountain', $tol- > dol-$ 'become full'. However, this change was not universal. Turkish has retained to the present day a considerable number

of lexemes with initial t , such as tuz 'salt', $taş$ 'stone'. The situation is further complicated by the fact that some lexemes which appear in Old Ottoman texts with d - subsequently underwent the reverse change of devoicing, e.g. $depe > tepe$ 'top', 'hill', $düken- > tüken-$ 'be exhausted'.

- 4 Development of rounded vs. unrounded harmony. At an early stage, any of the four front vowels e , i , $ö$, $ü$ could be followed by either of the two high front vowels i , $ü$, and any of the four back vowels a , i , o , u could be followed by either of the two high back vowels i , u . There was then a gradual change to a situation in which the vowels of the non-first syllables – whether part of the primary stem or a suffix – had to conform to the preceding vowel in terms of roundedness vs. unroundedness. Thus e and i were followed only by i , $ö$ and $ü$ only by $ü$, a and i only by i , and o and u only by u . In general, the only exception was that where the first syllable of a disyllabic stem had the vowel a and the second syllable began with a labial consonant the vowel of the second syllable, if high, tended to be the rounded u rather than the unrounded i , e.g. $\check{c}amur$ 'mud', $qabuq$ 'shell'.

In terms of bound morphemes, this process involved a change from a situation in which high-vowelled suffixes were divided into two classes, those that were invariably rounded-vowelled and those that were invariably unrounded-vowelled, to a situation in which all high-vowelled suffixes had the same four allomorphs – two rounded and two unrounded – and conformed to the vowel of the preceding syllable in terms of their labiality.

Note that the fourfold allomorphic alternation incorporating front–back and rounded–unrounded harmony became definitively established only in New Ottoman. It superseded the Old Ottoman bifurcation between suffixes with l and suffixes with U .

If in the following survey both unrounded and rounded forms of a morpheme are shown, this means that before the establishment of rounded vs. unrounded harmony the allomorphs appear to have been in free variation.

- 5 Voicedness vs. unvoicedness of stops is a particularly problematical area of Ottoman phonology and morphophonology.

In modern Turkish, stops in word-final position are unvoiced, and this is clearly reflected in the orthography adopted in 1928. In the Ottoman period, words of Arabic and Persian origin retained their original orthography, and where this indicated a final b , j , or d it is impossible to be certain how this was pronounced in a Turkish context. Where a particular lexeme had become part of the vernacular, as in the case of $kitāb$ 'book' ($> kitap$), it may even have had a learned pronunciation which retained the original voiced stop, and a 'popular' one in which the final consonant was unvoiced.

As far as native elements are concerned, the picture is very confused. The only commonly occurring bound morpheme ending in a stop, the converb suffix *-(y)Ubp*, was usually spelled with *b* in the Arabic script, and this is reflected in the standardised nineteenth-century orthography. In the case of lexical items, however, there was very little consistency of practice, and one can even find the same lexeme being spelled in different ways within one manuscript. The only generalisation that can be made is that the tendency to represent a final stop as voiced is greater in lenes (see pp. 32, 98), e.g. *qaB* 'container', *a(γ)aJ* 'tree', *uJ* 'end', 'frontier', *jurD* 'home country'. These are the consonants that become voiced in today's Turkish before a suffix beginning with a vowel. The use of the grapheme representing the voiced stop is relatively uncommon in fortes, e.g. *hep* 'all', *iç* 'inside', *it* 'dog'. These are the final consonants that are stable today, i.e. do not become voiced before a suffix beginning with a vowel.

Final *-q* and *-k* present a rather different picture. As noted above, their voiced counterparts were disappearing in intervocalic and syllable-final positions during the Old and Middle periods of Ottoman. The consonant *g* was not regularly differentiated from *k* in the orthographic system. Our understanding of the situation can therefore only be based upon how final velar stops are represented. If Old Ottoman *-q* seems to be in free variation with *-γ* in a few lexemes such as *aq* 'white', this is often due to sandhi before an initial vowel of the following word.

Voicing of final stops before a vowel-initial suffix seems to have taken place in rather more nominal stems, and considerably more verbal stems, than is the case in modern Turkish. The difference is primarily to be observed in stems with final *-t*, e.g. *buludu* 'cloud', *südü* 'milk' (accusative case, with labial harmony assumed).

In Old and Middle Ottoman the orthographic evidence of *-q* and *-γ* points to a tendency towards voicing of the final stop where the following word began with a vowel, as in *Merdāneliy eyler* 'He acts bravely', *olmay içün* 'in order to be', *çoy ise* 'if it is much'. Traces of this sandhi phenomenon are found as late as the nineteenth century, when the converb produced by the addition of the postposition *ile* 'with' to the infinitive in *-maq* was still regularly written *-mayıla*.

The verbal stems which underwent voicing of a final stop before a vowel-initial suffix included all causatives in *-t*, e.g. *araD-* 'cause to search', *qabarD-* 'cause to puff up', some other stems in *-D*, e.g. *işid-* 'hear', *yaraD-* 'create', and the stem *biraG-* 'leave'. In the latter case the demise of *γ* evidently led to the relatively early end of pre-vocalic devoicing. In the case of the *-D* stems it seems to have continued right into New Ottoman. There is a trace of it even today in the deverbal noun *yaradılış* 'creation', although all the productive forms of this verb are now derived from the fixed stem *yarat-*.

Another question concerns suffix-initial stops following an unvoiced consonant. Throughout the Ottoman period, all suffixes which in modern

Turkish begin with *D* and are regularly unvoiced following an unvoiced final consonant in the stem were written with an invariable <d> in all contexts. Examples are the locative suffix *-DA*, the predicate marker *-DUr* or *-Dir*, and the simple past marker *-DI*. It is impossible to know exactly what significance to attach to the insistence on the voiced grapheme as far as the Old Ottoman period is concerned, but it is quite clear from the 'transcription texts' that by the Middle Ottoman period this was merely an orthographic convention, and that the actual pronunciation of the initial stop in these suffixes was unvoiced after an unvoiced consonant, just as it is today.

The situation is somewhat more complicated for suffixes which begin with *J* (= modern Turkish *C*). The unvoiced alternate *ç* is sometimes explicitly recognised in the orthography, e.g. *çöpçe* 'like a straw' (sixteenth century). In Old Ottoman *ç* can even be found following a voiced consonant, as in *gözçi* 'spy', *bençileyin* 'like me'.

Morphology

This section aims to present the commonly occurring inflectional morphemes. Borrowed elements of Arabic and Persian morphology, which were essentially part and parcel of the 'learned' lexicon, are not discussed. Even within the elevated registers such morphemes were – with rare exceptions – not applied to native Turkish stems.

For reasons of space there is no discussion of derivational morphology, except in relation to denominal nominal-forming suffixes, because of the centrality of some of these to the grammatical system.

Any morpheme which is still productive in Turkish is recorded here without any comment on distribution or meaning, unless it seems to have come into existence only during the Ottoman period or/and its semantic range was significantly different from what it is today. Only the older phonological state of each morpheme is shown. Any morpheme which is cited here in a different form from that occurring in modern Turkish can be assumed to have reached its modern state by about the end of the eighteenth century. Information is also given about morphemes which are found during only part of the Ottoman period, and have not survived into modern Turkish.

Nominals

Denominal Nominal-forming Suffixes

The allomorphs *-JI* and *-çI* of the suffix *-JI* are in free variation in Old Ottoman, regardless of whether or not the stem-final consonant is unvoiced, e.g. *avji* or *avçi* 'hunter'. The principal function of *-rAK*, which was not productive after the fifteenth century, is to express the comparative degree of an adjective or adverb. Where no element of comparison is present, the effect is either to reduce or to augment the intensity of the description expressed, e.g. *ajıraq* 'rather bitter', *gökrek* 'bluish', *ufaraq* (from *ufaq*) 'very small'. The

suffix *-(I)mrAK* also appears in the form *-(U)mtIK* in Old and Middle Ottoman, e.g. *qaramtiq* 'blackish', 'nearly black'. The suffix *-sayi ~ -sayu*, not encountered after the Middle Ottoman period, has the same function, e.g. *qarasayi*, *qarasayu* 'blackish', 'nearly black'.

Other suffixes are *-IİK ~ IUK*, *-IU*, *-sUz*, *-jAK ~ -jUK*, *-jUGAz* (equivalent of modern *-CAGIz*) and *-(I)msI*, *-sl*.

Declension

The order of nominal inflectional suffixes (plural–possessive–case) is the same as in modern Turkish. So is the form of the plural suffix *-IAr*. Just as in modern Turkish, only nouns are declinable. Suffixes may be added to adjectives only when the latter are being used with substantival function.

Apart from occasional non-West-Oghuz elements in the oldest texts, nominal inflectional morphology is almost identical with that of modern Turkish. The main differences are those which form part of the general processes of phonological change.

Possessive Suffixes

The possessive suffixes are the following: 1p.sg. *-(U)m*, 2p.sg. *-(U)ŋ*, 3p.sg. *-(s)I(n)*, 1p.pl. *-(U)mUz*, 2p.pl. *-(U)ŋUz*, 3p.pl. *-IArI(n)*. Examples: *eliim* 'my hand', *başuŋ* 'your head', *gözi* 'his eye', *qapusi* 'its door', *atlarumuz* 'our horses'. The 'pronominal *n*' shown in parentheses at the end of the third-person suffixes occurs only where the possessive suffix is followed by a case suffix, including a zero accusative suffix, e.g. *Qılıjın(i) çekdi* 'He drew his sword', *evinde* 'in his/her house'.

Case Suffixes

The case suffixes are the following: nominative *-Ø*, genitive *-(n)Uŋ*, accusative *-(y)I*, dative *-(y)A*, locative *-DA*, ablative *-DAn*. Examples of noun paradigms: nominative *qiz* 'girl', *yol* 'road', *evi* 'his house', genitive *qizuŋ*, *yoluŋ*, *evinüŋ*, accusative *qizi*, *yoli*, *evin(i)*, dative *qıza*, *yola*, *evine*, locative *qızda*, *yolda*, *evinde*, ablative *qızdan*, *yoldan*, *evinden*. For the variability of the accusative marking after a third-person possessive suffix, exemplified in *evin(i)*, see below (p. 197).

The Suffix *-KI*

The points of difference from modern Turkish with respect to the suffix *-KI* are the following. In Old Ottoman it was suffixed mainly to the locative case and to expressions of time, rather than to the genitive case. In Old and Middle Ottoman it took the form *-KI* in back-vowelled contexts. In other words, not only was its vowel subject to front vs. back harmony but also its initial stop underwent voicing, e.g. *ardındayılar* 'those behind him' (fifteenth century), *yarınqı gün* ('tomorrow's day') 'tomorrow', *şu qarşuyi qara dā* 'that dark mountain opposite' (both sixteenth century).

Comparative Suffix

The comparative degree of an adjective or adverb was expressed in Old Ottoman by *-rAK*, e.g. *şekerden tatluraq* 'sweeter than sugar' (fourteenth century).

In the absence of an explicit object of comparison – expressed, as in this example, by the ablative case – it is virtually impossible to distinguish the comparative sense from the intensifying or reductive sense mentioned above (p. 187), e.g. *tézrek varayın* 'I will go more/very/rather quickly' (fifteenth century). In its comparative function *-rAK* was replaced, not by any other bound morpheme but by the adverb *daği*, which in time became the modern *daha*, e.g. *Biraz daği ilerü varalum* 'Let us go a little further' (fourteenth century).

As in modern Turkish, the comparative degree of an adjective or adverb could also be expressed simply by juxtaposition of the nominative with the ablative-marked object of comparison, e.g. *Benden yaşlu bir kimesne yoqdur* 'There's no one older than me' (fifteenth century).

There was no separate superlative morpheme in Old Ottoman, but in the context of a genitive–possessive construction *-rAK* effectively acquires a superlative sense, e.g. *çulqlaruŋ begenilmişreği* 'the (most) admired of qualities' (fourteenth century).

The modifying particle *en < eŋ*, which is used to express the superlative degree in modern Turkish, seems to have come into use gradually from the fifteenth century.

Pronouns

Personal Pronouns

In the case of personal pronouns, the main differences from modern Turkish are the following. The third-person singular pronoun is *ol* in the nominative, *an-* in all oblique cases. The third-person plural pronoun is *anlar*.

Demonstrative Pronouns

As for the demonstrative pronouns, the main differences from modern Turkish are the following. The forms *şol* and *ol* were for centuries more common than their synonyms *şu* and *o*. The oblique forms of *ol* were based on *an-*, as above.

Table 10.1 Personal pronouns

	1p.sg.	2p.sg.	3p.sg.	1p.pl.	2p.pl.	3p.pl.
nom.	ben	sen	ol	biz	siz	anlar
gen.	benüm	senüŋ	anuŋ	bizüm	sizüŋ	anlaruŋ
acc.	beni	seni	anı	bizi	sizi	anları
dat.	bağa	sağa	ağa	bize	size	anlara
loc.	bende	sende	anda	bizde	sizde	anlarda
abl.	benden	senden	andan	bizden	sizden	anlardan

In the earliest period, the plural forms of all three pronouns were usually without *n*, e.g. *bular*, *şular*, *olar*. The locative forms *bunda*, *anda* were often used in the wider spatial senses of 'here' and 'there', respectively. The ablative *andan* could have, in addition to the spatial meaning 'from there', also the temporal sense of 'after that', 'then'.

Reflexive Pronouns

In addition to *gendü* (> *kendi*), the morpheme *öz* also functioned as a reflexive, e.g. *Öz yolüñi yine özüñsin uran* 'It is you yourself who attack (= rob travellers on) your own road' (fifteenth century). In pronominal function only, there was a third reflexive morpheme *gendöz* ~ *gendüz* (< *kendü* + *öz*), e.g. *Buldum ol sanduqda ben gendözümü* 'I found myself in that chest' (fifteenth century). The third-person singular and plural of the pronoun *gendü* usually had no possessive suffix, and did not take the 'pronominal *n*'. However, *gendüsi(n)* is also encountered from quite an early date. The following example shows both forms in use side by side: *Şey'üñ gendüsine gendüden yaqınraq kimse yoqdur* 'There is no one closer to the thing itself than itself' (sixteenth century).

Other Pronouns

Standard forms in Old Ottoman include the indefinite pronouns *kimesne* 'someone, anyone', *nesne* 'something, anything', the interrogative pronouns *kim* 'who' and *ne* 'what', and the generalised personal pronoun *kişi* 'one', 'a person'.

Determiners

Determiners include the indefinite article *bir* and other commonly occurring quantifiers such as *her* 'every', *bir qaç* 'a few', *bir az* 'a little', *bir miqdār* 'some', *çoq* 'much', 'many', *nije* 'much', 'many', *qamu* 'all', *dükeli* 'all', *jümle* 'all'. The quantifiers *qamu* and *dükeli* are obsolete after Middle Ottoman. Interrogative determiners are *qaç* 'how much, how many' and *qanqı* (> *hangi*) 'which'.

Verbs

The combinability and sequence of verbal suffixes is essentially the same as in modern Turkish.

Voice

In the area of voice modifications, the differences from modern Turkish are very minor. The principal suffixes are the cooperative-reciprocal *-(I)ş*, the reflexive *-(I)n*, the passive *-Il*, *-(I)n*, and the causative *-Ar*, *-Ur*, *-DUr*, *-GUr*, *-(I)t*. Some verbs which in modern Turkish form their causatives in *-Dir* are found with *-Ur* in older Ottoman, e.g. *artur-* 'cause to increase', *gëyür-* 'cause to put on (a garment)', of which the modern forms are *arttır-* and *giydir-*. The morpheme *-GUr* is found only on a few stems, e.g. *ërgür-* 'cause to reach', *turyur-* 'cause to stop', 'make stand' (cf. *erdir-*, *durdur-*).

Negation and Possibility

The verbal negative suffix is *-mA*. Possibility is expressed by the auxiliary verb *bil-* ('know') following a converbial form in *-(y)A*, *-(y)I* or *-(y)U*, the latter two alternatives being found only down to the sixteenth century, e.g. *Döne bildi* '(S)he was able to return' (fifteenth century), *Bili bilevüz* 'We should be able to know' (fifteenth century).

The usual way of expressing impossibility is to suffix *-(y)A* or (down to the sixteenth century) *-(y)I* or *-(y)U* to the verb stem before the negative suffix, e.g. *Dil söyleyümez* 'The heart cannot speak' (fourteenth century), *Bulımadum* 'I could not find' (fifteenth century). In Old Ottoman another, much less common, way of expressing impossibility was to use the negative stem of *bil-*, e.g. *Geçe bilmen* 'I cannot go through' (fifteenth century).

Phraseological Constructions with Auxiliary Verbs

The auxiliaries *ët-* (> *et-*), *eyle-*, *qıl-* (all meaning 'do'), and *ol-* 'be', 'become' were used, as in modern Turkish, as a means of incorporating loan elements into the verbal system. In Ottoman the overwhelming preponderance of elements so incorporated were Arabic verbal nouns. The auxiliary with which they were most often combined in such phraseological constructions was *ët-*, e.g. *teslīm ët-* 'surrender', *irsāl ët-* 'send'. Combinations of Arabic participial forms with *ol-*, e.g. *nāil ol-* 'obtain', *zāyi ol-* 'be lost', were next in order of frequency. Compositions involving Persian elements, e.g. *şād ol-* 'rejoice', were always far less numerous. In the more elevated styles of Middle and New Ottoman prose, loan-based constructed verbs all but replaced Turkish lexical verbs, and in non-finite contexts it was even possible to dispense with the Turkish auxiliary (see p. 201). In the 'rationalised' Ottoman of the nineteenth century most of the Arabic derivatives were preserved, many acquiring additional meanings relating to aspects of the modern world, scientific, technological, political, etc.

Actional Modification with Auxiliary Verbs

The type of construction expressing actional modifications and involving a converb form of the main verb followed by an inflected verb functioning as auxiliary (see pp. 42, 113–14) was much more prevalent in Old Ottoman than it is today, and a greater range of auxiliaries were in productive use. They could all occur following a converb in *-(y)A*, *-(y)I* or (occasionally) *-(y)U*.

Those most commonly occurring are: *gel-* < 'come', expressing a habitual state of affairs which has obtained for a long time, e.g. *mā-teqaddemden fuqarāya verilü gelmiş māl* 'the money which, from of old, has (always) been given to the poor' (sixteenth century); *gör-* < 'see' meaning 'take care to', e.g. *Işbu uzaq yollar için yarayunuz qılı görüñ* 'Take care to make your preparations for these long journeys' (fifteenth century), *tur-* < 'stand', expressing a state, e.g. *Ne turursun, inegüñ ayu yimiş öli turur?* 'Why are you standing (there)? Your cow has eaten poison, and is dying' (fifteenth century); *vër-* (> *ver-*) < 'give' and *yaz-* < 'fail to', used as in modern Turkish.

Less common or less productive are *düş-* < 'fall', *qo-* < 'put', both roughly equivalent in meaning to *vēr-*, *git-* < 'go', *yat-* < 'lie', both with durative meaning, and *qal-* < 'be left', which is used as in modern Turkish.

Finite Verb Forms

Personal Markers

The personal markers of the imperative are: 1p.sg. *-(y)AyIn*, *-(y)AyIm*, 2p.sg. *-θ*, *-gil/-yil*, 3p.sg. *-sUn*, 1p.pl. *-(y)AlUm*, 2p.pl. *-(U)ḡUz*, 3p.pl. *-sUn(lAr)*. The personal suffixes occurring after the markers *-DI* ~ *-DU* and *-sA*, and the copula particles *idi* and *ise* were: 1p.sg. *-m*, 2p.sg. *-ḡ*, 3p.sg. *-θ*, 1p.pl. *-k/-q*, (in conditional forms earlier *-(v)Uz*), 2p.pl. *-ḡuz*, 3p.pl. *-θ*, *-lAr*.

The following personal suffixes are attached to thematic markers except that of the imperative, to the copula particle *imiš*, and to the head of a nominal predicate: 1p.sg. *-vAn*, *-(y)Am* > *-(y)Um*, *-(y)In*, 2p.sg. *-sIn*, 3p.sg. *θ*, *-Durur* > *-DUr*, 1p.pl. *-(v)Uz* > *-(y)Uz*, 2p.pl. *-slz*, 3p.pl. *θ*, *-DUrUr* > *-DUr(-lAr)*. The aorist and the optative do not take third-person suffixes.

The oldest first-person singular form *-vAn* was already being overtaken by *-(y)Am* in the earliest period, but survived in literary registers until the sixteenth century. The development *-(y)Am* > *-(y)Um* is assumed to have been influenced by the possessive suffix. The suffix *-(y)In*, which has specifically west Anatolian connections, occurs down to the seventeenth century. The change of the first-person plural form *-(v)Uz* > *-(y)Uz* began during the fifteenth century.

The third-person form *-DUrUr* had developed in Oghuz from the aorist of *tur-* 'stand'. The shortened form *-DUr* was already standard in Anatolia by the fourteenth century, with *-DUrUr* lingering on in Ottoman as a mainly poetic variant.

Copula Particles

As in modern Turkish, unaccentable copula particles can be added to nominal and to some verbal predicates to alter the tense, mood, or syntactic independence of the predication. In most cases, they occur between the thematic marker and the personal marker. Thus, *idi* signals past tense, and *imiš* indirectivity, whereas *ise* forms conditional ('if'), and *iken* forms temporal clauses ('while').

In older Ottoman the distinction between the free and bound forms of these morphemes cannot be clearly drawn. Particularly in nominal predicates there is clear graphic evidence of the persistence of a transitional form in which, despite suffixation, the initial vowel was retained, e.g. *varidi* 'there was'. Whether suffixation occurred or not, a preceding final stop would be voiced.

Thematic Suffixes of Tense, Aspect and Mood

A number of morphemes produce stems of tense, aspect, mood. This finite verbal system differs in certain ways from that of modern Turkish in terms of structure, morphology and semantics.

The present in *-r*, *-Ar*, *-Ur*, the so-called aorist, expresses intraterminal (imperfective) aspect, broad or generalised validity, and various types of modality. The negative of the aorist was originally conjugated with *z* replacing the *r* of the positive conjugation in all persons: 1p.sg. *bilmezven*, *bilmezem*, 2p.sg. *bilmezsin*, 3p.sg. *bilmez*, 1p.pl. *bilmezüz*, 2p.pl. *bilmezsiniz*, 3p.pl. *bilmezler*. The transition to the modern first-person forms *-mAm* (singular) and *-mAYlz* (plural) began in the sixteenth century.

The form *-(A)yUr(Ur) ~ -(I)yUr(Ur)* is derived from the combination of converb in *-(y)I* with the aorist of the auxiliary *yürü-* 'walk'. It first appeared in the fourteenth century, and acquired its shortened form *-(I/A)yür* by about the end of the fifteenth century. It eventually developed into a focal intraterminal (progressive) and into the modern form *-(I)yör*. Before the full establishment of this form as a finite item the semantic range of the aorist also covered cases of higher focality (see p. 43), e.g. *Ne turursun?* 'Why are you standing (there)?', *Nije gündür otlardum ot* 'For many days I had been eating grass' (fourteenth century).

The simple past tense morpheme was *-DI* in the third person but *-DU* in the first and second persons, e.g. *Aldi* '(S)he took', *Alduq* 'We took'.

The finite form in *-mlš* expressed post-terminal (perfect) aspect, or indirective past tense (see pp. 45, 214). It did not always have indirective meaning in Old and Middle Ottoman. This is especially true of the first and second persons, where the meaning was often simply perfect, e.g. *İlm oqımaya gelmişem* 'I have come to study' (fourteenth century), *Ani görmüşmişin?* 'Have you seen him?' (fourteenth century).

The finite form based upon the converb in *-(y)Ub* was a past tense of post-terminal (perfect) aspect. It had two distinct structures: a short one occurring mainly in the third-person form *-(y)UbdUr*, and a longer one in which the suffix *-DUrUr* behaves as if it were still an auxiliary verb, being fully inflectable. The perfect sense of both forms is seen in examples such as *Şikāyeti nedür? Niçün gelübdür?* 'What is his complaint? Why has he come?' (fourteenth century), *Ben pīr olupdururam, unuttum* 'I have grown old, I have forgotten' (sixteenth century). Both these structures disappeared from the written language after the sixteenth century. (However, *-ip dur-* forms with durative actional content are still used in informal registers of modern Turkish.)

The form in *-(y)IsAr* was an Old Ottoman future tense. It was already in decline in the fifteenth century, and occurs only sporadically in the sixteenth century, e.g. *Vaqtüm geldi öliserem bīgümān* 'My time has come, I am surely going to die' (fourteenth century).

The prospective form in *-(y)AsI*, which survives in modern Turkish as a means of expressing curses, in Old Ottoman mainly expressed unmodalised future time reference. It occurred more commonly as a participle than as a finite verb, but examples such as the following are encountered: *Yarın sinde hem eyle olasısin ki qarınjadan alu qalasısin* 'And tomorrow, in the grave,

you will be in such a state that you will be more helpless than an ant' (fourteenth century).

The prospective form in *-(y)AjAK* appeared first in the fourteenth century as a participle. By the fifteenth century it had developed a full verbal paradigm, but its use as a finite form was long avoided in the more elevated registers. Right down to the seventeenth century writers aspiring to elegance of style preferred the optative for simple predictive statements, using *-(y)AjAK* only in non-finite forms and in periphrastic constructions with *ol-*.

As for the imperative, the first-person singular in *-n* was, from the start, less widely used than that with *-m*, and had specifically west Anatolian connections. It died out after the sixteenth century. The second-person singular form was \emptyset , the optional suffix *-Gİl* being rarely used after the fifteenth century. Old Ottoman examples are 1p.sg. *Vereyin* 'Let me give!', 2p.sg. *Varyıl* 'Go!', 3p.sg. *Baqmasun* 'Let him not look!', 1p.pl. *Bilişelüm* 'Let us get to know each other!', 2p.pl. *Dutuğ* 'Hold it!', *Bekleğüz* 'Wait!', *Inanmaguz* 'Do not believe it!'.

The optative in *-(y)A* expressed various types of modality and future time reference. It was irregular in the first-person singular, in that instead of *-(y)Am* it had *-m*: *bilem* 'I should know', *olmayam* 'I should not be'. The past form was often contracted from *-Aydl* to *-Adl*, e.g. *Istemeyedüm* 'I would not have wanted', *Sanaduğ* 'You would have thought'. Only a few traces of the optative survive in modern Turkish. Its first- and third-person forms had gone out of regular use by the seventeenth century, and its second-person forms, which once functioned, *inter alia*, as a formal imperative, are now almost abandoned. The rather disparate functions of the optative – optative modality, future time reference, modalised future, etc. – have been dispersed among other available forms.

The necessitative in *-mAIU* first appeared in the fourteenth century, but did not acquire the status of an autonomous finite form until about the end of the fifteenth century. In the earlier period it occurred either as a participle or as part of a composite predicate in which *ol-* carried the inflections. In such usages it often had prospective aspect, rather than necessitative modality meaning, i.e. was equivalent to *-(y)AjAK*, e.g. *Eger Sultān Murād üzereñe gelmelü olursa biz anı deryādan bu yağa geçürmezüz* 'If Sultan Murad should be intending to come against you, we will not let him cross over the sea to this side' (fifteenth century).

The conditional in *-sA* formed counterfactual and open (future-referring) conditional clauses and expressed optative modality.

Non-finite Verb Forms

The morphemes forming verbal nouns are essentially the same as in modern Turkish: *-mAK*, *-mAKIK*, *-mA*, *-(y)İş*, *-DUK*, *-(y)AjAK*. There are, however, some structural or morphological differences. Down to the sixteenth century *-(y)AsI* also sometimes occurred as a verbal noun. This form displays the

peculiarity that, where the third-person possessive suffix and accusative marking are required, only the suffix *-n* is added: *Yā Ömer, bu özri Tagrı senden qabül eylesin bilmezdim* 'Umar, I did not think that God would accept this excuse from you' (fourteenth century). In the Old Ottoman period the morpheme *-mA* did not produce verbal nouns. It existed only as one of the suffixes for deriving lexical nominals from verbal roots. In Old Ottoman the verbal noun in *-mAK* could occur with a possessive suffix, e.g. *Senden sorarlar . . . qiyāmet gününden ki qaçandur anuğ gelmeği* 'They will ask you about the day of resurrection, when its coming is' (fifteenth century). It could also take the genitive case, as in *Aylamayın hiç assisi yoq* (lit. 'Crying has no benefit') 'There's no point in crying' (fourteenth century). The *-DUK* form of **i-* 'be' was in regular use in Old and Middle Ottoman, and still sometimes occurred in New Ottoman. It could be suffixed to the preceding word, and subjected to front vs. back harmonisation, e.g. *Sen İbrāhīm qandayduyın bilür misin?* 'Do you know where Ibrahim is?' (sixteenth century).

The morphemes forming participles, verbal adjectives which can be used attributively, are the same as in modern Turkish: *-(y)An*, *-(A/U)r*, *-mAz*, *-mİş*, *-DUK*, *-(y)AjAK*, *-(y)AsI*. As in modern Turkish, all these forms can occur without a nominal head.

As for converbs, the primary ones, involving no personal markers, are produced by a number of morphemes. The suffix in *-(y)Ub* had a discourse-advancing, but no semantic modifying function. The variant *-(y)UbAn* disappeared from the standard language after the sixteenth century. In Middle and New Ottoman writings the converb in *-mAKIA* was either discourse-advancing (propulsive; see p. 64), like *-(y)Ub*, or it expressed causal modification.

As for the converb in *-(y)A*, *-(y)I*, *-(y)U*, the forms in *I* and *U* are not standard after the sixteenth century. Down to that period all forms often occurred singly – as opposed to the doubled usage in modern Turkish – and could express either manner or purpose, e.g. *Dijle ırımayın yürüyü geçdi* 'He crossed the river Tigris on foot' (lit. 'walking'), *Size sığını geldüm* 'I have come to take refuge with you' (both fifteenth century). The suffix *-(y)ArAK* expressed manner, but could also be discourse-advancing. Down to the sixteenth century this also occurs in the forms *-(y)IrAK* and *-(y)UrAK*.

Other converb markers are *-mAKsIzn* 'without', *-mAdIn* > *-mAdAn*, *-mAzdAn* 'without'/'before', *-ken*, *iken* 'while', *-DIKÇA* 'whenever', etc., *-mAK içün*, *-mAGA* 'in order to', *-DUKdAn başqa* 'apart from', *-(y)AjAK yërde* 'instead of', *-mAKdAnsa* 'rather than', *-(y)AlI*, *-(y)AldAn* 'since'. (Concerning the pronunciation of *d* and *ç* see pp. 186–7). The latter form died out during the Middle Ottoman period.

The form *-DIKdA* 'when', which is not common in standard modern Turkish, was still in use in the early twentieth century. In older Ottoman the converb in *-(y)InjA* often meant 'until', but by the New Ottoman period the meaning of 'when' had all but taken over, leaving 'until' to be expressed by

the extended form *-(y)InjAyA qadar*. The markers *-(y)IjAK*, *-(y)UjAK*, *-(y)InjAK* were the semantic equivalents of modern *-(y)InjA* 'when' in the Old and Middle Ottoman periods. They were sometimes extended by the addition of *-Iz*, *-Az* or *-zIn*. The commonest form, *-(y)IjAK*, still enjoyed some currency in the nineteenth century. The marker *-mAGIn* 'because' is attested throughout the Ottoman period, but was obsolescent by the nineteenth century. The suffix *-JA* 'as if' was added to *-r*, *-mlš* and *-(y)AjAK* stems. In Old Ottoman the form *-JA* was also used in this function. Of the two forms *-(y)AjAK qadar* and *-(y)AjAKIAyIn* 'enough to', the second is confined to Old and Middle Ottoman.

Adverbs

While many nouns and adjectives can be used as adverbs without any morphological indication, there are certain suffixes which function as adverbial markers.

The suffix *-JA*, which has important adverbial functions in modern Turkish also, was much more freely used in Old Ottoman, where its distribution shows striking differences from today's usage. In the oldest texts the consonant is usually shown as unvoiced, even after vowels. In Old Ottoman the nominal to which *-JA* was attached could be a complex nominal phrase, e.g. *Müslümānlar ol kāfirleri iki gendüleri görürlerdi* 'The Muslims saw those infidels as twice as numerous as themselves' (fifteenth century), *Geçligi yedi bu yerjedür* 'Its area is seven times (that of) this place' (fifteenth/sixteenth century).

The suffix *-(y)In* was, in comparison with its status in modern Turkish, where this former case suffix survives only in a few frozen forms such as *kışın* (*qışın*) 'in winter', slightly more productive in Old Ottoman. Among the time adverbials that occur are those in which the suffix has been added to *vaqt* 'time' in converbial constructions, e.g. *dönejek vaqtin* 'when (s)he was about to return', *šāh qatına getürdükleri vaqtin* 'when they brought (me) into the king's presence'. Other types of adverbials could also be formed, e.g. of manner: *Imdi qoğ kim ajin ve susuzin anda qirilsunlar* 'Now leave (them), so that they may die there of hunger and thirst' (< 'hungrily and thirstily') (fourteenth century).

The suffix *-IAyIn*, which in modern Turkish is again found only in a handful of (mainly time) adverbials, such as *sabahleyin* 'in the morning', was quite productive in Old Ottoman. It could be attached to any kind of nominal phrase, with the meaning of 'like', e.g. *kömürleyin* 'like coal', *aqar sulayin* 'like running water'. If the noun to which it was suffixed had the third-person possessive suffix, the 'pronominal *n*' would be added first: *anasinlayin* 'like her mother'.

Where *-IAyIn* was suffixed to a person-marked *-DUK* verbal noun, it produced a converb with either temporal ('when') or equative ('as') meaning: *Firuna qoyasin, baliq bišdüginleyin çiqarasin* 'Put it in the oven, (and) when

the fish is cooked take it out' (fifteenth century), *kim menziletüm senüñ göğlüğ içinde umduyumlayin ola* 'that my place may be in your heart, as I hope'.

The suffix *-JIAyIn* also meant 'like', and was added mainly to pronouns: *benjileyin* 'like me', *šunlarjilayin* 'like these'.

The suffix *-rA* had a very limited range of productivity (in Old Ottoman only) as a directional suffix, almost exclusively for indicating the part of the body at which a blow is struck: *depere* 'on the head', *ayzira* 'on his mouth', *qarnira* 'in his abdomen'.

Of the interrogative adverbs only those found in Old Ottoman will be mentioned here: *ničün* 'why', *niče*, *nije*, *niše*, *nite* 'how', *qačan* 'when', *qanjaru* 'where' (allative), *qani* 'where' (locative), *qanda* 'where' (locative, allative).

Syntax

The ways in which the syntax of Ottoman differed from that of modern Turkish are of two basic types. On the one hand there are differences arising from changes in the syntactic structure of Turkish itself. Changes in syntactic rules have actually been very slight, but because sentence structure is also to some degree a matter of stylistic choice one can certainly identify some major changes in the way that written Turkish texts, at least, have been constructed. Of far greater impact, however, are differences arising from the use, at certain periods and in certain registers, of syntactic structures modelled on those of Persian, which were subsequently rejected.

Changes in Syntactic Structure

As regards case marking in the nominal phrase, the phenomena listed below are features of Old and Middle Ottoman.

Accusative case marking was quite often omitted where the definite direct object bore a possessive suffix. In the third person, where this practice was so common as to be almost the rule in some texts, the 'pronominal *n*' was still retained even where the *-I* was dropped, with the result that *-(s)In* and *-IArIn* functioned as accusative marking.

Similarly, it was quite common for the genitive case not to be used to mark the relationship between a nominal and a following quantifier suffixed with a third-person possessive suffix. Instead, both the nominal and the quantifier took on the case marker, in the following examples the accusative marker: *Ani qamusini arqandan birayayduğ* 'You would have left it all (lit. 'it, all of it') behind you' (fourteenth century), *Hisarları kimin müdārāyila tapdurdi ve kimin yaymayi lan aldi* 'He subjugated some of the castles by diplomacy, and some by plundering' (lit. 'the castles, some of them') (fifteenth century).

It was initially almost standard practice for the subject of a person-marked verbal noun in *-DUK*, *-(y)AjAK* or *-(y)AsI*, even where this subject had

specific or definite reference, not to be provided with genitive case marking.

In Old and Middle Ottoman the locative case was often used with verbs expressing 'putting', 'striking', etc., which in the modern language always take the dative. With the demonstrative pronoun *an-*, when used with reference to place, it was the locative form, not the dative, that was regularly used for 'motion towards', e.g. *Anda vardum* 'I went there'. The locative form *qanda* 'where' could similarly be used in an allative sense.

Contrary to the situation in modern Turkish, it was common, right down to the end of the Ottoman period, for the *-mAK* verbal noun to have a grammatical subject, e.g. *Üç aya degin qapudān paša veyā imdād gelmek mutasavver mi?* 'Is it conceivable that the Grand Admiral or reinforcements will come within three months?' (eighteenth century). In modern Turkish this structure has been replaced by the *-mA* verbal noun in a genitive–possessive construction.

The predicative use of the adjective *gerek* 'necessary' is one of the standard ways of expressing obligation in modern Turkish. The subject of such a sentence is a nominal clause, and if the obligation is predicated of a specified entity the noun clause has to be of the type involving the *-mA* verbal noun in a genitive–possessive construction. Ottoman usage differed from this in the following ways. The nominal clause functioning as the subject of the sentence could be constructed as in *Kişi haddini bilmek gerek* 'One should know one's place' (traditional saying). Where the obligation was predicated of the speaker or the addressee, first- or second-person predicate marking would be added to *gerek*, e.g. *Sen bile olmak gereksin* 'You must be there together (with us)' (fifteenth century), *Ben bilmek gerek idim* 'I ought to have known' (seventeenth century).

A quite different construction, widely used throughout the Ottoman period, but represented in modern Turkish only by the form *olsa gerek*, involved *gerek* as the main predicate of a conditional sentence. The sentence as a whole could express obligation, e.g. *Leşker devşirüb varsa gerek idi, yine qarār edüb oturdi* 'He should have levied troops and gone, but he stayed where he was and did nothing' (fifteenth century). This use is unknown in New Ottoman. The sentence could also express a statement about an event which is or was expected to occur, e.g. *Bilejük tekvuri Yar-Hisār tekvurinuñ qizün olsa gerekidi* 'The lord of Bilejik was going to marry the daughter of the lord of Yar-Hisar'.

The linking of clauses through the use of *-(y)Ub* converbs was a feature of Anatolian Turkish texts from the beginning. In the Old Ottoman period, both the converb clauses themselves and the sentence as a whole were usually readily processable, being simple in structure and of manageable length. Alongside the other major changes that came about with the transition to Middle Ottoman there was a marked increase in the complexity (in terms of clause structure) and overall length of sentences. It was not just the semantically vague *-(y)Ub* converb that was employed in clause concatena-

tion, but a range of other converbs with, for example, temporal or causal modifying force. The effect of a mixed succession of converbs of all types without sufficient anchorage by finite verbs was to make it impossible to follow precisely the logical, temporal and aspectual connections between clauses. Nevertheless, the penchant for this kind of sentence structure acquired such a hold on the ruling elite that even the Tanzimat reformers were not able to oust it entirely. However, from the mid-nineteenth century onwards teachers, journalists and those concerned with the practical business of drafting modern laws and regulations did succeed in developing a new type of sentence structure, essentially that of modern Turkish, in which clarity was given overriding priority and a balance was achieved between subordinate clauses and finite predicates.

Persian-type Clauses

Imitation of Persian syntactic structures was not confined to the more high-flown styles of Ottoman writing. This is particularly true of subordinate clauses, where the relative distribution of copies of Persian constructions and of Turkish-type constructions in any given work is essentially a function of the date of composition, not of the text type or intended readership. In the Old Ottoman period, evidently as a result of the use of Persian literary models, including the translation of existing works, much use was made of imitations of subordinate clauses of the Indo-European type, preceded by a copied conjunction and ending with a finite verb. This was a reflection of the fact that Anatolian Turkish had not yet developed a strong written tradition of its own. In time, Turkish writers developed confidence in their ability to use their own language as a medium of literature and instruction. Hence in Middle Ottoman there is a noticeable decrease in the use of copies of Persian-type constructions, and a corresponding increase in the use of Turkish nominal, converbial and participial clauses.

The examples chosen for illustration here are taken from fifteenth-century texts. The clause types they represent are by no means exhaustive.

Clauses based on finite forms occur primarily with verbs such as 'know', 'say', 'order', 'wish'. They are preceded by the conjunction *ki*, copied from Persian, or the Turkish *kim*, which is the interrogative 'who', e.g. *Sen bilürsin ki yoldaşlarıñ sağa yār olmayajaqdur* 'You know that your comrades are not going to help you', *Destür vèrdi kim Qaraja-Hisāri varub yayma ve tālān ède* 'He gave permission for him to go and raid and plunder Karaja-Hisar'. The use of the optative in the last example is typical where the clause expresses a directive or commissive speech act, a command, request, oath, etc. A similar construction involving *Buyurdum ki* 'I (hereby) command that' was the standard formula in Ottoman imperial edicts of the classical period.

Temporal clauses were typically introduced by *çünki* or *qačan ki(m)*, e.g. *çünki sabah oldi* 'when it was morning', *qačan kim pādīšāh nāmeyi diğledi* 'when the sultan had read the letter'.

Copies of Persian conditional clauses were relatively uncommon. They were introduced by the Persian conjunction *eger* 'if' or its shorter form *ger*, e.g. *Ger dilerisiz bulasiz oddan nejat* 'If you wish, you may find salvation from the fire (of hell)', *Eger hisar alindi, hos* 'If the castle has been taken, (that's) good'.

Relative clauses were signalled by the relative pronoun *ki(m)*, e.g. *Bu turkler ki gelub bunda tavattun eddiler, etrafa el uzatmaya basladilar* 'These Turks who have come and settled here have begun to raid the surrounding area'.

In purpose clauses the introductory conjunction was either *ki(m)* or *ta ki(m)*, and the verb was always either optative or first- or third-person imperative, e.g. *At saldi kim ol dayi basin qurtara* 'He rode off, so that he too might save his life'.

Persian-type Nominal Phrases

Those writers who sought to use their literary skills not for the edification of the populace at large but in order to flatter potential patrons and to join in the intellectual and aesthetic activities of an increasingly refined urban elite turned once more to Persia, this time in search of models of stylistic elegance. In order to emulate the highly sophisticated rhymed prose of their Persian masters, they copied its syntactic structures, and also the Arabo-Persian lexicon on which it was based.

The Persian nominal phrase has the basic structure head + modifier. Whether the modifier is a noun or an adjective, the relationship between it and the head is signalled by the suffixation of a so-called *izafat* (izafet) marker to the head noun. In Turkish this marker was pronounced *-(y)I*. Arabic loanwords in Persian were used within these structures with their own morphology and micro-syntax.

The pluralisation of Arabic nouns involved a change in the shape of the whole word, e.g. *memleket* 'country' > *memalik* 'countries', *kitab* 'book' > *kutub* 'books'.

There were rules for gender and/or number agreement between noun and adjective, one of which was that plural nouns denoting inanimates required the feminine singular form of an adjective. Thus the name commonly given by the Ottomans to their state in the pre-modern period, 'the Ottoman dominions' was *memalik-i osmaniye*, the final *ye* being the Arabic feminine ending.

In rhetorical prose a favoured type of adjectival epithet was itself a compound, produced by the juxtaposition (without linking vowel) of two items, typically either noun + noun, adjective + noun or noun + verbal adjective. Examples of these types of epithets, each in construction with a head noun, from a sixteenth-century text: *padišah-i alem-penah* [monarch:IZAFAT world-refuge] 'the monarch (who is) the refuge of the world', *a'da-yi bed-nihad* [enemies:IZAFAT bad-character] 'evil-natured enemies', *südde-i sa'adet-bahşuğuza* [threshold:IZAFAT happiness-giving:2PL.POSS.DAT.] 'to your happiness-giving threshold [standard metonym for 'court']'.

Basic sequences of head + modifier could be extended by the addition of further modifiers, each linked to the previous one with a *-(y)I*, or, in the case of a coordinate noun, by the conjunction *U*.

Nominal phrases constructed on the Persian model took Turkish suffixes, as required, attached to the right-most item, which was more often than not an adjective. They could also be incorporated within a Turkish genitive-possessive construction, as shown in the following nineteenth-century example: *devlet-i aliyemizin bidayet-i zuhurundan beri* 'since the beginning of the emergence of our exalted state'.

The Syntactic Role of Arabic Verbal Nouns

Apart from the importation into the learned register of entire prepositional phrases, used adverbially just as they stood, e.g. *bilâ istisnâ* 'without exception', the influence of Arabic syntax on Ottoman was slight compared with that of Persian. However, mention should be made of the way in which Arabic verbal nouns, which had become such an integral part of the Ottoman verbal system, could be used to form converbial clauses even without the support of a Turkish auxiliary. In such constructions, which became particularly prevalent in the elevated styles of the nineteenth century, the Arabic verbal noun behaved like a verb, taking complements and modifiers just as if the auxiliary had been present, e.g. *din ü devlet ve mülk ü milleti ihyâ için* 'in order to revive religion and the state, the empire and the nation'. Here the Arabic verbal noun *ihyâ* 'revival' forms, with the Turkish postposition *için* 'for', a converbial expression of purpose (cf. *-mAK için*), and takes an accusative-marked direct object.

Turkification of Ottoman Syntax

The final phase of transition from New Ottoman to modern Turkish is linked with the rise of national consciousness among Turkish intellectuals as the empire entered a stage of rapid disintegration from the 1870s. A new emphasis on the Turkishness of the language was encouraged by growing awareness of its genetic relationship with Central Asian Turkic, all but forgotten for centuries. The first really organised campaign for the Turkification of Ottoman was launched by a group of young writers in 1911. Their principal target of attack was not lexical borrowings but the application, in Ottoman, of Arabic and Persian grammatical rules, and particularly those which, like the *izafat*, joined together sequences of non-Turkish words in structures alien to Turkish syntax. They were not bothered about the *ki*-type clause constructions in Turkish, because these and other surviving imitative usages of Persian conjunctions had become part of the spoken idiom, as they still are today, with their own particular pragmatic functions.

In their campaign to extirpate Persian nominal-phrase syntax and Arabic nominal morphology from the literary language, these Young Turk reformers were remarkably successful. Apart from such things as names of contemporary Ottoman institutions, which were invariably Persian-type constructs, the

literary and journalistic writings of the closing years of the Ottoman era are stylistically indistinguishable from those of the first decade of the Republic. In grammatical terms, at least, they contain virtually nothing that a Turk of today would find strange.

Further Reading

- Adamović, M. (1985) *Konjugationsgeschichte der türkischen Sprache*, Leiden: Brill.
- Bulut, C. (1996) *Evliya Çelebis Reise von Bitlis nach Van. Ein Auszug aus dem Seyahatname. Interpretierende Transliteration, kommentierte Übersetzung und sprachwissenschaftliche Bemerkungen*, Dr. Phil. thesis, University of Mainz.
- Buri-Gütermann, J. (1972) *Der Satzbau in der Sprache der osmanischen Urkunden aus der Zeit von Mehmed Fatih bis Süleyman-i Qanuni*, Vienna: Verlag Notring.
- Deny, J. (1921) *Grammaire de la langue turque (dialecte osmanli)*, Paris: Éditions Ernest Leroux.
- (1941) *Türk dili grameri (Osmanlı lehçesi)*, trans. Ali Ulvi Elöve, Istanbul: Maarif Vekilliği. (Includes emendations and additions by the translator.)
- Hazai, G. (1973) *Das Osmanisch-Türkische im XVII. Jahrhundert: Untersuchungen an den Transkriptionstexten von Jakab Nagy de Harsány*, Budapest: Akadémiai Kiadó.
- (1982) 'Present-day views on Ottoman historical grammar and some of its prevailing problems', *Archivum Ottomanicum* 7, 269–83.
- Johanson, L. (1990) 'Historische Grammatik', in G. Hazai (ed.) *Handbuch der türkischen Sprachwissenschaft*, I, Budapest: Akadémiai Kiadó, 74–103. Includes a very full bibliography.
- (1991) *Linguistische Beiträge zur Gesamtturkologie*, Bibliotheca Orientalis Hungarica 37, Budapest: Akadémiai Kiadó.
- (1992) 'Periodische Kettensätze im Türkischen', *Wiener Zeitschrift für die Kunde des Morgenlandes* 82: 201–11.
- (1994) 'Ein türkisches Erzählsatzmuster', in D. Bellmann (ed.) *Gedenkschrift Wolfgang Reuschel*, Stuttgart: Steiner, 165–74.
- Mansuroğlu, M. (1959) 'Das Altosmanische', in J. Deny et al. (eds) *Philologiae Turcicae Fundamenta* I, Wiesbaden: Steiner, 161–82.
- (1968) *New Redhouse Turkish-English Dictionary*, Istanbul: Redhouse Press. (Includes Ottoman as well as modern Turkish.)
- Prokosch, E. (1980) *Studien zur Grammatik des Osmanisch-Türkischen unter besonderer Berücksichtigung des Vulgärosmanisch-Türkischen*, Freiburg: Schwarz.
- Römer, C. (1981) 'Der Einfluss der Übersetzungen aus dem Persischen auf die Entwicklung des Osmanischen im 14. und 15. Jahrhundert', *Wiener Zeitschrift für die Kunde des Morgenlandes* 73: 89–114.
- Timurtaş, F. K. (1977) *Eski Türkiye Türkçesi: XV. yüzyıl*, Istanbul. (Old Ottoman phonology and grammar, based on fifteenth-century poetical texts.)
- (1963–74) *XIII. Yüzyıldan beri Türkiye Türkçesiyle yazılmış kitaplardan toplanan tanıklarıyla tarama sözlüğü*, 7 volumes, Ankara: Türk Dil Kurumu. (A dictionary of obsolete native Turkish words and suffixes, and of obsolete meanings of surviving native Turkish words and suffixes, with examples gathered from texts ranging from the thirteenth to nineteenth centuries.)
- Waetzoldt, I. (1978) *Zu den osmanischen Verbformen des 16. Jahrhunderts nach dem Mecmû'-ı Menâzil des Matrâqçı Nasûh*, Freiburg: Schwarz.

11 Turkish

Éva Á. Csató and Lars Johanson

Introduction

Turkish is the most widely spoken member of the Oghuz branch of Turkic (Chapter 1). Modern Standard Turkish is the descendant of Ottoman Turkish (Chapter 10), developed on the basis of the Istanbul dialect.

Turkish is the best documented Turkic language. Several good grammars and textbooks are available in, for instance, English, French, German, Russian, Norwegian, Turkish and Croatian. The present chapter will, therefore, focus on some chosen topics and refrain from detailed accounts of many issues that can easily be looked up in reference books.

Sound System

Vowels and Consonants

The Turkish vowel system is characterised by the oppositions front vs. back, high vs. low and unrounded vs. rounded. The eight vowels written as *ü, i, ö, e, u, ı, o, a* (see p. 207) utilise all the possible combinations of these features for distinctive purposes. The vowel *e* is realised as [e] or as the more closed variant [e̞], e.g. in *köpeği* [køpe̞i]. The so-called 'closed e', *é*, a sound between *e* and *i* which had a distinctive function at earlier historical stages, has lost it in modern Standard Turkish.

There is a strong prevalence of low rounded vowels, *o* and *ö*, not occurring in non-first syllables. Exceptions include the verbal suffix *-(Ø)Iyor*, diminutive forms of personal names, e.g. *Fatoş* ← *Fatma*, and copied lexical items, e.g. *lodos* 'south wind' ← Greek *nótos*.

Turkish also has eight corresponding long vowels. In native words they

Table 11.1 Short vowels (in Turcological notation)

	Front		Back	
	unrounded	rounded	unrounded	rounded
high	i	ü	ı	u
low	e	ö	a	o

result from the loss of a voiced velar fricative, which is still indicated in the orthography by the letter *ğ*, so-called 'yumuşak *ğ*', e.g. *dağ* [dɑ:ɣ] 'mountain', *düğme* [dy:me] 'button'. Moreover, many lexical items copied from Arabic and Persian contain long vowels, e.g. *hal* [ha:l] 'state'.

Since there are no diphthongs in Standard Turkish, original diphthongs in copied lexical elements are rendered by a sequence of a vowel + consonant + vowel, e.g. *tuvalet* 'lavatory' ← French *toilette*.

High unrounded vowels in non-accented syllables may be reduced or totally dropped, e.g. *biraktı* [b⁽ⁱ⁾raktı], *memnuniyet* [memnu:n⁽ⁱ⁾ ʔet]. This is also often observed in the pronunciation of Turks speaking foreign languages, e.g. German *türkische* 'Turkish' [tyrk⁽ⁱ⁾ʃɛ]. Some low vowels may also be reduced, e.g. *gazete* 'newspaper' [gɑzɛ].

Standard Turkish has five pairs of voiced and voiceless stops and affricates, three pairs of fricatives, two nasals, three liquids and two approximants. Consonant length has distinctive function; see the minimal pairs *eli* 'his/her hand' vs. *elli* 'fifty'; *atu* 'his/her horse' vs. *attı* 'threw'. Initial stops are aspirated, though not as strongly as in English. The voiced labiodental *v* is pronounced with less friction than the corresponding English sound. It can be pronounced as [w̥] when preceding a rounded vowel, e.g. *kavun* [kawun] 'melon'. It also tends to be deleted after rounded vowels in syllable-final position, e.g. [kɑ:] for *kov-* 'drive away, persecute'.

The alveolar *n* is pronounced as velar [ŋ] before *g* or *k*, e.g. *hangi* [hangɯ] 'which'. The old phoneme *ŋ* (see pp. 106, 184) has basically merged with *n*.

The consonant *j* [ʒ] has a limited distribution, occurring only in words of non-Turkic, mostly Persian and French, origin, e.g. *jest* 'gesture' ← French *geste*. Native words usually do not begin with *c* [dʒ], *j* [ʒ], *f*, *l*, *m*, *n*, *r*, *z*. Though *p* occurs in some genuine Turkish words, e.g. *piş-* 'cook', its frequency is low.

Front and back members of three consonant pairs alternate regularly in front and back syllables, namely (in Turcological notation) *k* vs. *q*, *l* vs. *ʔ*, *g* vs. *ğ* (see Table 11.2) or (in phonetic notation) [k] vs. [k̟], [l] vs. [l̟], [g] vs. [g̟].

In colloquial speech mainly, *h* is omitted at the end of closed syllables, and the vowel preceding it is lengthened, e.g. *Mehmet* [mɛ:mɛt]. It is also often

Table 11.2 Consonants (in Turcological notation)

	Labial	Alveodental	Prepalatal	Postpalatal	Velar	Glottal
stop	p, b	t, d		k, g	q, ğ	
fricative	f, v	s, z	ʃ, ž			h
nasal	m	n				
affricate			ç, j			
glide			y			
liquid			l, ʔ r			

dropped between vowels, e.g. *hastahane* [hastɑ:ne] 'hospital'.

The letter *ğ* represents various phonetic phenomena of controversial phonological nature. It may lack phonetic realisation or just signal the length of the preceding vowel, e.g. *sağ* [sa:] 'healthy', *öğren-* [œ:ren] 'learn'. It also has a front variant pronounced [i], e.g. *öğle* 'noon' [œ:lɛ] or [œ:lɛ]. A form such as *değil* 'not' may thus be pronounced [de:il], [de:il] or [di:l]. A weak semivowel [w̥] may occur in intervocalic position after *o*, e.g. *soğuk* [sɔwuk] 'cold'.

Syllable Structure

Consonant clusters in initial position are, with some exceptions, not allowed. Thus, in copied lexical items such clusters are split by means of prothetic or epenthetic vowels, e.g. *istasyon* 'station', *kulüp* 'club'. Consonant clusters are also split in final position, e.g. *isim* 'name' ← Arabic *ism*.

Syllable structure is further characterised by intrasyllabic front vs. back harmony, according to which syllables are lexically marked as either front or back. This quality is signalled by both the vowel and the consonant(s) of the syllable. The principles governing this phonological phenomenon can thus be best described at a suprasegmental level. It is not the frontness vs. backness of the individual segments, but the front vs. back categorisation of syllables that has distinctive function.

Certain consonant pairs such as *k* vs. *q*, *g* vs. *ğ*, *l* vs. *ʔ* have a greater phonetic capability of signalling front vs. back quality than others. Thus, the monosyllabic stems *kül* 'ash' and *kol* 'arm' are phonetically realised as [kyɫ] and [kɔʔ].

In copied lexical items, consonants may be dominant in signalling the quality of a syllable, e.g. the front [l] in the word *hali* [ha:li] 'its state' ← *hal* 'state' ← Arabic. Syllabic harmony is disguised by the Latin orthography, which does not mark front vs. back quality in consonants.

Morphophonological Variation in Primary Stems

The stops *b*, *d*, *g* and the affricate *c* are devoiced in syllable-final position. This explains the alternation of voiced and devoiced stem consonants in many lexical items. Thus, *kitap* 'book' has *p* in word-final position and before suffixes beginning with a consonant, e.g. *kitaplar* 'books', *kitapta* 'in the book', but *b* occurs if the consonant is syllable-initial, i.e. before vowels, e.g. *kitabım* 'my book'. Compare *ağaç* 'tree' vs. *ağaca* 'to a/the tree'. In many stems, a final *k* alternates in the same way with an element written *ğ*, e.g. *etek* 'skirt', *etekler* 'skirts', *eteğim* [eteim] 'my skirt' (see above).

Another kind of alternation occurs in stems that are bisyllabic in syllable-final position, but otherwise lose their second vowel, e.g. *oğuldan* 'from the son', *oğlu* 'his/her son'. Words with alternating stems include the type *isim* 'name' < *ism*, whose consonant cluster is split in syllable-final position, e.g. *isimden* 'from the name', but not when the second consonant is part of a new syllable, e.g. *ismi* 'his/her/its name'.

Morphophonological Variation in Suffixes

Intersyllabic Front vs. Back Harmony

The principle of intersyllabic front vs. back harmony means that syllables within a phonological word form are either all front or back ('palatal harmony'). Although it is a very strong tendency, it is not consistently observed in all polysyllabic lexical items. Exceptions include native words such as *anne* 'mother', *kardeş* 'brother', *elma* 'apple' and numerous words copied from other languages such as *günah* 'sin', *oksijen* 'oxygen'.

Front vs. back harmony governs alternation between different forms of suffixes. Most suffixes have alternative forms depending on whether the stem they are attached to is front or back, e.g. the plurals of *kedî* 'cat' and *balık* 'fish': *kediler* 'cats', *balıklar* 'fishes'. Suffixes containing a high vowel have four possible realisations, because a choice between a rounded and an unrounded suffix-vowel ('labial harmony') comes in addition to the front vs. back alternation. Thus, the genitive forms of *kedî* 'cat', *balık* 'fish', *kuş* 'bird', *öküz* 'ox' are *kedinin*, *balığın*, *kuşun*, *öküzün*. The roundedness of the suffix vowel is determined by the corresponding feature of the last stem vowel. Thus, labial harmony is 'vowel harmony'.

If the preceding stem lacks intersyllabic front vs. back harmony, the quality of its last syllable determines the choice of suffix variant. Thus, *anneler* 'mothers' contains a front variant of the plural suffix, and *elmalar* 'apples' a back one. Though the primary stem *dal-* is back and *sil-* is front, the forms *dalıyorum* 'I sink' and *siliyorum* 'I wipe' contain a back and rounded variant of the first-person singular personal suffix, *yor* being a back syllable containing a rounded vowel.

There are a few invariant, non-harmonic suffixes such as *-ki* and *-ken*, e.g. *onunki* 'his/hers', *uyurken* 'while sleeping'. In the intraterminal suffix *-(Ø)Iyor*, the vowel *o* is non-harmonic, e.g. *dalıyor* 'sinks', *ölüyor* 'dies'.

Most cases of assimilation are progressive. Cases of regressive assimilation are observed when certain suffixes with initial *-(y)A*, e.g. the prospective suffix *-(y)AcaK* and the participle suffix *-(y)An*, follow a stem ending in a low vowel, e.g. *bekle-* 'wait', *alma-* 'not take'. In these cases, *e* is raised to *i*, and *a* to *ı*, e.g. *bekliyecek* 'will wait', *almıyan* 'not taking' (written *bekleyecek*, *almayan* according to current orthography).

Consonant Assimilations

The tendency to maintain a (C)VC syllable structure explains some suffix alternations. Thus, the dative suffix is *-yA* after vowel stems, e.g. *ataya* 'to a/the father' ← *ata* 'father', but *-A* after consonants, e.g. *ata* 'to a/the horse' ← *at* 'horse'.

An interesting phenomenon is that words written with stem-final *ğ* behave like consonant stems, though they do not contain a consonant any more, e.g. *dağa* [da:ɑ] 'to a/the mountain' (not **daya*) ← *dağ* [da:ɣ]. A particular case

is represented by *-(Ø)Iyor*. When it is attached to a stem ending in a vowel, this vowel and not the suffix vowel is dropped, e.g. *bekl-iyor* 'waits' ← *bekle-* 'wait', *ar-iyor* 'seeks' ← *ara-* 'seek'.

Suffix-initial stops and affricates assimilate to the preceding stem-final consonants with regard to the feature voiced vs. voiceless. Suffixes beginning with a consonant have both voiced and voiceless variants, e.g. the simple past suffix *-DI* in *attı* 'threw' ← *at-* 'throw' and *aldı* 'took' ← *al-* 'take'.

Prosodic Phenomena

Pitch accent is normally on the last syllable of native lexical stems and expanded forms of them containing accentable suffixes, e.g. *elmá* 'apple', *elmalár* 'apples', *elmalardán* 'from apples'.

The distribution of dynamic stress, marked with ' in front of the syllable, is less predictable. It often falls on the first syllable, in particular if the latter is heavy (closed or with a long vowel), but can also move to other syllables.

In certain lexical stems of non-Turkic origin, the pitch accent coincides with the dynamic accent on a non-final syllable, the suffixes remaining unaccented e.g. *'Ánkara* 'Ankara', *'Ánkara'da* 'in Ankara'.

Certain suffixes such as the negative *-mE*, copula particles and the derivational suffix *-CE*, do not carry an accent themselves, but cause both kinds of accent to coincide on the preceding syllable, e.g. *'Yápmayacağım* 'I am not going to do it', *Ma'cárım* 'I am Hungarian', *Al'mánca* 'in German'.

Accent has a distinctive function in cases such as *yapmá* [do:INF.] vs. *'Yápma* [do:NEG.IMP.] 'Do not do (it)!', *güzelcé* 'rather nice', *gü'zélce* 'nicely'. A special vocative accent, used when calling the attention of a listener, deviates from the usual pattern, for instance, by a high pitch on the first syllable e.g. *'Káhveci* 'Coffee-seller!' ← *kahveci*. Compounds tend to give prominence to their first component, e.g. *Pa'zár günü* 'Sunday', *'başbakan* 'prime minister', *'árzet-* 'present'.

Orthography

The Latin alphabet, which replaced the previous Arabic script, was introduced in 1928. It marks frontness and backness of syllables only on vowels and has no means of indicating corresponding differences in consonants.

The alphabet contains 29 characters, including eight vowel signs, *a, e, ı, i, o, ö, u, ü*, and 21 consonants, *b, c, ç, d, f, g, ğ, h, j, k, l, m, n, p, r, s, ş, t, v, y, z* (for *ğ* see p. 205). The two vowel characters with diacritics, *ü* and *ö*, designate the high and low rounded front vowels [y] and [œ], respectively. The letter *ı* (without a dot) stands for the back, high, unrounded vowel [i̯], written *î* in Turcological notation. The upper-case form of the dotted *i* is also written with a dot, *İ*, e.g. *Istanbul*.

Pronunciation of some special characters: *c* [dʒ] as in *just*, *ç* [tʃ] as in *chat*, *g* as in *get*, *j* [ʒ] as in French *Jean*, *ş* [ʃ] as in *shut*, *z* as in *zoo*.

The apostrophe is used to separate proper nouns from suffixes, e.g. *Atatürk'ün* [Atatürk:GEN.] 'Atatürk's' and to write glottal stops in words

copied from Arabic, e.g. *Kur'an* 'the Koran'. A circumflex is sometimes used in words of Persian and Arabic origin to indicate that the syllable is front, e.g. *gâvut* [ɑ: wu] 'infidel', or that a vowel is long, e.g. *resmî* [resmi:] 'official'. The latter use is today restricted to distinguish word forms that would otherwise be homographs, e.g. *resmî* 'official' vs. *resmi* 'its picture'.

Morphology

Turkish morphology is characterised by its synthetic nature providing rich possibilities for expanding primary stems by means of a large number of relatively unchangeable and clearcut suffixes with rather clearly definable semantic functions. Word forms are easily segmentable and may contain derivational suffixes as well as a high number of different morphemes expressing grammatical notions. The average number of morphemes in a Turkish word is about four times higher than the corresponding figure for English.

The suffixes mostly have different phonologically predictable realisations depending on harmonic and other phonotactic properties of the stems. They are cited here in the standardised forms explained in pp. xx–xxiii. Derivational suffixes precede inflectional ones. An exception is the adjectival derivational suffix *-ki*, which follows declined nominals and can again be followed by declensional suffixes, e.g. *burada-ki-ler-den* [here:LOC.:DER.:PL.:ABL.] 'from the ones here'. Suffixes may develop different functions in different syntactic positions. Combinations of morphemes may gain functions not directly derivable from the functions of their components.

Nominal and verbal suffixes constitute strictly separate classes. With very few exceptions, one class is only attached to nominal stems, and the other only to verbal stems. A denominal and a deverbal suffix may have similar semantic functions, e.g. the denominal nominal *-CI* and the deverbal nominal *-(y)lci*, both indicating occupation or profession, e.g. *süt-çü* 'milkman' ← *süt* 'milk', *sür-ücü* 'driver' ← *sür-* 'drive'.

Nominals

The main nominal categories are nouns, adjectives and pronouns. Nouns and adjectives are not distinguished morphologically. Nouns can, for instance, be used as attributes, and adjectives can function as nouns. A particular lexical item is classified as adjective if it is dominantly used attributively, and is used with comparative and superlative markers. Adjectives may be used adverbially without any modification. The adverbial function may be underlined by repetition, e.g. *yavaş yavaş* 'slowly' ← *yavaş* 'slow'.

Nominal categories can be declined for number by means of a morphologically marked plural vs. an unmarked singular. There are six possessive suffixes marking the person and number of the possessor. There are six cases: five marked oblique vs. an unmarked nominative. The resulting declension paradigm contains 2 × 6 × 6, i.e. 72 forms. The order of suffixes is derivation

+ number + possessive + case, e.g. *öğrencilerimizden* [learn:DER.:PL.:POSS.1PL.:ABL.] 'from our pupils'.

Since there is no grammatical gender in Turkish, pronouns and pronominal suffixes are always ambiguous; thus the third-person pronoun *o* can be interpreted as 'he', 'she' or 'it'.

Derivation

One group of derivational suffixes used to build nouns from nouns has different diminutive meanings, e.g. *-CIK* and *-CAĞIZ* as in *Sinancık* 'little Sinan', *kızcağız* 'poor little girl'. Adjectives are derived with the very productive suffix *-li* 'having, provided with', e.g. *Norveçli* 'Norwegian' ← *Norveç* 'Norway', *şekerli* 'sweet' ← *şeker* 'sugar', *atlı* 'horseman' ← *at* 'horse'. The privative suffix *-siz* expresses the opposite, i.e. 'lacking, without', e.g. *parasız* 'without money'. Abstract nouns are derived with *-lik*, e.g. *boşluk* 'emptiness' ← *boş* 'empty', *parasızlık* 'impecuniousness, (the state of) not having money'.

A special case is represented by derivations based on the juxtaposition of repeated forms in which the first consonant of the second element is modified, e.g. *kitap mitap* 'books and similar things' ← *kitap* 'book'.

Declension

The only productive plural suffix in Turkish is *-lar*. Forms not marked for plurality are ambiguous between singular and generic readings unless the context excludes one of the two interpretations. Thus *Kitap okuyor* [book read:İYOR.PRES.] can be rendered in English both as '(S)he reads a book' or '(S)he reads books'. Singularity can be expressed by adding the indefinite article *bir* 'a(n)', e.g. *Bir kitap okuyor* [a book read:İYOR.PRES.] '(S)he reads a book' (see pp. 220–1).

The plural suffix indicates individual plurality and may be interpreted as referring to individual occurrences, e.g. *sular* [water:PL.] 'portions/masses of water'. Plural suffixes can also be attached to the inherently plural personal pronouns *biz* 'we' and *siz* 'you': *bizler* and *sizler*. The plural of a proper name means the person and his/her family, e.g. *Sinanlar* 'Sinan and his family'. The collective suffix *-gil* can also be used in some registers with the same meaning, e.g. *Ahmetgil* 'Ahmet and his family'. The plural suffix may also succeed the possessive suffix in expressions such as *annemler* 'my mother and the rest of my family'.

Possessive suffixes have the following forms: 1p.sg. *-(I)m*, 2p.sg. *-(I)n*, 3p.sg. *-(s)I(n)*, 1p.pl. *-(I)mIz*, 2p.pl. *-(I)nIz*, 3p.pl. *-lArI(n)*. Note that the last form, *taşları* in Table 11.3, is ambiguous between the readings 'his/her/its stones', 'their stone' and 'their stones'.

The five cases indicated by suffixes are genitive *-(n)In*, accusative *-(y)I*, dative *-(y)A*, locative *-DA*, ablative *-DAn*. Third-person possessive suffixes are followed by a 'pronominal *n*' when a case suffix is attached. As a result, some forms are ambiguous between second- and third-person readings, e.g.

Table 11.3 Possessive suffixes

Possessor	Possessed	
	Singular	Plural
1sg	taşım 'my stone'	taşlarım 'my stones'
2sg.	taşın	taşların
3sg.	taşı	taşları
1pl.	taşımız	taşlarımız
2pl.	taşınız	taşlarınız
3pl.	taşları	taşları

taşında 'on your stone' ~ 'on his/her/its stone'; see Table 11.4.

The cases mentioned so far can be regarded as core cases. Their suffixes are accentable, do not combine with each other, and require preceding third-person possessive suffixes to take on a 'pronominal *n*'. A few other suffixes that are often regarded as case suffixes display deviant properties, e.g. inability to carry pitch accent. These peripheral cases include the equative in *-CA* 'in the manner of, according to, by', e.g. *Ankara radyosunca* 'by/according to Radio Ankara', with 'pronominal *n*'. The suffixed form of the postposition *ile* 'with', *-(y)la*, does not require a 'pronominal *n*' in preceding third-person possessive suffixes, e.g. *eliyle* 'with his/her hand'. When added to certain pronouns, it governs a core case, namely the genitive, e.g. *benimle* [I:GEN.:with] 'with me'.

Old instrumental suffixes are preserved in certain adverbs of time, e.g. *'yazın* 'in summer', *'kışın* 'in winter'. An old directive suffix is found in adverbs such as *'sónra* 'later', which may be followed by a core case suffix, e.g. *'sónradan* 'later'.

Table 11.4 Case suffixes

	Singular		Plural	
	Singular	Plural	1p.sg. possessive	3p.sg. possessive
nom.	taş 'stone'	taşlar 'stones'	taşım 'my stone'	taşı 'his/her/its stone'
gen.	taşın	taşların	taşımın	taşının
acc.	taşı	taşları	taşımı	taşını
dat.	taşa	taşlara	taşuma	taşına
loc.	taşta	taşlarda	taşımnda	taşında
abl.	taştan	taşlardan	taşımndan	taşından

Adjectives

Comparatives and superlatives of adjectives are only formed analytically with the particles *daha* and *en*, e.g. *daha ayıp* 'more shameful', *en ayıp* 'most shameful'. The particle *daha* may be absent in comparative constructions in which the thing compared is in the ablative, e.g. *senden (daha) kuvvetli*

[YOU:ABL. (more) strong] 'stronger than you'. The meaning 'less than' is expressed by *az* 'little, few', e.g. *senden (daha) az kuvvetli* 'less strong than you'.

Intensive adjectives are formed by doubling the first syllable and adding – or replacing its final consonant with *-p, m, s* or *r*, e.g. *apaçık* 'wide open' ← *açık* 'open', *bomboş* 'quite empty' ← *boş* 'empty', *büsbütün* 'entirely' ← *bütün* 'whole', *çır(d)çıplak* 'stark naked' ← *çıplak* 'naked'.

Pronouns

Table 11.5 Paradigms of personal pronouns

	Nominative	Genitive	Accusative	Dative	Locative	Ablative
1p.sg.	ben	benim	beni	bana	bende	benden
2p.sg.	sen	senin	seni	sana	sende	senden
3p.sg.	o	onun	onu	ona	onda	ondan
1p.pl.	biz	bizim	bizi	bize	bizde	bizden
2p.pl.	siz	sizin	sizi	siz	sizde	sizden
3p.pl.	onlar	onların	onları	onlara	onlarda	onlardan

The polite second-person form is *siz* 'you', used to address one or more persons.

Reflexive pronouns are formed with *kendi* 'own': *kendim* 'myself', *kendin* 'yourself', *kendi* 'himself/herself/itself', *kendimiz* 'ourselves', *kendiniz* 'yourselves', *kendileri* 'themselves'. A polite way of speaking about a third person is to use *kendisi* instead of a personal pronoun, e.g. *Kendisi geldi* '(S)he has arrived'. This use signals that the speaker's deictic perspective is subordinated to the perspective of the person talked about.

Absolute forms of possessive pronouns are formed by adding *-ki* to the genitive of personal pronouns: *benimki* 'mine', *seninki* 'yours', *onunki* 'his/hers/its', *bizimki* 'ours', *sizinki* 'yours', *onlarınki* 'theirs'. The 'pronominal *n*' is added to *ki* when followed by case suffixes, e.g. *benimkinden* 'from mine'.

Reciprocal pronouns are formed with *birbir* + possessive suffix, e.g. *birbirimiz* 'each other of us'.

There are three demonstrative pronouns, of which *şu* 'this here, the following' is, as it were, a grammaticalised counterpart of the gesture of pointing at something. *Şu* refers forward in text. The other pronouns, *bu* 'this' and *o* 'that', basically represent different degrees of proximity. *O* is also used as a personal pronoun. The demonstrative pronouns have adjectival-adverbial forms such as *şöyle* 'like this here, like the following', *böyle* 'like this', *öyle* 'like that', and serve as bases for the derivation of place expressions such as *burası* 'this place', *oradan* 'from there'.

Numerals

The cardinals up to 'ten' are *bir, iki, üç, dört, beş, altı, yedi, sekiz, dokuz, on*. The tens from 'twenty' on are *yirmi, otuz, kırk, elli, altmış, yetmiş, seksen, doksan*. 'Hundred' is expressed by *yüz*, 'thousand' by *bin*, and 'million' by *milyon*. *Kırk* 'forty' means 'indefinitely many, umpteen' in some idiomatic expressions, e.g. *kırk yulda bir* 'once in a blue moon'. The other cardinals are formed by juxtaposition, e.g. *on bir* 'eleven', *dört yüz elli* 'four hundred and fifty'.

Ordinals are formed with *-(I)ncl*, e.g. *birinci* (also *ilk*) 'first', *onuncu* 'tenth' and *altmışıncı* 'sixtieth'. In expressions of fraction the denominator is in the locative, e.g. *yüzde bir* [hundred:LOC.one] 'one per cent'. Distributives are formed with *-(ş)Ar*, e.g. *birer* 'one each', *altışar* 'six each'.

Indefinite quantifiers include *çok* 'many, much', *az* 'few, little', *fazla* '(too) much, more', *birkaç* 'a few, several', *birçok* 'a good deal', *bazı* 'some'.

Personal Markers

A set of unaccented suffixes is added to nominal predicates, representing the number and person of the first actant: *-(y)Im* 'I', *-sIn* 'you', *Ø* 'he/she/it', *-(y)Iz* 'we', *-sInIz* 'you', *-lAr* 'they', e.g. *Türküüm* 'I am a Turk'. The addition of *-Dir* in the third-person singular and plural may, in formal style, just mark the predicative function, e.g. *Türktür* '(It is a fact that) (s)he is a Turk'. In informal style, it may express presumption, e.g. *Türktür* '(S)he must be a Turk' or 'I guess (s)he is a Turk'.

Simple past, conditional and indirective forms of the obsolete verb *i-* are used as copula particles, e.g. *idim, idin, idi, idik, idiniz, idiler* 'I was', etc., *isem, isen, ise, isek, iseniz, iseler* 'if I were, would be', etc., *imişim, imişsin, imiş, imişiz, imişsiniz, imişler* 'I obviously am/was', etc. These forms also have unaccentable and harmonic suffixed variants, e.g. *Hastaymışım < Hasta imişim* 'I am/was obviously ill'.

Verbals

Verb forms, simple and periphrastic ones, can express grammatical notions of actionality (durativity, etc.), voice (passive, reflexive, causative, cooperative-reciprocal), deontic modality (possibility, impossibility, necessitative), negation, viewpoint (intraterminal, post-terminal, prospective, etc.), mood (indicative, imperative, optative, conditional), tense (present, past), interrogation, epistemic modality (indirective, mnemonic, presumptive, etc.), followed by elements indicating person and number of the first actant.

Many of these categories can combine with each other. The possible combinations cannot be described in a simple scheme because of their high number. The following is not a detailed account of this rich system, but a summary of its main characteristics and central forms.

Formation of Verbal Stems

Verbal stems may be derived from nominal stems by productive suffixes such as *-lA*, e.g. *temizle-* 'cleanse' ← *temiz* 'clean', *kutla-* 'celebrate, congratulate' ← *kut* 'good luck, fortune'.

Another way of deriving verbal stems from nouns is compounding, i.e. using an auxiliary verb (pro-verb) such as *et-* or *yap-* 'do'. The latter technique is regularly applied when copying non-Turkish lexical items, e.g. *memnun et-* 'satisfy' (← Arabic), *spor yap-* 'do sport' (← French *sport*). Such compounds form one syntactic constituent and can, as a rule, only be separated by particles such as *DA* 'also' and *bile* 'even'.

The voices are cooperative-reciprocal, causative, reflexive and passive. Reciprocal stems are formed with *-(I)ş*, e.g. *görüş-* 'see each other, meet' ← *gör-* 'see'. Reciprocity can also be expressed in an analytic way with the reciprocal pronoun, e.g. *birbirini gör-* [each other:POSS3.:ACC. see] 'see each other'.

The most productive causative suffixes are *-Dir* and *-t*, the latter occurring after stems ending in vowels, *l* and *r*, e.g. *kestir-* 'make somebody cut' ← *kes-* 'cut', *yıkAt-* 'make somebody wash' ← *yıkA-* 'wash'. A group of monosyllables take *-Ir*, e.g. *bitir-* 'terminate' ← *bit-* 'come to an end'. Several causative suffixes may be combined, e.g. *bitirt-* 'make somebody finish something'.

The so-called passive suffixes are *-il* and *-(I)n*, the latter occurring after stems ending in *l* and vowels, e.g. *içil-* 'be drunk' ← *iç-* 'drink', *okun-* 'be read' ← *oku-* 'read', *alın-* 'be taken' ← *al-* 'take'.

So-called reflexives are also formed with *-(I)n*, e.g. *yıkan-* 'wash oneself' ← *yıkA-* 'wash'. Since the latter form may also mean 'be washed', the opposition between reflexives and passives is not clear-cut. Many so-called reflexives are not reflexive in a functional sense, but have lexicalised meanings, e.g. *sevin-* 'be glad' ← *sev-* 'love'. Reflexivity can be expressed unequivocally in an analytic way with the pronoun *kendi*, e.g. *kendini yıka-* [self:POSS3.:ACC.wash] 'wash oneself'.

Numerous suffix combinations are possible, e.g. reciprocal + causative + passive as in *seviştiril-* 'be caused to love each other'. One restriction prevents causative suffixes from succeeding passive suffixes if the latter actually have passive meanings (see p. 56).

Verbal stems are negated with *-mA*, which is unaccentable except in some negative aorist forms, e.g. '*Gélme* 'Do not come!', *Gelméz* '(S)he will not come'.

The suffix *-(y)Abil-*, expressing possibility or ability, goes back to the combination of the converb in *-(y)A* with the auxiliary *bil-* 'know', e.g. *verebil-* 'be able to give'. A syntactic property reflecting its earlier compound nature is that the particle *DA* 'also' may intervene, *gele de bildi* [come:A.CONV. also know:DI.PST.] 'could come, too'. Forms of impossibility are formed with *-(y)AmA-*, e.g. *vereme-* 'cannot give'. In this complex, *-(y)A* can synchronically be taken as an allomorph of *-(y)Abil-*. Negated verbal stem may also

occur before the possibility suffix, e.g. *Gelmeyebilir* [COME:NEG.:POSSIBILITY:R.PRES.] 'It may be that (s)he does not come'.

Lexical verbs have inherent actional contents. For example, transformatives denote telic actions (implying a natural turning-point), whereas non-transformatives lack this feature. The interpretation of viewpoint and tense forms is dependent on such actional meanings (see p. 42).

Periphrases containing a converb of the lexical verb and an auxiliary verb may be used to indicate modifications of inherent actionality, e.g. *yazıp dur-* [write:CONV. stand.POSTVERB.] 'keep on writing', *salıver-* [release:CONV.:give.POSTVERB.] 'release (suddenly), (just) let go'. The use of such 'postverbs' is relatively restricted in standard Turkish compared with the rich systems of actionality modifiers in other Turkic languages.

Finite Forms

To form finite items, thematic suffixes, e.g. *-DI* for simple past and *-miş* for indirective past, are attached directly to the verbal stem. Personal markers and copula particles are added to them, e.g. *Gel-di-niz* 'You have come' ← *gel-* 'come'. The system of finite forms will be discussed below.

Aspectotemporal Forms

Intraterminal forms include the usual present tense in *-(Ø)Iyor* (see p. 43): *İçiyorum, İçiyorsun, İçiyor, İçiyorumuz, İçiyorsunuz, İçiyorlar* 'I drink', etc. The present in *-mAktA(dIr)* often has a more focal meaning close to the English progressive, *Yazmaktayım* [write:MAKTA.PRES.:1SG.] 'I am writing'. The so-called aorist in *-(V)r* is a present with predominantly modal meanings, denoting events which the subject referent is ready to carry out or which are characteristic of him/her/it, e.g. *İçerim, İçersin, İçer, İçeriz, İçersiniz, İçerler* 'I drink, will drink, am one who drinks', etc.

One of the numerous past tense forms is the simple past in *-DI*. It competes with the complex forms *-(Ø)Iyordu* and *-(V)rdI*, based on *-(Ø)Iyor* and *-(V)r* + the past copula particle *idi*. While the simple past is a rather neutral way of presenting a past event, e.g. *Çıktı* '(S)he left', *-(Ø)Iyordu* expresses focused (focal) intraterminality, and typically refers to an event as going on at a past orientation point, e.g. *Çıkiyordu* '(S)he was (just) leaving'. The latter form does not say anything about the completion of the event, and thus leaves the question open whether the first actant referent actually left or not. The form in *-mAktAydi*, consisting of *-mAktA* + *idi*, is used similarly. The aorist past in *-(V)rdI* expresses a non-focal intraterminal view: the event is not necessarily going on at the past orientation point chosen, but is rather typical of that point, e.g. *Beşte çıkardı* [five:LOC.leave:R.COP.:DI.PPST.] '(S)he used to leave/would leave at five'.

The form *-miş* combines *-miş* with the past copula particle, designating a post-terminal view in the past, often corresponding to an English pluperfect, e.g. *İçmişti* '(S)he had drunk'. It lacks all indirective nuances. The same is true of the combination of *-miş* and *-DIr* as a past tense form in formal registers

(administrative texts, newspapers). In the spoken language, *-miş* may have presumptive meaning, e.g. *İçmiştir* 'I guess (s)he has drunk'.

The stem *-miş* may be extended by auxiliary verbs such as *ol-* 'be, become' and *bulun-* 'be', e.g. *Seçilmiş oldu* '(S)he therewith was elected', *Ölmüş bulunuyor* '(S)he is in the state of having died'.

The complex *-DIydI*, consisting of the simple past and the past copula particle, expresses a 'mnemonic' past, often implying something like 'as far as I can remember', e.g. *İçtiydik* 'We drank (at that time, long ago)'.

The suffix *-(y)AcAK* signals a prospective view, sometimes with a modal meaning of obligation. It can be used to refer to future events, e.g. *İçeceğim, İçeceksin, İçecek, İçeceğiz, İçeceksiniz, İçecekler* 'I will/shall drink', etc. There are corresponding past forms, consisting of *-(y)AcAK* and the past copula particle, e.g. *İçecekti* '(S)he would/should/was to drink'.

Modal Forms

The optative forms of *yap-* 'do', 'let me do', etc.:

Table 11.6 Optative

	Singular	Plural
1p.	yapayım	yapalım
2p.	yapasın	yapasınız
3p.	yapa	yapalar

The imperative forms of *koş-* 'run' are: 2p.sg. *Koş* 'Run!', 2p.pl. *Koşun(uz)* 'Run!', 3p.sg. *Koşsun* 'Let him/her/it run!', 3p.pl. *Koşsunlar* 'Let them run!'.

As for the necessitative in *-mAIlI*, it can also be used in impersonal constructions, e.g. *Burada çok çalışmalı* [here:LOC. much work:NEC.] 'Here one has to work a lot'.

When added directly to a lexical verbal stem, *-miş* refers to past events with an indirective (e.g. inferential) meaning, e.g. *İçmişim, İçmişsin, İçmiş, İçmişiz, İçmişsiniz, İçmişler* 'I obviously drank', etc. Indirectivity can also be expressed by adding the copula particle *imiş* to nominal forms. This particle is temporally neutral, referring to both present and past events, e.g. *Çıkiyormuş* [leave:İYOR:COP.MIŞ] '(S)he is/was obviously leaving', *Yazmaktaymışım* [write:MAKTA. COP.MIŞ:1SG.] 'I am/was obviously writing'.

Interrogation

Yes/no questions are marked with the unaccentable interrogative marker *-mI*. It precedes personal markers of the pronominal type and copula particles, e.g. *Aç mısınız?* [hungry:Q.:2PL.] 'Are you hungry?', *Beni seviyor musun?* [love:İYOR.PRES. Q.:2SG.] 'Do you love me?', *Aç değil miydiniz?* [hungry not Q.:COP.:DI.PST.:2PL.] 'Were you not hungry?', but it succeeds

other personal markers, e.g. *Uyumadın mı?* [sleep:NEG.:DI.PST.:2SG.Q.] 'Did you not sleep?'

Non-finite Forms

Turkish is rich in verbal nouns, verbal adjectives, and verbal adverbs. Some of the forms used can function as heads both in finite and non-finite clausal constructions and in noun phrases.

Among the verbal nouns, the ones in *-DİK* and *-(y)AcAK* are used, as a rule, with possessive suffixes, e.g. *oturduğu* 'his/her/its sitting', *yediğin* 'your eating', *oturacağı* 'his/her (future) sitting', *yiyeyeğin* 'your (future) eating'. Infinitives are formed with *-mA*K and *-mA*, e.g. *pişirmek* 'cook', *alma* 'take'. The shorter form can be used as a predicate core in fully-fledged non-finite clauses and carry subject representative suffixes. The verbal noun in *-(y)İş*, which also has this full predicative potential, mostly expresses the manner in which the action is carried out, e.g. *söyleyiş* 'way of speaking' ← *söyle-* 'say, speak'.

Participles are adjectival forms of verbs, formed with suffixes such as *-(y)An*, e.g. *oturan* '(who is) sitting' ← *otur-* 'sit, live', *yiyen* '(who is) eating' ← *ye-* 'eat'. The forms in *-DİK* and *-(y)AcAK* mentioned under verbal nouns are also used as adjectival forms, e.g. *oturacağın ev* 'a/the house you will live in'. The post-terminal participle in *-mİş* lacks all indirective meaning when used as a verbal adjective. In present-day Turkish, the so-called aorist participle occurs only in lexicalised forms, e.g. *yazar* 'writer' ← *yaz-* 'write', negative form: *çukmaz* 'blind alley' ← *çık-* 'go out, leave'.

Adverbial forms of verbs, converbs, are formed with suffixes such as *-(y)A*, *-(y)İp*, *-(y)ArAk*, *-(y)IncA* and *-(y)ken*. The suffixes *-(y)A*, *-(y)ArAk* and *-(y)ken* present events in an intraterminal perspective, as ongoing, e.g. *koşa koşa* 'running', *gülerek* 'laughing', *giderken* 'while going', whereas *-(y)IncA* presents an event terminally, at the moment it reaches its relevant limit, e.g. *girince* '(at the moment of) entering'. The functions of the converb in *-(y)İp* will be dealt with below.

Conditionals, which are also adverbial verb forms, are formed with *-sA* attached directly to the verbal stem, e.g. *yapsa* 'if (s)he happens to do' (hypothetical condition). Complex conditionals are formed with *-sA* + a copula particle, e.g. *yapsaydı* 'if (s)he would do/would have done' (counterfactual condition), or with thematic stems of the verb + a conditional copula particle *ise*, e.g. *geliyorsa* 'if (s)he comes', *gelmişse* 'if (s)he has come' (real condition).

Indeclinable Word Classes

Adverbs are of different types. One type is derived from nominal forms by means of *-CA* (see p. 210), e.g. *çocukça* 'in a childish way' ← *çocuk* 'child'. Many adverbs are accented on the first syllable, e.g. *'ancak* 'only'. Numerous adverbs have been copied from Arabic and Persian, e.g. *daima* 'always' ← Arabic.

Enclitic particles, e.g. *DA* 'and, also', *ki* 'that' etc., *bile* 'even', *ise* ~ *-(y)sA* 'as for', and the copula particles, serve different functions, which are dealt with under other sections in this chapter. When following immediately upon the first word of a clause, the so-called 'Wackernagel position', certain particles function as clause junctors, e.g. *Çok da yakışıklı* 'And he is very good-looking', *Bir de güzel bir kız var* 'Furthermore, there is a pretty girl there'. Some enclitic particles such as *DA* obey the rules of front vs. back harmony, e.g. *ben de* 'me too', *o da* '(s)he too'.

A number of elements may occur as self-contained utterances, interjections, including greetings, e.g. *Mérhaba* 'Hello', conversational formulas, e.g. *Lütfen* 'Please!', *Háydi* 'Come on!', *Évet* 'Yes', *Háyr* 'No', *İnşallah* 'God willing', and other emotional expressions, e.g. *Váh, vah* 'Too bad', *A'mán* 'Oh, mercy!'

The few conjunctions (free junctors) used in Turkish are mostly copied from Persian and Arabic, e.g. the coordinative conjunction *ve* 'and'. Subordinate clauses are not introduced by conjunctions, but marked by non-finite verbal forms as subjunctors (see pp. 229–33). Junctors such as *çünkü* 'because', *ama* 'but' and *eğer* 'if' will be discussed below.

Syntax

Turkish has an elaborate and highly regular system of sentence constructions of a left-branching, i.e. head-final, type. Some typological characteristics are: Turkish is a nominative-accusative language (see p. 52). Modified categories are preceded by their modifiers. Non-head constituents may be scrambled relatively freely in clauses. Heads of constructions are marked, e.g., for case and number. Pronominal arguments, both first actants and other actants, are not necessarily expressed in clauses as syntactic subjects and non-subjects. Thus no dummy elements such as English *it* are used to fill syntactic positions.

Nominal Phrases

With regard to external syntactic properties, nominal phrases are characterised by being able to function as arguments of predicates. They can carry case suffixes, participate as constituents in nominal compounds, and function as complements of postpositions.

Some non-finite clausal constructions based on non-finite verb forms – headless relative clauses, complement clauses – share these external properties, but differ from nominal phrases by having an internal clausal structure (see pp. 59–63, 229–30).

The basic word-order property of nominal phrases is the final position of the head category, i.e. determiners, complements and modifiers precede the lexical head. The latter can be succeeded by certain particles.

The head of the nominal phrase may be a noun, *doktor* 'doctor', a nominal compound, *kadın doktoru* 'gynaecologist' or an apposition expressing

identity, e.g. *kadın doktor* 'female doctor'.

Nominal phrases may contain determiners, quantifiers, attributive modifiers. There is no definite article. The numeral *bir* 'one' is used as an indefinite article, e.g. *güzel bir kız* [pretty a girl] 'a pretty girl'. The indefinite article is not used in predicative expressions such as *Öğretmenim* 'I am a teacher'.

Demonstrative pronouns can be used in definite expressions, e.g. *şu/bu/sokakta* 'in this (here)/this/that street'. There are also other quantifiers restricting the referential properties of nominal phrases, e.g. *başka* 'other', *diğer* 'other', *öbür* 'the other, the next', *aynı* 'same', *bütün* 'whole, all', *tüm* 'whole', *her* 'each, every', *herbir* 'each one', *herhangi bir* 'whichever'.

Determiners and Quantifiers

Some determiners and quantifiers, e.g. *diğer* 'other', indicate specificity of the nominal phrase. A specific nominal phrase is in the accusative when functioning as direct object. With *başka*, which also means 'other', it is possible to say both *Başka olaylar gördüm* [other event:PL. see.DI.PST.:1SG.] 'I saw other (non-specific) events' and *Başka olayları gördüm* [other event:PL.ACC. see.DI.PST.:1SG.] 'I saw other (specific) events'. If the determiner *diğer* is used, only the specific reading is possible, e.g. *Diğer olayları gördüm* [other event:PL.ACC. see.DI.PST.:1SG.] 'I saw the other (specific) events'.

Quantifiers precede adjectival modifiers and follow demonstrative pronouns, e.g. *bu iki büyük karpuz* 'these two big water melons', *birkaç ilginç konu* 'some interesting topics'. The order of determiners and modifiers is relatively free. The indefinite *bir* 'a(n)' may both precede and follow adjectival modifiers, e.g. *güzel bir kız* and *bir güzel kız* 'a pretty girl', the latter order being compulsory when *bir* is used as the numeral 'one'.

The indefinite *bir* can also modify plural nominals to express different unspecified entities, notably objects of vague sensory perception, e.g. *bir şeyler* 'some (different, unspecified) things', *bir yerler* 'some places', *bir zamanlar* 'some times (long ago)', *bir ışıklar* 'some lights', *Biraz önce bir sesler duydum* [a little before a sound:PL. hear:DI.PST.:1SG.] 'A while ago I heard some (unidentifiable) sounds'.

Attributive Modifiers

Types of attributive modifiers are adjectives, e.g. *güzel yemek* 'good food', adjective phrases, e.g. *gerçekten çok güzel yemek* 'really very good food', relative clauses, e.g. *sana pişirdiğim yemek* [YOU:DAT. COOK:DIK.PART.POSS. 1SG. food] 'food which I prepared for you', adverbial expressions, e.g. *taştan duvar* [stone:ABL.wall] 'wall of stone', *akıllıca bir plan* [wise:DER. plan] 'an intelligent[ly made] plan', and postpositional expressions, e.g. *insan-şey ilişkisi üzerine bir araştırma* [man thing relation:POSS3. about.POSTP. an investigation] 'an investigation into the relation between man and things'. A special type of adjectival modifiers are 'bahuvrihi' compounds (see p. 62), e.g. *gözü açık adam* [eye:POSS3. open man] 'a clever man'.

Genitive constructions consist of a genitive attribute and a head noun carrying a possessive suffix representing the person and number of the possessor, e.g. *evin numarası* [house:GEN. number:POSS3.] 'the number of the house', *benim oğlum* [I:GEN. son: POSS. 1SG.] 'my son'.

This construction may be discontinuous, the attribute being separated from the head by intervening elements. In colloquial Turkish, the possessor may occur in postverbal position, e.g. *Numarasını unuttum evin* [number:POSS3.:ACC. forget:DI.PST.:1SG. house:GEN.] 'I have forgotten its number, the house's'. In some registers, the possessive suffix may be missing in constructions denoting family relations or not referring to real ownership, e.g. *Mehmed'in Fatma* [Mehmet:GEN. Fatma] 'Mehmed's (daughter/wife/sister) Fatma', *bizim köy* [we:GEN. village] 'our village'.

Partitive meaning can be expressed with the ablative and a possessive suffix, e.g. *adamlardan biri* [man:PL.:ABL. one:POSS3.] 'one of the men', with the genitive construction, *adamların biri* [man:PL.:GEN. one:POSS3.] 'one of the men', or with a possessive suffix, e.g. *birimiz* [one:POSS. 1PL.] 'one of us'.

Compounds

A genuinely Turkish way of forming new lexical items is represented by nominal compounds, in which two nouns express one notion together, the last member carrying a third-person possessive suffix, e.g. *yatak odası* [bed room:POSS3.] 'bedroom', *Pazar günü* [Sunday day:POSS3.] 'Sunday'. A compound may also contain another compound, e.g. *Türk Dil Kurumu* [Turk language society:POSS3.] 'The Turkish Language Society', in which case the possessive suffix marks both the compound *Dil Kurumu* 'language society' and its composition with *Türk* 'Turk'. This method is, in some specific layers of the vocabulary, replaced by pure juxtaposition of nouns, e.g. in geographical names such as *Edirnekapı* 'The Edirne (Adrianople) Gate' instead of *Edirne kapısı*, and in names of dishes, institutions, etc.

The order of constituents is the reverse in the possessive so-called *izâfat* construction, copied from Persian and still found in lexicalised forms, e.g. *nokta-i nazar* [point:COMPOSITIONAL ELEMENT view] 'point of view' (see p. 244).

Relative Clauses

Relative clauses are based on participles, which function as subjunctors fulfilling tasks comparable to those of English subordinative conjunctions, e.g. *gelen adam* [come:AN.PART. man] 'a/the man who comes' (cf. pp. 61–3). Reflexive pronouns may function as resumptive pronouns, e.g. *kendisiyle konuştuğum adam* [himself:with speak:DIK.PART.POSS. 1SG man] 'the man I talked to in person'. Since the distinction between restrictive and non-restrictive relative clauses is not encoded formally, the corresponding interpretation is context-dependent.

Relative subjunctors are of two types: (1) possessive subjunctors with a

possessive personal marker representing the first actant, and (2) non-possessive subjunctors lacking a personal marker. The participles in *-DIĞI* and *-(y)AcAĞI* are of the first type, that in *-(y)An* is of the second type.

The principles determining the choice between the two types can be summed up as follows.

The only absolute restriction which can be accounted for in terms of structural properties is that possessive subjunctors cannot be used if the head noun is coreferential with the first actant of the relative clause. In this case, only a non-possessive subjunctor is allowed, e.g. *okula giden kız* [school:DAT. go:AN.PART. girl] 'the girl who goes to school'.

The same rule applies when the head noun is coreferential with a genitive attribute of the first actant in the relative clause, e.g. *kızı hasta olan kadın* [daughter:POSS3. ill be:AN.PART. woman] 'the woman whose daughter is ill'. In *bize güveneceği şüpheli olan adam* [we:DAT. trust:ACA.K.PART.:POSS3. doubtful be:AN.PART. man] 'a/the man about whom it is doubtful whether he will trust us', the complement clause [(his) trusting us] is the subject of the relative clause. The head noun *adam* 'man' is coreferential with the first actant, i.e. the potential genitive attribute of the subject clause. Thus, only the non-possessive subjunctor can be used.

If the head noun is not coreferential with the first actant of the relative clause, both subjunctor types may in principle be chosen. A possessive subjunctor is normally chosen when the head noun is coreferential with an entity different from the first actant of the relative clause, e.g. *oturduğum yer* [live:DIK.PART.:POSS.1SG place] 'a/the place where I live'.

However, if the subject referent of the relative clause is not specific, the non-possessive subjunctor may be used, e.g. *su bulunan yer* [water be:AN.PART. place] 'a/the place where there is water'. If the verb of the relative clause is transitive, non-possessive subjunctors are avoided in present-day Turkish, e.g. **At yiyen ot* [horse eat:AN.PART. grass] with the attempted reading 'Grass that horses eat'.

If the subject referent is specific, only a possessive subjunctor can be used, e.g. *suyun bulunduğu yer* [water:GEN. find itself:DIK.PART.:POSS3. place] 'a/the place where the water is to be found'.

Both subjunctor types can occur in headless relative clauses. The non-possessive subjunctor is normally used with first-actant reference, the possessive one otherwise, e.g. *Beni arıyanı bilmiyorum* [I:ACC. seek:AN.PART.:ACC. know:NEG.:İYOR.PRES.:1SG.] 'I do not know (the one) who sought me', *Sana söylediklerine inanma* [you:DAT. say:DIK.PART.:PL.:POSS3.:DAT. believe:NEG.] 'Do not believe what (s)he has told you'.

The subjunctor *-(y)AcAK* may be used in a diathetically neutral way and also have modal meanings of destination or obligation, e.g. *yiğitlik gösterecek gün* [manhood show:ACA.K.PART. day] 'the day to show (one's) manhood', *içecek su* [drink:ACA.K.PART. water] 'water to drink'. Other subjunctors would necessarily be in the passive in corresponding constructions.

The Syntax of Case-Marking

Subjects of finite clauses are in the nominative. When the subject of a complement clause immediately precedes the verb, it may be in the nominative or the genitive depending on whether its referent is specific or not, e.g. *Misafir geldiğini duydum* [guest come:DIK.PART.:POSS3.:ACC. hear:DL.PST.:1SG.] 'I have heard that a guest/guests has/have come' vs. *Misafirin geldiğini duydum* [guest:GEN. come:DIK.PART.:POSS3.:ACC.hear:DL.PST.:1SG.] 'I have heard that the guest has come'. Subjects in other positions are assigned the genitive.

Objects immediately preceding the verb may be in the accusative or in the nominative depending on whether they are specific or not, e.g. *Kitabı okuyorum* [book:ACC. read:İYOR.PRES.:1SG.] 'I read the book' vs. *Kitap okuyorum* [book read:İYOR.PRES.:1SG.] 'I read a book/books' (see p. 209). Objects in other positions are, as a rule, in the accusative. In some cases, nominative objects may also be found sentence-initially.

Nominal phrases functioning as complements are assigned the nominative, accusative, dative, locative or ablative by some complement-taking category, a verb, a noun, a postposition or an adverb. The choice is often easily predictable from the semantics. Possessor attributes in genitive constructions are in the genitive.

Subjects in non-finite clauses that are based on possessive subjunctors and do not function adverbially are, as a rule, in the genitive, e.g. in complement clauses such as *Arkadaşım geldiğini duydum* [friend:POSS.1SG.:GEN. come:DIK.PART.:POSS3.:ACC. hear:DL.PST.:1SG.] 'I have heard that my friend has arrived' and in relative clauses such as *Arkadaşım geldiği tren* [friend:POSS.1SG.:GEN. come:DIK.PART.:POSS3. train] 'a/the train by which my friend comes/has come'.

Subjects of adverbial clauses are normally in the nominative even if they have possessive subjunctors, e.g. *Arkadaşım geldiğinde çok sevindim* [friend:POSS.1SG come:DIK.PART.:POSS3.:LOC. very be glad:DL.PST.:1SG.] 'I was very glad when my friend arrived'. However, adverbial clauses based on postpositions may also behave like nominal clauses with regard to case assignment of the subject, e.g. *Ali'nin bildiği kadarıyla* [Ali:GEN. know:DIK.PART.:POSS3. as far as] 'as far as Ali knows'. The syntactic choice between the nominative and the genitive may correspond to semantic differences, e.g. *babam odaya girdiği gibi* [father:POSS.1SG room:DAT. enter:DIK.PART.:POSS3. as] 'as soon as my father enters/entered the room' vs. *babamın konuştuğu gibi* [father:POSS.1SG.:GEN. speak:DIK.PART.:POSS3. like] 'like my father talks'.

Nominal phrases functioning adverbially may be assigned the dative, ablative and locative cases. Some adverbials, e.g. temporal expressions of duration, may be in the nominative, e.g. *Bir hafta çalışmadım* [a week work:IMPOSSIBILITY:DL.PST.:1SG.] 'I could not work for a week'.

Postpositional Phrases

Prototypical postpositional phrases consist of a nominal phrase followed by a postposition. Four main types can be distinguished on the basis of case-marking properties.

The syntax of the first type is similar to that of genitive constructions (see pp. 112–13). The postpositional construction *evin içinde* [house:GEN. IN POSTP.:POSS3.LOC.] ‘in the house’ is formed like *evin sahibi* [house:GEN. owner:POSS3.] ‘the owner of the house’. Both constructions include a genitive attribute and a possessive suffix attached to the head noun. An important difference concerns the case marking of the possessor. The possessor of a genitive construction must be in the genitive and is otherwise interpreted as a compound, e.g. *ev sahibi* [house owner:3POSS.] ‘house-owner’. In a postpositional phrase, the first nominal may also be in the nominative, e.g. *ev içinde* [house IN.POSTP.:POSS3.LOC.] ‘in a/the house’. The genitive case is, however, obligatory with pronouns, e.g. *onun içinde* [it:GEN. IN.POSTP.:POSS3.LOC.] ‘in it’.

Spatial nouns forming postpositions of this type include *alt* ‘underside’, *üst* ‘upper side’, *ara* ‘space between’, *arka* ‘space behind’, *ön* ‘space in front’, *karşı* ‘opposite side’, *yan* ‘side’, etc. Different cases are assigned depending on the semantics, e.g. *üstüne* [upper side.POSS3.:DAT.] ‘onto’, *üstünde* [upper side.POSS3.:LOC.] ‘on’, *üstünden* [upper side.POSS3.:ABL.] ‘from the top of’.

In the second type postpositions are lexicalised with a particular case suffix, e.g. *tarafından* ‘by, on the part of’, *bakımından* ‘from the point of view of’, *yüzünden* ‘because of’, *sayesinde* [shadow:POSS3.:LOC.] ‘thanks to’, *yerine* ‘instead of’, *uğruna* ‘for the sake of’. With certain postpositions of this type, the nominal can only be in the nominative unless it is a pronoun, e.g. *Mehmet hakkında* [Mehmet about.POSTP.:POSS3.:LOC.] ‘about Mehmet’. The possessive suffix attached to the postposition agrees with the nominal, e.g. *(benim) hakkında* [I:GEN. about.POSTP.:POSS.1SG.:LOC.] ‘about me’.

In the third type, the nominal is in the nominative unless it is a pronoun, and no productive possessive or case suffix is attached to the postposition, e.g. *çiçek gibi* ‘like a/the flower’, *benim gibi* ‘like me’. This type also includes *için* ‘for’, *kadar* ‘as much as’ and *ile* ‘with’.

The fourth type does not carry any possessive or case suffix, and takes a nominal in the dative or ablative, e.g. *köye doğru* [village:DAT. towards.POSTP.] ‘towards (a/the) village’, *bundan dolayı* [this:ABL. because.POSTP.] ‘because of this’. Such case assignment is typical of postpositions developed from adverbial expressions or converbs such as *göre* (‘seeing’) ‘according to’ ← *gör-* ‘see’, e.g. *radyoya göre* [radio:DAT. according to.POSTP.] ‘according to the radio’.

Postpositions function – similar to case suffixes – as grammatical markers on the preceding nominals. Some postpositions have developed suffixed variants. Thus, the postposition *ile* ‘with’ displays the harmonic variant *-(y)la*, e.g. *arabyla* ‘by car’, *uçakla* ‘by air’. The Ottoman counterpart of *için* ‘for’ had a suffixed variant *-çün*, which is now obsolete.

Verbal Phrases

The core of a verbal phrase may consist of a simple and/or compound lexical verb, e.g. *oku-* ‘read’ and *devam et-* ‘continue’. It can also be an extended, ‘analytical’ stem containing an auxiliary such as *ol-* ‘become, be’ and *bulun-* ‘be, find itself’, e.g. *vermekte bulun-* [give:INF.:LOC. be] ‘be giving/give’, *verecek ol-* [give:PROSP. be] ‘be going to give’. As noted above, there are also periphrastic forms, containing a converb of the lexical verb and a postverb, which modifies the actional content, e.g. *yazıp dur-* [write:IP.CONV. stand.POSTVERB.] ‘keep writing’. Periphrastic forms are contiguous, though some particles such as the interrogative *mi* and *DA* ‘also’ may intervene. The modified lexical verb precedes its modifier. Many periphrastic forms are conceived of as words, which may also be reflected in the orthography, e.g. *Uyuyakaldı* [sleep:A.CONV.:remain.POSTVERB.:DI.PST.] ‘(S)he fell/remained asleep’. The process of grammaticalisation is even more developed in *Düşeyazdı* [fall:A.CONV.:fail.POSTVERB.:DI.PST.] ‘(S)he almost fell’, whose element *yaz-* is not used as a lexical verb (‘fail’, etc.) any more.

The Sentence

Finite sentences may consist of one nominal or verbal predicate core, all other constituents including the subject, being optional, e.g. *Biliyor* [know:İYOR.PRES.] ‘[(S)he] knows [it]’. Whether an actant is expressed as a syntactic subject, object etc. or not depends on semantic and discourse–pragmatic factors.

Nominal predicates may carry a suffix representing the person and number of the first actant, e.g. *Yazarım* [writer:1SG.] ‘I am a writer’, *Yazarız* [writer:1PL.] ‘We are writers’. Other copulas are copula particles such as *idi* ‘was’ and forms of auxiliary verbs such as *ol-* ‘become, be’, e.g. *İstanbul’-daydık* [Istanbul:LOC.:COP.:PST.:1PL.] ‘We were in Istanbul’, *Orada olabilir* [there:LOC. be:POSSIBILITY:R.PRES.] ‘It can be there’.

While the person and number of the first actant, the potential overt subject, is mostly encoded in personal markers, non-first actants are often established on the basis of discourse–pragmatic information. Pronouns can be used for special communicative functions such as focusing, e.g. *Bunu ben biliyorum* [this:ACC. I know:İYOR.PRES.:1SG.] ‘I know this.’

Different types of subjectless clauses can be distinguished. A pronominal subject may be dropped. Some thematic stems may be used impersonally, e.g. *Onu ziyaret etmeli* [it:ACC. visit:NEC.] ‘One has to visit it’. Predicates based on passives of intransitives may lack a first actant, e.g. *Girilmez* [enter:PASS.:NEG.R.PRES.] (‘it is not entered’) ‘It is not permitted to enter’. Note again that there is no dummy subject in Turkish. A special case is represented by certain converb clauses which necessarily share the first actant with the matrix clause, e.g. *Oturup konuşmaya başladı* [sit down:IP.CONVERB talk:INF.:DAT. begin:DI.PST.] ‘(S)he sat down and began talking’.

Subordinated clauses are embedded by means of action nouns, participles

and converbs with postpositive subjunctors. Subjunctors may be simple suffixes or complex forms with case suffixes, postpositions and/or lexical items. The converb *diye* 'saying' has been grammaticalised as a junctor to indicate the insertion of finite clauses.

Constituent Order

Turkish has a non-strict SOV order in main clauses and a strict SOV order in constituent clauses. This unmarked order of major constituents implies that the subject is the topic and that the object is a part of the comment, which presents new information.

Non predicate-final order patterns, traditionally called *devrik cümle* 'inverted sentence', are used in various communicative functions and may be motivated by different factors. Indefinite subjects tend to be in immediate preverbal position, e.g. *Bunu kimse anlamaz* [this:ACC. nobody understand:NEG.R.PRES.] 'Nobody will understand this'. One or several constituents may follow the finite predicate, e.g. *Kimse anlamaz bunu* 'Nobody will understand this' (see also below).

While this order is normally not licensed in subordinate clauses, exceptions are observed in spoken variants. Thus, in *Girip odaya konuşmaya başladı* [enter:IP.CONV. room:DAT. speak:INF.:DAT. begin:DI.PST.] 'Entering into the room (s)he began to speak', the adverbial constituent *odaya* 'into the room' succeeds the converb of the subordinated clause.

In imperative clauses, the most emphatic constituent, the verb, is often in initial position, e.g. *Gel buraya!* [come here:DAT.] 'Come here!' Questions may also start with a verb, e.g. *Götüreyim mi seni onlara?* [take away:OPT.:1SG. Q. you:ACC. they:DAT.] 'Shall I take you to them?'

Certain sentence adverbials and particles tend to take either initial or final position, e.g. *Ya gelmezse?* [or come:NEG.R.PRES.:COND.] 'And what if (s)he does not come?', *Hasta değil ya?* [ill not or] 'S(h)e is not ill, is (s)he?', *Yağmur yağacak herhalde* [rain rain:PROSP. probably] 'It will probably rain.'

The Communicative Structure of the Sentence

The usual communicative structure of the sentence is topic preceding comment. Topic constituents are often separated from the rest of the sentence by a short pause, marked by a comma in writing. Topicalisation can be signalled by the copula particle *ise* or by the expression *-(y)A gelince*, e.g. *Ben ise bunun bir sakıncası yok* [I COP.:COND. a drawback:POSS3. not existent] 'As for me, it does not have any drawback'.

Focus constituents take the position immediately preceding the predicate, e.g. *Ali'yi dün ben gördüm* [Ali:ACC. yesterday I see:DI.PST.:1SG.] 'I saw Ali yesterday' as an answer to *Ali'yi dün kim gördü?* [Ali:ACC. yesterday who see:DI.PST.] 'Who saw Ali yesterday?' A contrastively focused constituent also takes this position, e.g. *Ali'yi o değil, ben gördüm* [Ali:ACC. (s)he not I see:DI.PST.:1SG.] 'Not (s)he, but I saw Ali yesterday'. Clefting in English and similar languages has its counterpart in relativisations such as *Seni arayan*

odur [you:ACC. look for:AN.PART. (s)he:DIR.] 'It is (s)he who is asking for you'.

A frequent type of construction which may be used by the speaker as a strategy to gain time puts the dummy category *şey* 'thing' in a referent-introducing or focusing preverbal position, and the missing lexical expression in postverbal position, e.g. *Şey alayım, peynir* [thing buy:OPT.:1SG. cheese] 'I shall buy, what is it called, cheese'.

A constituent following the predicate core can represent presupposed or old information, e.g. *Anası geldi çocuğun* [mother:POSS3. come:DI.PST. child:GEN.] 'The mother of the child came', *Biliyorum ben* [know:1YOR. PRES:1SG. I] 'I know (it)'. Afterthoughts can also be added to the sentence, which may result in word orders hardly acceptable in preverbal position, e.g. *Geldi anası çocuğun* [come:DI.PST. mother:POSS3. child:GEN.] 'She came, the child's mother', but **Anası çocuğun geldi* [mother:POSS3. child:GEN. come:DI.PST.].

Definiteness does not have any constant morphological or syntactic marker. The interpretation of a nominal phrase with respect to definiteness is dependent on the context and possibly on intonational patterns. The general tendency for clause-initial expressions to be interpreted as definite may also be observed in Turkish.

Referential categories encoded by morphological and syntactic means are individual plurality, marked by the plural suffix, individual singularity, indicated by the indefinite article, and specificity. Specificity of a direct object can be marked in immediate preverbal position by the accusative, e.g. *Mektubu aldım* [letter:ACC. get:DI.PST.:1SG.] 'I have got the letter', *Bir mektubu aldım* [a letter:ACC. get:DI.PST.:1SG.] 'I have got a [specific] letter'. Corresponding sentences with a nominative object are *Mektup aldım* [letter get:DI.PST.:1SG.] 'I have got a letter/letters' and *Bir mektup aldım* [a letter get:DI.PST.:1SG.] 'I have received a letter'. This function of accusative-marking is essentially limited to the preverbal position, since objects in other positions must, as a rule, be in the accusative.

Case-marking of subjects plays a similar role in many non-finite clause types. The immediate preverbal position allows a systematic relation between genitive-marking and specificity. Assignment of the genitive signals specificity of the subject referent of a complement clause, e.g. *Dün misafirin geldiğini gördük* [yesterday guest:GEN. come:DIK.PART.:POSS3.:ACC. see:DI.PST.:1PL.] 'We saw that the guest/a particular guest arrived yesterday'. When the subject is in the nominative, its referent is mostly interpreted as non-specific, e.g. *Dün misafir geldiğini gördük* [yesterday guest come:DIK.PART.:POSS3.:ACC. see:DI.PST.:1PL.] 'We saw that a guest/guests arrived yesterday'.

Agreement Rules

There is no agreement in number or case between an attribute and a modified nominal head. In genitive constructions, however, agreement between the

genitive attribute and the head is marked by possessive suffixes on the latter. Nouns following cardinal numerals appear, as a rule, in the singular, e.g. *iki eşek* 'two donkeys'. Plural can, however, be used in proper names and other fixed expressions, e.g. *Yedi Cüceler* 'the Seven Dwarfs'.

In the first and second persons, there is agreement between pronominal subjects and verbal personal markers in person and number. The third person is mostly represented by a zero personal marker. Plural agreement is not obligatory. It is less frequent with inanimate subject referents and with subjects standing close to the predicate. A morphological restriction is that a word form cannot contain more than one plural suffix. A *-lar* suffix on a nominal predicate is thus ambiguous between marking the plural of the nominal phrase and representing a third-person plural subject. Though the issue has not been studied systematically, it seems that accentuation may disambiguate the construction, since the nominal plural suffix is accentable and the personal marker tends to be unaccented, e.g. *Öğretmenler* [teacher:PL.] '(These are the) teachers' vs. *Öğretmenler* [teacher:COPI.PART:PL.] '(They are) teachers'.

Expression of 'Being' and 'Having'

'Being' in the copulative sense is expressed with personal markers attached to the predicate, e.g. *Buradayım* [here:LOC.:1SG.] 'I am here', in the third-person singular \emptyset , e.g. *Burada* 'It is here'. Tense and mood may be modified with copula particles and verbs, e.g. *Buradaydım* [here:LOC.:COPI.DI.PST.:1SG.] 'I was here'. The negation is expressed by *değil* 'not', e.g. *Burada değilim* [here NOT:1SG.] 'I am not here'.

Existential predications are expressed with the adjectives *var* 'existent' and *yok* 'non-existent', e.g. *Para var* [money existent] 'There is money', *Sorun yok* [problem non-existent] 'There is no problem'. These elements are also used to express 'having', e.g. *Param var* [money:POSS.1SG. existent] 'I have money', *Param yok* [money:POSS.1SG. non-existent] 'I have no money'. If the possessor is expressed, it is in the genitive, *Memleketin parası yok* [country:GEN. money:POSS3. non-existent] 'The country does not have money'. Non-finite counterparts of *var* and *yok* are formed with *ol-* 'become, be' and the negation *olma-*, e.g. *param olmadığı* [money:POSS.1SG. be:NEG.DIK. PART.:POSS3.] '(the fact) that I have no money'.

Non-declarative Sentence Types

Polar interrogative sentences are formed with the unaccentable interrogative particle *-mi*. In verb-final position, it questions the whole sentence, e.g. *Siz eve gidiyor musunuz?* [you home:DAT. go:IYOR.PRES. Q.:2.PL.] 'Do you go home?' Otherwise its scope is narrow, not extending beyond the constituent it is added to, e.g. *Siz eve mi gidiyorsunuz?* [you home:DAT. Q. go:IYOR.PRES.:2.PL.] 'Do you go home?', *Siz mi eve gidiyorsunuz?* [you Q. home:DAT. go:IYOR.PRES.:2.PL.] 'Do you go home?'

Interrogative pronouns include *kim* 'who', *ne* 'what', *neden* 'why', *niçin*

'why', *nerede* 'where', *hangi* 'which', etc. They either take the focus position immediately before the verb, or stand in sentence-initial position, e.g. *Sana bunu kim söyledi?* [you:DAT. this:ACC. who tell:DI.PST.], *Kim sana bunu söyledi?* 'Who told you this?'

Embedded interrogative sentences are formed with special compounds consisting of an affirmative and a negative form of the same lexical verb, e.g. *Gidip gitmeyeceğimi sordu* [go:IP.CONV. go:NEG.:ACAK.PART.:POSS.1SG.:ACC. ask:DI.PST.] 'He asked whether I would go or not', where the converb *gidip* ← *git-* 'go' is followed by the negated action noun *gitmeyecek*.

The Syntax of Negation

As noted above, verbals are negated with *-mA*, and nominals with *değil* 'not', and existential predications with *yok* 'non-existent'. The last two elements can be followed by copulas, e.g. *Evde yoktu* [home:LOC. non-existent:COPI.DI.PST.:3SG.] '(S)he was not at home', *Evde değildim* [home:LOC. not:COPI.PST.:1SG.] 'I was not at home', *Ölecek değilim* [die:ACAK.PART. not:1SG.] 'I am not to die', *Dönmeyecek değilim* [return:NEG.:ACAK.PART. not:1SG.] 'It is not the case that I am not to return' (double negation). Contrasted constituents are also negated with *değil*, e.g. *Dün değil, geçen hafta geldi* [yesterday not last week come:DI.PST.] 'It was not yesterday but last week that (s)he came', *Ali değil, Mehmet geldi* [Ali not Mehmet come:DI.PST.] 'It was not Ali but Mehmet who came'.

The use of the particle *hiç* 'nothing, no ... at all' gives more emphasis to negations, e.g. *Hiç param yok* [no money:POSS.1SG. non-existent] 'I do not have any money at all', *Hiç kimse gelmedi* [no anybody come:NEG.:DI.PST.] 'Nobody has come'.

Coordination

A genuine Turkish way of expressing 'and' relations is based on the postposition *ile* 'with' attached to the first element, e.g. *onunla ben* [he:GEN.:with I] '(s)he and I', *kaşla göz arasında* [eyebrow:with eye between:POSS3.:LOC.] 'lit. between the eyebrow and the eye', i.e. 'in a trice'. The first constituents are postpositional phrases syntactically subordinated to the second constituents. Case suffixes are only added to the head, e.g. *Dostumla beni davet etti* [friend:POSS.1SG.:with I:ACC. invite:DI.PST.] '(S)he invited my friend and me'.

Syntactic coordination involves the use of the copied conjunctive *ve* 'and', e.g. *Sinan ve Ali* 'Sinan and Ali'. Case can be assigned to both or only the second element, e.g. *Sinan('i) ve Ali'yi gördüm* [Sinan (:ACC.) and Ali:ACC. see:DI.PST.:1SG.] 'I saw Sinan and Ali'.

Sharing of declension suffixes ('suspended affixation') is also possible between other coordinated elements, e.g. *-(y)Im in iyi çalışmalar diler, selamlarımı sunarım* [good work:PL. wish:R.PRES. greeting:POSS.1SG.:ACC. forward:R.PRES.:1SG.] 'I wish good work and send my greetings'. Predications

of equal rank are often combined by means of non-modifying converbs (see pp. 64–5, 231–2).

Various types of coordination are signalled by simple junctors such as *ya* 'or', *ama* 'but' and discontinuous junctors such as *DA ... DA* 'both ... and', *ya ... ya* 'either ... or', *ne ... ne (de)* 'neither ... nor', e.g. *hayat ya ölüm* 'life or death', *ben de sen de* 'both me and you', *ya ben ya sen* 'either me or you', *ne ben ne de sen* 'neither me nor you'.

Diathetic Relations

Passive clauses are primarily used to present events without expressing who or what carries them out. This means, in grammatical terms, that the first actant of the non-diathetic verb is not a potential overt subject of the passive verb. A second actant of a non-diathetic transitive verb is promoted to first actant of the corresponding passive verb and may thus function as grammatical subject. Thus, in *Adam görüldü* [man see:PASS.:DI.PST.] 'The man was seen', *adam* 'man' is the first actant of *görül-* 'be seen' ← *gör-* 'see'. Since passive intransitives lack a first actant, clauses based on them are subjectless, e.g. *Burada güzel yaşıyor* [here fine live:PASS.:İYOR.PRES.] 'One lives well here', 'Life is fine here' (see p. 55).

The first actant of the non-diathetic verb may be expressed, in the corresponding passive sentence, as an adverbial phrase marked with *-CA* or *tarafından* 'on the part of, by', e.g. *Topkapı Sarayı hangi padişah tarafından yapıldı?* [Topkapı Sarayı which sultan by.POSTP. make:PASS.:DI.PST.] 'By which sultan was the Topkapı Sarayı built?', *Türk Dil Kurumu'nca yayımlanan bir sözlük* [Turkish Language Society:CA.DER. publish:PASS.:AN.PART. a dictionary] 'a dictionary published by the Turkish Language Society'. Such agent expressions are much less frequently used than English ones (with *by*).

The argument structure of a causative verb is expanded in comparison with the non-diathetic form. The causative of a non-transitive verb is transitive, e.g. *bitir-* 'terminate' ← *bit-* 'come to an end'. The causative of a transitive verb introduces a new first actant, the causer or initiator of the action, while the first actant of the non-diathetic form is demoted and mostly expressed by a dative phrase, e.g. *Ali salatayı bana yedirdi* [Ali salad:ACC. I:DAT. eat:CAUS.:DI.PST.] 'Ali had me eat the salad'.

If the non-diathetic verb has a dative complement, the case assignment in the causative clause follows two patterns. The causee agent can be assigned the accusative and the dative of the complement is retained, e.g. *Çocuğu derse başlattım* [child:ACC. lesson:DAT. start:CAUS.:DI.PST.:1SG.] 'I made the child start the lesson' ← *Çocuk derse başladı* [child lesson:DAT. start:DI.PST.] 'The child started the lesson'. In the other pattern, the causee agent is assigned the dative and the original complement is in the accusative: *Dersi çocuğa başlattım* [lesson:ACC. child:DAT. start:CAUS.:DI.PST.:1SG.] 'I made the child start the lesson'. The choice of pattern is determined by discourse-pragmatic and semantic factors such as topicalisation, specificity and animateness.

Clause Combining

Clauses may be set in relation to each other in a discourse with (1) juxtaposition, (2) non-subordinating conjoining devices and (3) subordinating devices.

Pure juxtaposition as in *Ali gelemiyor. Çalışıyor* 'Ali does not come. He works' suggests a causal interpretation although this relation is not signalled explicitly. Conjunctional adverbs, or 'adjunctors', can specify the semantic relation between the juxtaposed clauses, e.g. *da ondan* 'that's why', *onun için* 'therefore', *şundan* 'for the following reason' in examples such as *Ali gelemiyor. Çalışıyor da ondan*. 'Ali cannot come. He is working, that is why', *Ali çalışıyor. Onun için gelemiyor* 'Ali is working. Therefore, he cannot come', *Ali şundan gelemiyor: çalışıyor* 'Ali cannot come for the following reason: he is working'. Patterns of these types are frequently used in spoken Turkish.

Another type of clause combining is conjoining. In conjoined structures, clauses combine in a syntactically non-subordinative way, with conjunctors to signal semantic relations between them. In *Ali gelemiyor çünkü çalışıyor* 'Ali cannot come for he is working', the causal conjunctive *çünkü* 'for' is used.

The syntactic relation between the clauses in these conjoined structures is different from the relation found in 'and' coordination and subordination. *Ali gelemiyor çünkü çalışıyor* does not function as one syntagm and cannot be subordinated as the complement of a verb such as *biliyorum* 'I know': *[*Ali gelemiyor çünkü çalıştığını*] *biliyorum*. This is, however, possible with coordinated structures as in [*Ali'nin geldiğini ve burada çalıştığını*] *biliyorum* [Ali:GEN. come:DIK.PART.:POSS3.:ACC. and here work:DIK.PART.:POSS3.:ACC. know:İYOR.PRES.:1SG.] 'I know that Ali has come and works here'. It is also possible with structures containing a subordinated clause [*Ali'nin gelip burada çalıştığını*] *biliyorum* [Ali:GEN. come:IP.CONV. here work:DIK.PART.:POSS3.:ACC. know:İYOR.PRES.:1SG.] 'I know that Ali has come and works here'.

The position of conjunctors is not restricted to the clause-initial position, e.g. *Çalışıyor çünkü* 'He is working, you see'. The conjunctive *ki*, which has a very vague semantic content, often belongs, as signalled by the intonation pattern, to the first of two conjoined clauses, e.g. *bir adam ki // ona güven olmaz* 'such a man that (he) cannot be trusted'. It can introduce a clause that adds a piece of information about an entity already mentioned, e.g. *bir adam ki bu adamı çok iyi biliyorum* 'a man, [namely] a man I know very well'. It can also be used as a particle without conjoining functions in the strict sense, e.g. in elliptic structures with suppression of the second clause, e.g. *O kadar güldük ki* 'We laughed that much'.

Types of Non-finite Clauses

Subordinated clauses are typically prepositive non-finitised clauses, based on action nouns, participles and converbs. The non-finite endings of these

verb forms act as syntactic subjunctors, i.e. can be regarded as bound complementisers.

Complement Clauses Non-finite clauses have clause-like internal structures with respect to their predicative potential. Some of them function as nominal phrases which can be assigned case and be complements of predicates. The following types can function as subject or object complements:

- 1 Clauses based on forms such as *-DİK-* and *-(y)AcAK*, e.g. *Öldüğü doğru değil* [die:DİK.PART.:POSS3. true not] 'It is not true that (s)he has died', *Geldiğini biliyorum* [come:DİK.PART.:POSS3.:ACC. know:İYOR.PRES.:1SG.] 'I know that (s)he comes/has come/came', *Geleceğini biliyorum* [come:ACAK.PART.:POSS3.:ACC. know:İYOR.PRES.:1SG.], 'I know that (s)he will come'.
- 2 Clauses based on infinitives, e.g. *Yola çıkmayı istemedim* [road:DAT. leave:INF.:ACC. want:NEG.:DI.PST.:1SG.] 'I didn't want to set out', *Burada çok çalışmak lazım* [here much work:INF. necessary] 'One has to work a lot here', *Gelmemi söyledi* [come:INF.:POSS.1SG.:ACC. say:DI.PST.] 'S(h)e said I should come'.
- 3 Clauses based on the non-finite form in *-(y)İş*, e.g. *Elinden sık sık kitabı düşürüşü beni güldürdü* [hand:POSS3.:ABL. often book:ACC. fall:CAUS.İŞ:3POSS.1:ACC. laugh:CAUS.:DI.PST.] 'His/her (way of) letting the book fall out of his/her hand again and again made me laugh'.

Non-finite clause constructions should not be confused with noun phrases whose head is a lexicalised verbal noun, e.g. *Ali'nin bugünkü ilginç konuşması* [Ali:GEN. today:DER. interesting speak:INF.:POSS3.] 'Ali's interesting speech today', where the head is *konuşma* 'speech' ← *konuş-* 'speak'. Since the attributive form *bugünkü* 'today's' cannot be used to modify verbals, it is clear that the noun phrase *Ali'nin bugünkü ilginç konuşması* has the same properties with respect to its internal structure as phrases headed by nouns. In a corresponding clausal construction headed by the non-finite verb form *konuşma* 'speaking', however, only the adverb *bugün* can be used: *Ali'nin bugün konuşması beni çok sevindiriyor* [Ali:GEN. today speak:INF.:POSS3.1:ACC. much like:REFL.:CAUS.:İYOR.PRES.] 'It makes me happy that Ali will speak today'.

Whenever there is a choice between *-DİK* and *-mA*, the clause based on *-DİK* is semantically marked as 'factive', e.g. *Dostunun beklediğini söyledi* [friend:POSS3.:GEN. wait:DİK.PART.:3POSS.:ACC. say:DI.PST.] '(S)he said that his/her friend waited', *Öldürüldüğü doğru* [die:CAUS.:PASS.:DİK.PART.:3POSS. true] 'It is true that (s)he is/was killed'. Corresponding clauses based on *-mA* have different meanings: *Dostunun beklemesini söyledi* [friend:POSS3.:GEN. wait:INF.:3POSS.:ACC. say:DI.PST.] '(S)he said that his/her friend should wait', *Öldürülmesi doğru* [die:CAUS.:PASS.:INF.:POSS3. right] 'It's right for him/her to be killed'.

Non-finite clauses also include relative clauses, which have already been dealt with (see pp. 219–20). Headless relative clauses may function as complement clauses *Söylediklerimi duymadı* [say:DİK.PART.PL.:POSS.1SG.:ACC. hear:NEG.DI.PST.] '(S)he did not hear what I said'. Infinitive clauses may also function as attributes in genitive constructions, e.g. *sinava girmemin amacı* [exam:DAT. enter:INF.:POSS.1SG.GEN. purpose:POSS3.] 'the purpose of my going up for the exam'.

Subordinated Finite Clauses Subordinated constructions usually contain non-finite verb forms. However, the utterance verb *de-* 'say' and verbs of belief such as *san-* 'believe' can take finite complements, e.g. *Ona hacet yoktur dedi* [it:DAT. need nonexistent:DIR. say:DI.PST.] 'There is no need for it, (s)he said'; cf. the corresponding non-finite construction *Ona hacet olmadığını söyledi* [it:DAT. need be:NEG.:DİK.PART.:POSS3.:ACC. say:DI.PST.] '(S)he said there was no need for it'.

Other utterance predicates can embed finite clauses by means of the subjunctor particle *diye* 'saying', e.g. *Geliyor mu diye soruyorum* [come:İYOR.PRES. Q. say:A.CONV. ask:İYOR.PRES.:1SG.] 'I ask: is (s)he coming?', *Okusun diye aldım* [read:IMP. say:A.CONV. take:DI.PST.:1SG.] ('I bought it saying (s)he should read it) 'I bought it in order for him/her to read it'.

Verbs such as *iste-* 'want' can take optative clauses, e.g. *Kendime geleyim istedim* [self:POSS.1SG.:DAT come:OPT.:1SG. want:DI.PST.] 'I wanted to come to myself again'. Some postpositions can take finite complements, e.g. *Aşık olmuştun gibi davranıyorsun* [in love be:MİŞ.PST.2SG. as.POSTP. behave:İYOR.PRES.2SG.] 'You behave as if you were in love'.

With matrix clauses containing certain verbs, e.g. *san-* 'believe' and *bil-* 'know', the embedded clause can have a finite verb form, e.g. *Seni öldü(n) sanıyordum* [you:ACC. die:DI.PST.:(2SG.) believe:DI.PST.:1SG.] 'I believed you to be dead' (so-called 'raising' constructions).

Converb Clauses Converbs function as predicative cores, expandable according to the valency pattern of the verb and linking predications to matrix constructions at different syntactic levels. They are formed with a specific set of subjunctors, simple markers, consisting of one suffix morpheme, or complex ones, containing more, even free, morphemes. Complex converb subjunctors may even be based on simple ones, e.g. *-(y)IncAyA kadar* 'until', consisting of *-(y)IncA*, the dative suffix and the postposition *kadar* 'up to'.

Non-modifying Converb Clauses The converb in *-(y)İp* represents a 'syndetic' type which is widespread in all Turkic languages (see pp. 64–5, 117). It is conceptually vague, encoding nothing more specific than a kind of 'and' relation. The content of the converb clause is interpretable as being on a par with that of the matrix predication. This kind of linking is, in spite of

the syntactic embedding, similar to the semantic function of the English coordinative conjunction 'and'. Thus, *-(y)Ip* forms can often be interpreted as though they carried the same suffixes as the predicate of the matrix construction. Three main types of linking with *-(y)Ip* are discernible, none of them expressing a circumstantial modification of the matrix construction.

The first type yields complex predications of the structure SUBORDINATE PREDICATION + SUPERORDINATE PREDICATION, where each predication may have its own first actant, whether expressed as a subject or not, e.g. *Silâhlar çekilip kavga başladı* [weapon:PL. draw:PASS.:IP.CONV. fight begin:DI.PST.] 'Weapons were drawn and the fight began'.

This type is strongly restricted in modern Turkish. The restrictions are partly due to the semantic roles ascribable to the subject referents and partly to coreferentiality relations. The construction may be used if the subordinate verb is a passive or an 'unaccusative' verb describing a non-volitional action, e.g. *bat-* 'sink', *öl-* 'die', *ol-* 'become, be', *çık-* 'appear'. The subject referents of such verbs are typically 'non-agentive', i.e. not being the instigator of the action, but rather affected by it.

The construction may also be used if the first actant of the converb construction (whether expressed or not) is coreferential with an actant of the matrix construction (whether expressed or not). There is often an expressed or implied possessive relationship between the actants, e.g. *Bu kitap yüz sayfa olup fiyatı iki bin liradır* [this book hundred page be:IP.CONV. price:3POSS. two thousand lira] 'This book contains one hundred pages and its price is two thousand lira', where there is a possessive relation between *kitap* 'book' and *fiyatı* 'its price', marked by the possessive suffix.

The second and most frequent type yields complex predicates of the structure SUBORDINATE PREDICATE + SUPERORDINATE PREDICATE, where a possible first actant, whether expressed as a subject or not, is shared and where each predicate core may be more or less expanded, e.g. *Biz de gidip şu hamamda yıkanacağız* [we also go:IP.CONV. this here bath wash:REFL.:PROSP.:1PL.] 'We will also go and wash ourselves in this bath'.

The third type yields complex verbs of the structure SUBORDINATE PREDICATE CORE + SUPERORDINATE PREDICATE CORE, in which only the converb, but not the matrix verb form, is expandable according to its normal predicative potential, e.g. *Ali yazıp durdu* [Ali write:IP.CONV. stand.POSTVERB.:DI.PST.] 'Ali kept on writing'. For this 'postverb' type, see pp. 42, 113–14.

Modifying Converb Clauses One of the numerous converb subjunctors that modify the matrix predication is *-(y)ArAk*, often used to describe circumstances of manner, e.g. *Koşarak geldi* [run:ARAK.CONV. come:DI.PST.] '(S)he came running'. In the modern language, *-(y)ArAk* has developed into a frequent converb which is taking over some functional domains previously occupied by *-(y)Ip*. The converb in *-(y)A* has a limited use today except in double forms such as in *Güle güle otur* [laugh:A.CONV. laugh:A.CONV. live:IMP.]

'Live happily!' (idiomatic expression uttered to somebody who has moved to a new house).

Converb clauses based on *-(y)ArAk* and *-(y)A* mostly have the same first actant as their matrix construction. If they do have a different first actant, it is on the same conditions as for the clauses based on *-(y)Ip*, e.g. *Geri kalan kıymadan köfte yapılarak lezzetli bir yemek hazırlamak da mümkündür* [back remain:AN.PART. minced meat:ABL. meat ball make:PASS.:ARAK.CONV. delicious a dish prepare:INF. also possible:DIR.] 'It is also possible to make a delicious dish by making meat balls of the minced meat that is left'.

The other converbs are exempt from such restrictions with respect to first actant choice. They will not be dealt with in detail here. There are special negative forms such as *-mAdAn* meaning 'not doing' or 'before doing', e.g. *Durmadan konuştu* [stop:MADAN.CONV. speak:DI.PST.] 'S(h)e spoke unceasingly'. A converb with the meaning 'as soon as' is formed by juxtaposition of the affirmative and the negative aorist, e.g. *girer girmez* [enter:R. enter:NEG.R.] 'as soon as ... enter(s)/entered'. The converb in *-(y)InçA* is terminal ('when', 'as soon as'), whereas the Ottoman predecessor *-(y)InçA* had terminative–limitative functions ('until', 'as long as'); cf. p. 195. Some converbs such as *-Diğl için* 'because [of the fact that]' include a possessive suffix marking the first actant, e.g. *Ayşe gelmediği için öfkeleni* [Ayşe come:NEG.:DIK:POSS3. because.POSTP. get angry:DI.PST.] 'He got angry because Ayşe did not come'. Note again that the subject of such adverbial clauses is in the nominative.

Conditional Clauses The verb in the protasis of conditional sentences is formed with *-sA* added to verbal stems and with the copula particle *ise* added to thematic stems. The use of *eğer* 'if' ← Persian *agar* is optional.

In clauses expressing hypothetical conditions, the simple *-sA* is used, e.g. *Gelse konuşuruz* [come:COND. speak:R.PRES.:1PL.] 'If (s)he should come, we will talk'. In clauses expressing irreal (counterfactual) conditions, *-sA* plus the past copula particle *idi* is used, e.g. *Zengin olsaydım sana bir yüzük alırdım* [rich be:COND.:COP.:DI.PST.:1SG. you:DAT. a ring buy:R.:COP.:DI.PST.:1SG.] 'If I were rich, I would buy you a ring', *Ölmeseydi tanışacaktık* [die:NEG.:COND.:COP.:DI.PST. get acquainted:PROSP.:COP.:DI.PST.:1PL.] 'Had (s)he not died, we would have got acquainted'. In clauses expressing real conditions, thematic stems plus the conditional copula particle *ise* are used, e.g. *Bakarsak görürüz* [look:R.:COP.COND.:1PL. see:R.PRES.:1PL.] 'If we look, we will see it'.

Concessive clauses are formed with the conditional and a particle such as *DA* 'also' or *bile* 'even', e.g. *Para gönderseler de almayacağım* [money send:COND.:3PL. also accept:NEG.:PROSP.:1SG.] 'Even if they should send money I am not going to accept it'.

Text Syntax

The grammatical system of Turkish provides the speaker with a great number of grammatical devices. The actual use of these devices is based on conventions that may allow the definition of sub-systems and discourse types. This differentiation is important for understanding how speakers of Turkish use the possibilities inherent in the grammar. In a given discourse type, a subset of possible forms is used, and the definition of their semantic functions depends on the actual possibilities of contrast given there. For discourse types based on the use of verbal forms, see Johanson 1971.

Questions of text linguistics, e.g. cohesion devices, anaphoric relations, are relatively little researched (see Slobin and Zimmer 1986; Van Schaak 1996). For propulsive (plot-advancing) discourse functions of verb forms, see Johanson (1998).

Lexicon

In addition to the native Turkic vocabulary, the Turkish lexicon is still characterised by a significant Arabic and Persian component (see Tietze 1990). The Turkish language reform has, however, weakened the dominance of this component (see Chapter 13). A great number of lexical elements have been copied into Turkish from Greek, Italian and French, recently also from English. For Turkish nautical terms of Italian and Greek origin, see Kahane, Kahane and Tietze 1988.

References and Further Reading

For more detailed information on relevant literature, the reader is also referred to the bibliographies given in the works listed below. An annual bibliography *Turkologischer Anzeiger/Turkology Annual* has been published since 1975 in Vienna and Budapest by G. Hazai and A. Tietze. A yearbook of Turkish linguistics, *Dilbilim arařtırmaları*, Ankara: Hitit Yayınevi, has been published since 1991 by G. Durmuřođlu, K. İmer, A. Kocaman and S. Özsoy. The proceedings of the biennial Turkish Linguistics Conferences have been published since 1986 (see Slobin and Zimmer 1986). A periodical covering various aspects of Turkish and Turkic linguistics is *Turkic Languages*, Wiesbaden: Harrassowitz.

- Banguođlu, T. (1986) *Türkçenin grameri*, Türk Dil Kurumu Yayınları 528, Ankara: Türk Tarih Kurumu.
 Brendemoen, B. and Hovdhaugen E. (1992) *Tyrkisk grammatikk*, Oslo: Universitetsforlaget.
 Čaušević, E. (1996) *Grammatika suvremenoga turskog jezika*, Zagreb: Hrvatska Sveučilišna Naklada.
 Deny, J. (1921) *Grammaire de la langue turque (dialecte osmanli)*, Paris: Éditions Ernest Leroux.

- (1955) *Principes de grammaire turque ('turk' de Turquie)*, Paris: Adrien-Maisonneuve.
 Ergin, M. (1972) *Türk dil bilgisi*, İstanbul: Minnetođlu Yayınları.
 Erguvanlı, E. (1984) *The Function of Word Order in Turkish Grammar*, Los Angeles: University of California Press.
 Gencan, T. N. (1966) *Dilbilgisi*, Türk Dil Kurumu Yayınları 243, İstanbul: Ahmet Saıt.
 Hazai, G. (1978) *Kurze Einführung in das Studium der türkischen Sprache*, Budapest: Akadémiai Kiadó.
 — (ed.) (1990) *Handbuch der türkischen Sprachwissenschaft 1*, Budapest: Akadémiai Kiadó, Wiesbaden: Harrassowitz.
 Johanson, L. (1971) *Aspekt im Türkischen. Vorstudien zu einer Beschreibung des türkeitürkischen Aspektsystems*, Studia Turcica Upsaliensia 1, Uppsala: Almqvist & Wiksell.
 — (1991) *Linguistische Beiträge zur Gesamtturkologie*, Budapest: Akadémiai Kiadó.
 — (1998) 'Viewpoint operators in European languages', in Ö. Dahl (ed.), *Tense and Aspect in the Languages of Europe*, Berlin, New York: Mouton de Gruyter.
 Kahane, H., Kahane, R. and Tietze, A. (1988) *The Lingua Franca in the Levant. Turkish Nautical Terms of Italian and Greek Origin*, İstanbul, Ankara, İzmir: ABS Kitabevi.
 Kononov, A. N. (1956) *Grammatika sovremennogo tureckogo literaturnogo jazyka*, Moscow-Leningrad: Izdatel'stvo Akademii Nauk SSSR.
 Kornfilt, J. (1990) 'Turkish and the Turkic languages', in B. Comrie (ed.), *The World's Major Languages*, New York, Oxford: Oxford University Press, 619–44.
 Lewis, G. L. (1967) *Turkish Grammar*, Oxford: Oxford University Press.
 — (1989) *Teach Yourself Turkish*, Sevenoaks: Hodder and Stoughton.
 Özsoy, S. (1996) *Türkçe konuşalım*, İstanbul: Boğaziçi Üniversitesi.
New Redhouse Turkish-English Dictionary (1968), İstanbul: Redhouse.
 Sezer, E. (1991) *Issues in Turkish Syntax*, Ph.D. thesis, Harvard University, Cambridge, MA.
 Slobin, D. I. and Zimmer K. (1986) *Studies in Turkish Linguistics*, Typological Studies in Language 8, Amsterdam, Philadelphia: John Benjamins.
 Swift, L. B. (1963) *A Reference Grammar of Modern Turkish*, Indiana University Publications, Uralic and Altaic Series 7, The Hague: Mouton.
Türkçe sözlük I-II (1988), Ankara: Türk Dil Kurumu.
 Tietze, A. (1963) *Turkish Literary Reader*, Indiana University Publications, Uralic and Altaic Series 22, Bloomington, The Hague: Mouton.
 — (1990) 'Die fremden Elemente im Osmanisch-Türkeitürkischen', in G. Hazai (ed.) 1990, 119–145.
 Underhill, R. (1976) *Turkish Grammar*, Cambridge (MA), London: MIT Press.
 Van Schaak, G. (1996) *Studies in Turkish Grammar*, Turcologica 28, Wiesbaden: Harrassowitz.
 Wendt, H. F. (1992) *Türkisch. Langenscheidts praktisches Lehrbuch*, Berlin, Munich, Vienna, Zurich, New York: Langenscheidt.

12 Turkish Dialects

Bernt Brendemoen

Introduction

Within all the scholarly disciplines comprised by the designation Turcology, study of language variation can be regarded as the oldest one. Thanks to Maḥmūd al Kāšyarī and his extraordinary work *Compendium of the Turkic Dialects*, we have a remarkably detailed picture of the different dialects and languages spoken by Turkic tribes in the eleventh century (see pp. 23–4). Nevertheless, dialectology in its modern sense, i.e. the study of local variation within one particular language, has been a neglected field within Turcology. Considering the fact that the Republic of Turkey until quite recently was the only country within the vast Turkic-speaking areas of the world with a Turkic variety as its national language, it could be expected that at least Anatolian dialectology would be a field where a lot of research could be done by the Turks themselves. Unfortunately this is not the case.

This lack of scholarly activity, or interest, in dialectology is especially striking if we take into consideration that *Türk Dil Kurumu*, 'The Turkish Language Society', in the initial years of the language reform in the 1930s, searched the different Anatolian dialects for pure Turkish words which could replace words of Arabic and Persian origin. School-teachers all over Turkey were encouraged to submit lists of dialect words, and the result is *Derleme Sözlüğü*, a 12-volume lexicon of Turkish dialects, or rather elements believed to be Turkish in those dialects. In fact, this dictionary is today also a gold mine for scholars who want to study words of foreign, e.g. Armenian or Greek, origin in Turkish dialects. However, its compilation had very little to do with dialectology.

Even later, the classification of the dialects into groups and sub-groups, or the compilation of a dialect map have not been regarded as tasks of The Turkish Language Society. One reason for the reluctance to support dialectological work may be that a general dialect survey has often been felt to be politically touchy. Another reason for the rather stepmotherly treatment of dialectology has to do with the low status dialects have in Turkish society. At the same time, the use and necessity of dialectological studies within a broader linguistic frame have not been fully recognised.

Nevertheless, a large number of valuable studies on different dialects have been carried out. The work done in dialectology up to now has mostly

consisted in collecting dialect material and preparing studies on single dialects. Comparative studies of dialect features aiming at describing variation within a greater area have been attempted to a very limited extent.

Due to this lack of exhaustive research, it is not yet possible to give an accurate survey of dialect features for all Turkish dialects. Likewise, a geographical classification of the dialects has to be rather tentative (see Kowalski 1934; Kral 1980; Boeschoten 1991; Karahan 1996). The following chapter is a synthesis of the most important Anatolian dialect features addressed by these scholars, weighed against my own views. The Balkan and Thracian dialects will not be dealt with here. Dialect forms will be compared with corresponding forms in Standard Turkish (ST), quoted in the official orthography.

Phonetics and Phonology

Vowels

In large parts of Anatolia, /*ɛ*/, an unrounded upper-mid front vowel corresponding to IPA [e], is preserved as a phoneme beside the unrounded mid-front /*e*/, allowing the existence of minimal pairs such as *ɛl* 'stranger, country' and *e*l 'hand' (see pp. 30, 90, 94–5, 183, 248). A wide range of *e* sounds is also found in standard Istanbul pronunciation, but without distinctive functions.

Tendencies towards reduced vowel inventories are common. Thus, Standard Turkish /*ö*/ and /*ü*/ frequently correspond to retracted back /*o*/ and /*u*/ or half-front vowels. The geographic distribution is complicated. In the whole of central Anatolia, retraction or semi-retraction seems to be almost a rule in initial syllables after front velar stops, e.g. *goz* 'eye' vs. ST *göz*. Whereas in central Anatolia the preceding stop usually does not retain any trace of palatalisation, it characteristically remains strongly palatalised in the province of Rize. In neighbouring Trabzon, retraction is very seldom after velar stops, but in other positions, especially word-initial, retraction or semi-retraction is also quite common here, e.g. *ordek* 'duck' vs. ST *ördek*, *uzum* 'grape' vs. ST *üzüm*.

In the same way, there is a strong tendency in Trabzon and Rize for *e* to correspond to ST *ö*, when followed by *r* or *l*, e.g. *el-* 'die' vs. ST *öl-*, and for *i* to correspond to *ü*, e.g. *irmaq* 'river' vs. ST *ırmak*.

Rounding caused by neighbouring labial consonants is very frequent all over Anatolia, e.g. *bül-* 'know' vs. ST *bil-*.

Consonants

The treatment of word-internal and final velar stops in back environments constitutes one of the most conspicuous isoglosses in Anatolia. Only western Anatolia, the Black Sea coast to the west of Sinop, and Trabzon and Rize have preserved *q*, while it has become a fricative *ç* or *ɣ* elsewhere.

Voicedness/non-voicedness of initial velar stops is also an important feature. These stops are very frequently voiced in back vowel environments in an area that seems to extend even further west than the previous feature, e.g. *ğarğa* 'crow' vs. ST *karga*. The distribution of voicedness vs. non-voicedness before front vowels is approximately the same as in Standard Turkish. Initial dental stops are subject to exactly the same voicing variation as the velar ones, e.g. *daş* 'stone' vs. ST *taş*, but this feature is less comprehensive than with the velars, even in areas where voicing of *q* seems to be especially frequent, e.g. in the northeast. Voicing variation in labial stops is less frequent. In most areas of Trabzon, initial stops, including labial ones, are unvoiced and unaspirated regardless of the quality of the subsequent vowel, at least when the word has a certain emphasis, e.g. *dur-* 'stand'.

Palatalisation of stops, especially of *k* and *g*, in the environment of front vowels, is considerably stronger in Rize, partly also in Trabzon, than in Standard Turkish. Such strongly palatalised stops have been transcribed as *d''* and *t''* by Räsänen, e.g. *d''üneş* 'sun' vs. ST *güneş*, *ast''ere* 'to the military' vs. ST *askere*. In the same area, there is also palatalisation in words such as *d'ed'e* 'grandfather' vs. ST *dede* and *an'n'e* 'mother' vs. ST *anne* (also palatalised).

In most east Anatolian dialects except for the very southeast as well as Trabzon and Rize, the originally voiced velar fricative that is mostly written as *ğ* and realised as *y* or \emptyset in Standard Turkish is pronounced as γ in back and as *g* in front environments, e.g. *ayaç* 'tree' vs. ST *ağaç*, *jiger* 'liver' vs. ST *ciğer*. The pronunciation γ also occurs sporadically in back environments in the central and west Anatolian dialects.

The old velar nasal η , which has become *n* in Standard Turkish, is preserved in most dialects except for the northeast Anatolian ones and certain dialect islands such as Balıkesir, Antalya and parts of Malatya. In Trabzon and Rize it has mostly become *n*, but has disappeared in some cases, e.g. *bāa*, *sāa*, *oa*, dative forms of *ben* 'I', *sen* 'you', *o* 'he, she, it' respectively. A different development is observed in other areas, e.g. in Urfa and Bartın, where the old η has sporadically developed into *y*, e.g. *babay* 'your father' vs. ST *baban*.

The liquid *r* is often dropped in syllable-final position, particularly in west Anatolia, most frequently in its central or northern parts. This feature diminishes the farther we move in the southern or eastern direction.

Consonant assimilations, which mostly affect liquids, are much more frequent in most Anatolian dialects than in Standard Turkish, e.g. *talle* 'field' vs. ST *tarla*, *ollar* 'they' vs. ST *onlar*.

Morphophonology

Front vs. back harmony in suffixes is mostly identical with the one found in Standard Turkish except in parts of Trabzon and Rize. Here, certain suffixes tend to be standardised as invariable front suffixes, e.g. *-ler* as a

third-person plural verbal suffix, less frequently the dative and locative suffixes. Others are fixed as non-harmonic back suffixes, e.g. the conditional suffix *-sa* and the negative suffix *-ma*. Similar examples are also found sporadically elsewhere.

In some areas, notably in some districts of Trabzon, labial harmony has been preserved at an archaic stage of development. As in Old Ottoman, most high vowel suffixes are either rounded or unrounded (cf. pp. 108, 185). This feature is complicated by the fact that *i* and *ü* tend to become *i* and *u* respectively, as stated above. Thus we find forms such as *tirenümi* 'my train (accusative)' vs. ST *trenimi*, *qoli* 'his arm' vs. ST *kolu*, *Yaptunuz* 'You did' vs. ST *Yaptınız*, but *Yaptile* 'They did' vs. ST *Yaptular*, *girdunsa* 'if you broke' or 'if you entered' vs. ST *kirdunsa*, *girdinse*. The choice of a suffix vowel may be decided by the stem and not necessarily by an intervening suffix, e.g. *qolina* 'to his arm' vs. ST *koluna*. Less strong reflexes of this earlier stage in the development of labial harmony are frequently found all over north Anatolia.

Morphology

In an area in and around the province of Denizli, the forms of the accusative and dative suffixes are reversed compared with Standard Turkish, being *-(y)A* and *-(y)I* respectively, e.g. *Bu elmaya vuranları bahşiş verilir* 'A reward is given to those who hit this apple' vs. ST *Bu elmayı vuranlara bahşiş verilir*. Since, for example, the third-person possessive suffix never appears with a low vowel and *beni* is found instead of *bana*, e.g. *Beni bir çay yap* 'Make me a cup of tea' vs. ST *Bana bir çay yap*, this seems to be an essentially morphological – and not a phonological – feature. The change might have been triggered by a phonetic development *e* > *i* in final position, and the case confusion caused by this might have been compensated for by a morphological change *-(y)I* > *-(y)A*, also affecting irregular forms such as the personal pronouns.

The present tense suffix, corresponding to Standard Turkish *-(I)yor*, constitutes one of the most clear-cut isoglosses in Anatolia. *-(I)yo* is the normal form in most of west and central Anatolia, while dialects further east have forms based on a high vowel suffix, mostly *-(I)yi(r)*. Along the Black Sea coast, also *-(I)yu(r)* and *-(I)yé(r)* are found. Examples: west Anatolia *Aliyon* and *Aliyörun*, central Anatolia *Aliyom*, Gaziantep *Aliym*, Diyarbakır *Aliyam*, Erzurum *Alīram*, Trabzon *Aluy(u)rum*, *Aliy(u)rum* (and also *Alayirum*, *Aliyam*). In the very southwest, present tense formation with auxiliary verbs such as *dur-* 'stand' and *gét-* 'go' is common, e.g. *Yatıpduruyun* 'I am lying', *Gédıpgéderik* 'We are going' (Demir 1993).

The first- and second-person forms of the copula also show considerable variation, particularly when suffixed to participial verbal stems. Thus, the first-person form, which is *-(y)Im* in Standard Turkish, is *-(y)In* in most parts

of western Anatolia, including the western Black Sea coast, Konya and Niğde, while it is *-(y)Im* in the central Anatolian dialects further to the east. In the easternmost and southeastern dialects, it has the same form as in Azerbaijanian, *-(y)Am*. In the west, with about the same distribution as *-(y)Im*, the first-person plural copula suffix is *-(y)Iz*, as in Standard Turkish, and *-z*, when added to the present in *-(I)yo*. In central and east Anatolia including the eastern Black Sea coast, it is *-(y)Ik* or *-k* in front vowel environments and usually *-(y)Iχ* or *-χ* in back vowel environments. Examples: west Anatolia *Gideriz*, *Yapıyoz*, east Anatolia *Giderik*, *Gideruk* (Rize, Trabzon), *Aliyuχ* (Erzincan), *Aliyiχ* (Diyarbakır). In the second-person singular, *-sIη* is the usual suffix in most parts of western and central Anatolia, while *-sIn* is found in the northeast, in Trabzon mostly as *-sun*. The east and southeast dialects – by and large the ones with *-(y)Am* in the first-person singular – have *-sAn*. However, with the present tense stem, most parts of west and central Anatolia use *-η*, e.g. *Gidiyoη* ‘You are going’ vs. ST *Gidiyorsun*, while the present tense suffix *-(y)i* found further east requires *-siη* or *-η*, e.g. Elazığ *Gidiysiη*, Gaziantep *Gidiyη*. Among the more eastern dialects, Erzincan has *Gidiysèn*. The Black Sea dialects show great variation, e.g. Giresun *Gidiy(i)sin* and *Gidiy(u)sun*.

Syntax

Compared with Standard Turkish, most dialects show a tendency to prefer less complex sentences. Hypotactical constructions are not very common. Direct speech is much more frequently used than indirect speech expressed by means of subordinate nominal clauses.

Notably in the northeast, constituent order is freer than in Standard Turkish. Thus, the post-verbal position can be used for focused constituents, particularly adverbial phrases expressing goal or purpose such as infinitives in the dative, e.g. Trabzon *Ik’i tãnesi gidiyi řere etmek almaa* ‘Two of them go to town to buy bread’ vs. ST *İki tanesi řehre ekmek almaya gider*.

Another important characteristic of the Trabzon and partly Rize dialects is that anaphoric pronouns are used in cases where Standard Turkish would normally dispense with them. In the Trabzon dialect, they are placed after the verb, e.g. *gardařumi daniy misun? – daniyrum oni* ‘Do you know my brother?’ – ‘I do [literally ‘I know him’]’, where Standard Turkish would have *Kardeřimi tanıyor musun? Tanıyorum* (without the pronoun *onu* ‘him’).

References and Further Reading

- Adamović, M. (1985) *Konjugationsgeschichte der türkischen Sprache*, Leiden: Brill.
 Boeschoten, H. (1991) ‘Dialectology’, in H. Boeschoten and L. Verhoeven (eds) *Turkish Linguistics Today*, Leiden: Brill, 150–65.
 Caferođlu, A. (1959) ‘Die anatolischen und rumelischen Dialekte’, in J. Deny et al. (eds) *Philologiae Turcicae Fundamenta* 1, Wiesbaden: Steiner, 239–60.

- Demir, N. (1993) *Postverbien im Türkentürkischen*, Wiesbaden: Harrassowitz.
 Karahan, L. (1996) *Anadolu ağızlarının sınıflandırılması*, Ankara: Türk Dil Kurumu.
 Korkmaz, Z. (1956) *Güney-Batı Anadolu ağızları*, Ankara: Ankara Üniversitesi Dil Ve Tarih Coğrafya Fakültesi.
 ——— (1990) ‘Anatolian dialects’, in G. Hazai (ed.) *Handbuch der türkischen Sprachwissenschaft* 1, Budapest: Akadémiai Kiadó, 388–413.
 Kowalski, T. (1929–30) ‘Einige Probleme der osmanisch-türkischen Dialektforschung’, *Rocznik Orientalistyczny* 7: 264–80.
 ——— (1934) ‘Osmanisch-türkische Dialekte’, in *Enzyklopaedie des Islam* 4, Leiden: Brill, 991–1010.
 Kral, P. (1980) *De in Turkije gesproken Turkse dialecten met een overzicht van de verwantschapstermen in Turkije*, MA thesis, Leiden University.
 Mollova, M. (1962) ‘Les ga-dialectes Turcs dans les Balkans et leur rapport avec les autres langues Turkes’, *Linguistique Balkanique* 4: 107–30.
 Räsänen, M. (1932) ‘Chansons populaires Turques du Nord-Est de l’Anatolie’, *Studia Orientalia* 4: 2, Helsinki.

13 The Turkish Language Reform

Bernt Brendemoen

A Successful Reform

Among all the reforms promulgated by Kemal Atatürk with the aim of westernising the Republic of Turkey, the alphabet and language reforms are probably the ones that – when regarding the situation today nearly 70 years after the reform was initiated – may be characterised as the most successful. The reforms in fields such as religion and civil law are frequently boycotted, notably in the rural districts. In the field of language, however, even the most conservative people today, whether they like it or not, use a vocabulary strongly influenced by the reform movement. Even if conservative groups in their publications and official speeches deliberately use old-fashioned language, this language is still extremely remote from what was common in similar texts before the reform. The fact that Ottoman poetry today is so unintelligible to high-school students that it has to be taught as foreign language texts, also clearly shows that the process of replacing Arabic and Persian lexical and syntactical elements with Turkish ones has been successful. However, in the 55 years the language reform was in progress from 1928–1983, it was probably the most discussed, and, in the eyes of a large part of the population, the most unpopular of all the reforms in the Republic of Turkey.

In the modern Turkish of the 1990s, there still prevails, to some extent, a lexical dichotomy, which nowadays rather manifests itself as an unusually great lexical variation. In order to explain this dichotomy, which was an essential difficulty for any learner of Turkish between the 1940s and 1970s, it is necessary to give a brief survey of the background and development of the reform.

Although the inadequacy of the Arabic writing system for expressing the sounds of the Turkish language had been pointed out as early as in the so-called *tanzimat* (literally 'reorganisation') period in the middle of the nineteenth century, it was not until nationalist ideology developed in the first decades of the twentieth century that radical circles launched the idea of adopting the Latin alphabet. This idea started to interest Atatürk long before

the Turkish Republic was established, and must have seemed particularly attractive to him both because the adoption of the Latin alphabet would signify a step towards westernisation and away from traditional Islamic values, and because of the nationalisation of the language would be the result of the subsequent language reform. For already when the alphabet reform was promulgated in 1928, Atatürk was aware that a profound language reform would be inevitable, both due to the inadequacy of the new Latin-based system for writing words of Arabic origin, and because the educational reforms launched almost simultaneously abolished Arabic and Persian as school subjects. Thus, the door to the traditions of the east was irrevocably closed.

The Work of the 'Turkish Language Society'

Atatürk's aim was to 'liberate' the Turkish language from foreign elements, or rather from Arabic and Persian elements, which represented the old culture from which he wanted to rescue the country and language, and to replace them with pure Turkish elements. In order to achieve this goal, the *Türk Dili Tetkik Cemiyeti*, 'the Turkish Language Society', was established in 1932. From 1936 onwards, it was called *Türk Dil Kurumu*, abbreviated TDK. Its methods in finding *Öztürkçe* ('pure Turkish') substitutes for the old Arabic and Persian words were mainly the following:

- 1 Reviving words found in older, particularly Old Ottoman and Pre-Ottoman, texts, such as *görenek* for *an'ane* 'tradition'.
- 2 Suggesting the use of words found in Anatolian dialects, e.g. *gözü* for *ayna* 'mirror'.
- 3 Introducing words from other Turkic and Altaic languages, such as *ulus* 'nation', a Mongolian word probably borrowed from Old Turkic, where it had the form *uluš*.
- 4 Enlarging the semantic sphere of already existing Turkish words, e.g. *çevir-* 'to turn around', which was 'enriched' with the additional meaning 'to translate', and *neden* 'why', which was used as a noun and given the meaning 'reason, cause' as well as its original meaning.
- 5 Introducing calques and literal Turkish translations of foreign terms, e.g. *bakan* 'minister', a participle of the verb *bak-* 'to look', for *nazır*, whose Arabic original is a participle with the same meaning.
- 6 A more dubious procedure for producing new words was the formation of compounds with prefixes, a morphological class not existing in the Turkic languages, such as *öngör-* 'foresee', to replace *derpiş et-*.
- 7 One of the most contestable methods has been the derivation of new words by means of productive or unproductive suffixes of Turkic or other origin, such as *-sAl*, e.g. *dilsel* 'linguistic', to replace *lisanî*, or *-(I)nl*, e.g. *ayrıntı* 'detail', to replace *teferruat*.

The methods mentioned above were not employed with the same intensity throughout the whole period of the reform, but were subject to the varying politics of the TDK. The much less controversial task of replacing morphological and syntactic structures of Arabic or Persian origin, e.g. the Persian *izāfat* (*izafet*) constructions (see pp. 118, 175, 200–1, 219), by genuinely Turkish constructions mostly met with no strong reactions, since most of them already had their Turkish equivalents.

Much of the work of the TDK, notably in the earliest years of the organisation, was rather devoid of scholarly value. Both the fact that Turcology was a new philological discipline and the lack of scholarly expertise in the different commissions of the TDK led to a good deal of amateurism and randomness in its early work. It should, however, be remembered that the task given to the TDK was one unprecedented in history and that it must have been extremely difficult to find suitable words to replace abstract words of Arabic and Persian origin. In some cases, this necessarily led to fanciful creations based more on imagination and free association than on words existing in real languages or dialects.

The 'Sun-language Theory'

There are strong indications that Atatürk was not entirely happy about the lists of proposed replacements for the Arabic and Persian words published by the TDK in the first years of the reform. In 1935 and 1936, a complete retreat was made with the introduction of the so-called *Güneş-Dil Teorisi*, the 'Sun-language Theory', based on a draft that Atatürk had received from an Austrian Serb, Dr Hermann F. Kvergić. According to this theory of language development, Turkish was the mother of all languages. Thus it was no longer necessary to search for pure Turkish words to replace Arabic and Persian ones, since the ultimate origin of these words and languages was Turkish anyhow. Needless to say, few issues in the history of the Turkish Republic have been ridiculed so much, in particular by foreign Turcologists, as this theory. A topic fervently discussed is whether Atatürk had it introduced primarily as a measure to stop further language purification, or if he really believed in it.

A Language Revolution

After Atatürk's death in 1938, the reform movement continued as if the 'Sun-language Theory' had never been put forth, and became not only a reform, but a revolution. The lack of a scholarly approach in the TDK did not prevent people – except for the active opponents against the reform – from accepting the neologisms. Newspapers, radio and particularly – from the 1970s onwards – television played an essential role for the general acceptance of the neologisms. Politics was also important. With the gradual democratisation of the country from the late 1940s onwards, the language movement gradually

became a party issue. Left-wing parties were strongly in favour of a continued reform, while conservative – religious or more right-wing – parties were against it. One would believe that nationalist parties such as the *Millî Hareket Partisi*, 'the National Movement Party', with its Turanist roots would have supported a Turkisation of the language. This may have been true of certain periods and certain circles, but the use of neologisms, particularly in the 1970s, came to be identified almost exclusively with the political left.

One reason for this is the important role played by Bülent Ecevit, the leader of the left-wing *Cumhuriyetçi Halk Partisi*, 'the Republican People's Party', who attained immense popularity during the 1970s. His conscious use of neologisms in his speeches was no doubt the main reason for the common acceptance of words such as *olanak* 'possibility', *olasılık* 'probability', and *sapta-* 'establish, prove', none of which has the semantic transparency otherwise usually found in neologisms that were accepted.

Political support from the government in power was in fact no prerequisite for the TDK to go on with its purification work, since Atatürk had bequeathed half of his fortune to the society. Nevertheless, with the exception of the fervent activities it used to show in periods of left-wing governments, its main efforts after c.1965 were more focused on the rather insurmountable task of creating a Turkish terminology for different fields of science and school subjects. Less effort was spent on finding *Öztürkçe* equivalents to words of Arabic and Persian origin, but this did not exclude the desire and hope that such substitutions would be found some day. In the field of science, the TDK has been quite successful, publishing numerous special dictionaries in which more or less fanciful Turkish equivalents of European terminology were suggested. In many fields, notably in the social and technical sciences, the number of *Öztürkçe* suggestions commonly accepted today is quite high.

Later Developments

With the military coup on 12 September 1980, persons and organisations that had been particularly opposed to the language reform tried to convince the government that the TDK should be closed down. In the autumn of 1983, the TDK was merged with the 'Turkish Historical Society' and a society for Atatürk studies into an institution named *Atatürk Kültür, Dil ve Tarih Yüksek Kurumu*, the 'Atatürk High Council for Culture, Language and History', and the former president and board members were replaced by more conservative philologists. Since then, the society has kept a rather low profile, concentrating on lexicography and publication activities. However, since the beginning of 1994, the society's periodical has also contained modest lists of *Öztürkçe* words suggested as equivalent to words and expressions from European languages.

Today, a neologistic vs. a conservative vocabulary may still be used by a person, a newspaper or a journal to display a political or religious vs.

antireligious attachment. But even in the most conservative publications, the amount of *Öztürkçe* words is quite high, due to the fact that several generations have passed since the reform started. Thus, no younger journalists today are capable of writing in the old language. Numerous new words that provoked fervent discussions and mockery when they were launched – in the earliest reform period or between 1940 and 1965 – are fully acceptable today and preferred by most people to their old equivalents. Tendencies to ridicule the reform hardly ever occur. This is especially striking in studies by foreign Turcologists commenting on the reform. While older scholars often mock the reform and characterise certain neologisms as 'terrible', 'unintelligible' and 'monstrous', later authors are much more careful in their characterisations, seeing that exactly these neologisms have become accepted against all odds.

Although the multi-faceted vocabulary of today's Turkish may be used to colour the language politically or religiously on a lexical level, it also presents the conscious writer with immense possibilities of stylistic variation without any necessary political or religious implications. The fact that the reform has yielded a vocabulary where words of different origins may co-exist shows that the storms accompanying the first phases of the reform have subsided and that the aim of the reform has been achieved to a great extent. Even if the *Öztürkçe* equivalents have not replaced all the old words, at least words of Turkish origin constitute the bone marrow of the contemporary lexicon, with the old words playing a marginal role.

One of the issues most frequently discussed at the peak of the reform was the generation gap created by the fact that parents had problems with understanding their children. The stabilisation of the language in the 1980s and the fact that the older generation who did not understand the new words is gradually dying away has rendered this argument invalid. Many people in Turkey today do not seem bothered by the similar gap that was created between the readers of today and Ottoman literature. The fact that Ottoman Turkish is a different language that can only be read with the help of a dictionary, seems to be accepted. However, the language reform has also created another cultural gap which is almost never discussed, namely between Turkey and the other Turkic-speaking nations of the world. Until the collapse of the Soviet Union, cultural contacts between Turkey and those nations were extremely limited. Now that close contact on all levels has been established, the language barrier makes itself clearly felt. If there had been no language reform, the Turks of Turkey would now have a large common vocabulary of Arabic and Persian origin with which they could make themselves understood with most of their sister nations. In the present situation, the possibilities of communication are rather restricted.

There are now circles in Turkey working for the creation of a common Turkic written language. It will be interesting to see to what extent Turkey in the next years will launch Turkish neologisms as replacements for words of Arabic and Persian origin in other Turkic languages.

Further Reading

- Boeschoten, H. (1991) 'The "language reform"', in H. Boeschoten and L. Verhoeven (eds) *Turkish Linguistics Today*, Leiden: Brill, 165–75.
- Brendemoen, B. (1990) 'The Turkish language reform and language policy in Turkey', in G. Hazai (ed.) *Handbuch der türkischen Sprachwissenschaft I*, Budapest: Akadémiai Kiadó, 454–93.
- Korkmaz, Z. (1974) *Cumhuriyet döneminde Türk dili*, Ankara: Ankara Üniversitesi.
- Lewis, G. L. (1984) 'Atatürk's Language reform as an aspect of modernization in the Republic of Turkey', in J. M. Landau (ed.) *Atatürk and the Modernization of Turkey*, Leiden: Brill, 195–213.
- (1997) 'Turkish language reform: the episode of the Sun-Language Theory', *Turkic Languages* 1, 25–40.
- Scharlipp, W. (1978) *Untersuchungen zur Morphologie und Substitution türkei-türkischer Neologismen*, Hamburg: Buske.
- Steuerwald, K. (1963–1966) *Untersuchungen zur türkischen Sprache der Gegenwart I–III*, Berlin: Langenscheidt.

14 Azerbaijanian

Claus Schönig

Introduction

Azerbaijani is very closely related to Turkish. Some scholars even take both to represent one language whose dialects cover a large area reaching from the Balkans to the Caucasus and into Iran. Modern Standard Azerbaijanian is based on the dialect of Baku, the capital of northern Azerbaijan, a dialect which itself is strongly influenced by Turkish.

The history of Azerbaijanian as a literary language is inseparably linked to the history of Ottoman Turkish. It is often possible to identify so-called old Anatolian Turkish texts as Old Azerbaijanian texts. Until the 1930s, written Azerbaijanian was very close to Ottoman, though it displayed more or less local features.

Azerbaijani was written in Arabic script up to 1929, when a Latin one was introduced. The end of the 1930s saw the change to a Cyrillic alphabet, which underwent numerous changes up to 1958, when its shape became final. This alphabet is based on the Russian alphabet and contains some additional signs. A new Latin-based modified script was adopted by the Republic of Azerbaijan in 1991.

Sound System

Phonetics and Phonology

Standard Azerbaijanian has the nine vowel phonemes /a/, /e/, /è/, /i/, /î/, /o/, /ö/, /u/ and /ü/. The symmetrical system of most modern Turkic languages, based on combinations of the features high vs. low, back vs. front, rounded vs. unrounded, is broken by the existence of a mid-unrounded front vowel /è/ in first syllables. Besides /è/, the vowels /o/ and /ö/ do not occur in suffixes.

While standard Azerbaijanian generally does not exhibit long vowels, vowel length may appear in conservative pronunciation of Arabic and Persian loanwords. In Arabic borrowings, vowel length sometimes substitutes for a glottal sound, *hamza* or *ʿain*, of the originals. These phenomena are expressed by an apostrophe in modern orthography, e.g. *me'na* 'meaning'.

As for the consonant phonemes of Standard Azerbaijanian, the stops include the labials /b/ and /p/, the dentals /d/ and /t/ and the velars /g/ and /k/.

The voiceless stops are aspirated. The front variant of /g/ is *g*, the back variants are *ğ* and *γ*. The front variant of /k/ is *k*, the back variant *χ*. A voiceless back stop *q* is lacking, except in dialects.

The other consonant phonemes are the voiceless affricate /ç/ and its voiced counterpart /j/, the sibilants /s/, /z/, /š/ and /ž/, the glide /y/, the nasals /n/ and /m/, the uvular fricative /h/, the liquids /r/ and /l/, and the labial fricatives /v/ and /f/. The liquid /l/ has, as in most Turkic languages, a front and a back variant. The labial fricatives /v/ and /f/ occur in loanwords and in some dialects instead of /b/ or /p/, mainly in syllable-final position, e.g. *-f* instead of *-p* in *öftüm* 'I kissed'. The consonant *ž* only occurs in loanwords.

Some typical correspondences between Azerbaijanian and Turkish consonantism may be mentioned. The nasal *m* corresponds to *b* in Turkish words containing a following nasal, e.g. *men* 'I' vs. Turkish *ben*. Azerbaijanian equivalents of Turkish words with *y*- sometimes lack this initial consonant, e.g. *il* 'year', *ürek* 'heart' vs. Turkish *yıl*, *yürek*. In dialects, /h/ sometimes appears as a 'cockney *h*' in words which otherwise begin with a vowel, e.g. *helbette* instead of *elbette* 'surely'. The fricative *-χ* often appears at the end of the first syllable in cases where Turkish exhibits *-q*, e.g. *çoχ* 'much' vs. Turkish *çok*.

Azerbaijani word accent is rather similar to the Turkish one. There is normally a pitch accent on the last syllable and often expiratory stress on the first one. Some suffixes, e.g. personal markers of pronominal origin and the negation suffix, move the accent to the syllable preceding them.

The melodious sentence intonation is influenced by Persian and rather different from the Turkish one. In spoken varieties, questions are often marked with rising intonation rather than by means of the interrogative particle.

Phonotactics and Morphophonology

The syllable structure in Azerbaijanian is similar to that of Turkish and other Turkic languages. A syllable can be open, e.g. *su* 'water', or closed by one consonant, e.g. *at* 'horse'. If it is closed by two consonants, the first one may not be a stop, at least in words of Turkic origin, e.g. *üst* 'above', *kürk* 'fur'. Loanwords may contain double initial consonants, e.g. *sport* 'sport', *Rza* 'Rizā (proper name)'. Initial *z*- appears in loanwords only. In the standard language, /r/ and /l/ can occur word-initially in loanwords without prothetic vowels.

There is at least partial devoicing of word-final consonants and suffix-initial consonants after voiceless consonants. This phenomenon is usually not reflected in the orthography, e.g. *өндү* 'kissed', since signs such as *т* and *к* are used rather to express aspirated fortis. Exceptions are found in some suffixes with initial *G* such as *-GIn*, e.g. *keskin* 'sharp'. Throughout this chapter, word forms will be quoted according to the conventions of the current orthography.

Before suffix-initial vowels, final lenis obstruents of polysyllabic and certain monosyllabic stems undergo morphophonemic shifts: $t \rightarrow d$, $k \rightarrow y$, g and $\chi \rightarrow \gamma$, $\check{c} \rightarrow \check{j}$, e.g. *söylet-* 'cause to speak' \rightarrow *Söyledir* '(S)he causes (somebody) to speak', *gét-* 'go away' \rightarrow *Gedir* '(S)he goes away', *çiçek* 'flower' \rightarrow *çiçeyim* 'my flower', *dodağ* 'lip' \rightarrow *dodayım* 'my lip', *çoğ* 'much' \rightarrow *çoğu* 'most of it', *güç* (written *güj*) 'force' \rightarrow *güjü* 'its force'. Most monosyllabic stems are not affected by this rule, since the obstruents in question are fortes (strong), e.g. *at* 'horse', *ata* 'to the horse' (see pp. 32, 98).

Certain bisyllabic stems lose the high vowel of the final syllable when a suffix with an initial vowel is added, e.g. *oğul* 'son' \rightarrow *oğlum* 'my son'.

Sound harmony is well developed, though some frequent suffixes are invariable, e.g. *-daş* 'fellow', *-ken* 'while'. Front vs. back harmony also affects older loanwords from Persian and Arabic. Though more recent loans from Russian are written according to Russian orthography, they may also, at least in lower sociolects, be pronounced according to the rules of front vs. back harmony. Labial harmony does not affect low vowels. Thus, in suffix syllables containing low vowels, front vs. back harmony unifies /a/ and /e/ into an archiphoneme /A/. In suffix syllables affected by both kinds of harmony, /i/, /i/, /u/ and /ü/ are unified into an archiphoneme /I/.

Morphology

Nouns and Adjectives

Word Formation

Suffixes deriving nouns from nominal stems include *-çi* for actor nouns, e.g. *yalançi* 'liar' \leftarrow *yalan* 'lie', *demirçi* 'blacksmith' \leftarrow *demir* 'iron', the diminutive suffixes *-çIK* and *-ça*, e.g. *uşağçığ* 'little child' \leftarrow *uşağ* 'child', *meydança* 'little square' \leftarrow *meydan* 'square', or *-daş* to denote a fellow participant, e.g. *emekdaş* 'colleague' \leftarrow *emek* 'profession'. The suffix *-LIK* is polyfunctional, forming abstracts, e.g. *yağşılığ* 'goodness' \leftarrow *yağşi* 'good', professions, e.g. *muellimlik* 'profession of a teacher' \leftarrow *muellim* 'teacher', and nouns of location and instrument, e.g. *kömürlük* 'coal cellar' \leftarrow *kömür* 'coal', *gözlük* 'glasses' \leftarrow *göz* 'eye'.

The most frequent suffixes deriving adjectives from nominal stems are *-li*, e.g. *atlı* 'provided with a horse, horseman' \leftarrow *at* 'horse', and the privative suffix *-siz*, e.g. *atsiz* 'not having a horse, horseless'. The suffix *-ki* derives relational adjectives such as *ağşamki* 'pertaining to the evening' \leftarrow *ağşam* 'evening', adjectives from locatives, e.g. *baydaki* 'located in the garden' \leftarrow *bayda* 'in the garden', and nominals from genitives, e.g. *bizimki* 'ours' \leftarrow *bizim* 'our'. Like possessive suffixes and demonstrative pronouns, *-ki* takes on the 'pronominal *n*' in oblique cases, e.g. *bizimki-n-den* 'from ours'.

A suffix *-LIK*, homophonous with the noun-deriving suffix *-LIK*, forms adjectives meaning 'good for ... concerning ...', e.g. *aylığ* 'for a month' \leftarrow *ay* 'month', *bizlik* 'concerning us' \leftarrow *biz* 'we'.

Of the numerous suffixes deriving nominal stems from verbal stems only a few can be mentioned here. The agentive noun is derived with *-(y)lji*, e.g. *aliji* 'one who buys, customer' \leftarrow *al-* 'take, buy'. The verbal noun in *-(y)lş* forms nouns mainly referring to the way of performance of an action, e.g. *giriş* 'entrance' \leftarrow *gir-* 'go in'. The verbal noun in *-mA* denotes the action as such and plays an important role in constructing non-finite clauses. It sometimes functions like a passive participle, e.g. *bölme* 'part' \leftarrow *böl-* 'divide'.

Comparison

A comparative degree of adjectives is mostly expressed by means of *daha* or the suffix *-rAK*, e.g. *daha büyük*, *büyükrek* 'bigger'. It can also be expressed without these elements if the standard of comparison is ablative-marked, e.g. *senden (daha) büyük* 'bigger than you'. Superlatives may be formed with the particles *en* and *lap*, e.g. *en güzel* 'nicest', *lap yağşı* 'best'. Detensive forms of adjectives are derived with *-(I)mtll*, e.g. *görmüzümtül* 'reddish' \leftarrow *görmüzi* 'red', *-sov*, e.g. *delisov* 'rather mad' \leftarrow *deli* 'mad', and the comparative suffix *-rAK*, e.g. *yağşırağ* 'rather good' \leftarrow *yağşı* 'good'. Intensive forms are formed with partial reduplication (see pp. 39–40), e.g. *tepteze* 'brand-new' \leftarrow *teze* 'new', *yamyaşıl* 'absolutely green' \leftarrow *yaşıl* 'green'.

Declension

Nominal inflectional suffixes are plural, possessive and case suffixes, attached to nominal stems in this order.

Like in Turkish, *su* 'water' has an oblique stem *suy-* before suffixes with initial vowel, e.g. genitive *suyun*, dative *suya*, accusative *suyu*, first-person possessive *suyum* 'my water', third-person possessive *suyu* 'its water'.

Plural of nouns is marked with *-lAr*, e.g. *év* 'house', *évler* 'houses'; *at* 'horse', *atlar* 'horses'. This marker is also used to express plurality in third-person possessive suffixes, pronouns and finite verbs.

Azerbaijani exhibits the Turkic standard set of six cases. Additional

Table 14.1 Case suffixes

		<i>év</i> 'house'	<i>oğ</i> 'arrow'	<i>ata</i> 'father'
nom.	Ø	<i>év</i>	<i>oğ</i>	<i>ata</i>
gen.	-(n)In	<i>évin</i>	<i>oğun</i>	<i>atanın</i>
acc.	-(n)I	<i>évi</i>	<i>oğu</i>	<i>atani</i>
dat.	-(y)A	<i>éve</i>	<i>oğa</i>	<i>ataya</i>
loc.	-dA	<i>évde</i>	<i>oğda</i>	<i>atada</i>
abl.	-dAn	<i>évden</i>	<i>oğdan</i>	<i>atadan</i>

Table 14.2 Possessive suffixes

		év 'house'	süd 'milk'	ata 'father'
1p.sg.	-(I)m	evim	südüm	atam
2p.sg.	-(I)n	evin	südün	atan
3p.sg.	-(s)I	evi	südü	atası
1p.pl.	-(I)mIz	evimiz	südümüz	atamız
2p.pl.	-(I)nIz	eviniz	südünüz	atanız
3p.pl.	-IArI	evleri	südleri	ataları

cases, not recognised as such in most grammars, include an instrumental in *-IA*, developed from the postposition *ile*, in dialects *-(I)nAn*, an equative in *-JA* and a terminal case in *-(y)AjAn*, in dialects *-(y)AtAn*, e.g. *bu vaxtajan* 'up to this time'.

The possessive suffixes are identical to those of Turkish. Possessive forms are declined like normal nouns, e.g. *atama* 'to my father', but the oblique forms of third-person suffixes take on a 'pronominal *n*', e.g. *ata-sın-a* 'to his father'.

Pronouns

The personal pronouns *men* 'I', *sen* 'thou', *o* 'he, she, it', *biz* 'we', *siz* 'you' and *onlar* 'they' are declined like nouns, though the oblique stem of *o* includes the 'pronominal *n*'. The first- and second-person singular dative forms are *mene* and *sene*, without the irregular back vowels of most other Turkic languages. If a pronoun precedes a noun as a genitive attribute, the noun is normally marked with the corresponding possessive suffix, e.g. *senin kitabın* 'your book'. The possessive suffixes may sometimes be omitted, mainly in the first person, e.g. *menim kitab(ım)* 'my book', *bizim teşkilat(ımız)* 'our organisation'.

The demonstrative pronouns *bu* 'this', *o* 'that' express nearness and distance. They have oblique stems with 'pronominal *n*', *bun-* and *on-*, to which plural and case suffixes are added, e.g. *bunlara* 'to these'. Though Azerbaijani has no definite article, *bu* and *o* are sometimes used in a similar way. The pronoun *hemin* 'the same' is borrowed from Persian.

Reflexive pronouns consists of combinations of *öz* 'self' and possessive suffixes, e.g. *özümüz* 'ourselves'. When the subject referent performs an activity related to him-, or herself, the pronouns may be reduplicated, e.g. *Men özüm-özümü hamıdan yaxşı tanışıram* 'Of all people I know myself best'. Interrogative pronouns are *kim* 'who', *ne* 'what' (oblique stem *ney-*), *hansı* 'which'. Indefinite pronouns are *kim ise* 'whoever' and *ne ise* 'whatever'. Quantifiers include *bütün* 'all, whole', *hamı* 'all', *bə'zi* 'some', *her, her bir, her jür* 'all, every', *filan* 'and the like, and so on, some', *eyni* 'same'. Negative pronouns, occurring with negative predicates, are formed with *həç*, e.g. *həç kim* 'nobody', *həç ne* 'nothing', *həç bir şey* 'nothing at all'.

Numerals

The numerals resemble those of Turkish. The cardinals of the first decade are *bir* 'one', *iki* 'two', *üç* 'three', *dörd* 'four', *bəş* 'five', *altı* 'six', *yəddi* 'seven', *sekkiz* 'eight', *doğguz* 'nine', *on* 'ten'. The tens are *iyirmi* 'twenty', *otuz* 'thirty', *qırq* 'forty', *elli* 'fifty', *altmış* 'sixty', *yətmış* 'seventy', *seksen* 'eighty', *doxsan* 'ninety'. 'Hundred' and 'thousand' are expressed with *yüz* and *min*, from which the designations of hundreds and thousands are derived in a multiplicative way, e.g. *dörd yüz* '400', *iki min* '2000'. As in most other Turkic languages, compound numerals are expressed in an additive way, starting with the highest value, e.g. *dörd min yəddi yüz on bir* '4711'.

Ordinal numerals are derived with *-(I)njl*, e.g. *ikinci* 'second', *üçüncü* 'third'. Instead of *birinci* 'first' the Arabic loanword *evvel* may be used. Distributive numerals are formed by means of repetition, e.g. *iki-iki* 'two by two'. Only *bir* takes on the distributive suffix *-Ar*, *birer-birer* 'one by one'. Collective numeral expressions can be formed with *-LIK*, e.g. *üçlük* 'trinity', or with plural possessive suffixes, e.g. *altımız* 'we six together'. Approximate numbers are expressed with juxtaposition, e.g. *bəş-altı güz* 'five or six girls'. The suffixes *-IArçA* and *-IArIA* are used to express an indefinite plurality of tens, hundreds or thousands, e.g. *yüzlerçe* ~ *yüzlerle güş* 'hundreds of birds'.

Copulas

The present copula is 1p.sg. *-(y)Am*, 2p.sg. *-sAn*, 3p.sg. *-dlr*, 1p.pl. *-(y)IK*, 2p.pl. *-sInIz*, 3p.pl. *-(dlr)IAr*. The past copula particle consists of the simple past of the defective verb **i-*. The copula particle *imış* signals indirectivity, e.g. inferentiality, but not anteriority. The conditional copula particle is *ise*. The word *dəyil* is used for negation, e.g. *dəyilik* 'we are not'. Existence is expressed by *var*, non-existence by *yox*.

Verbs

Derivation of Verbs

Verbs can be derived from nominal stems with suffixes such as *-IA*, *-IAn*, *-IAt* and *-IAş*, e.g. *terle-* 'sweat' ← *ter* 'sweat', *sözleş-* 'speak to each other' ← *söz* 'word, speech', *alovlan-* 'catch fire, flame up' ← *alov* 'flame'. The suffixes *-(A)l* and *-Ar* are often used for deadjectival derivation, denoting the acquisition of a quality, e.g. *daral-* 'become narrow' ← *dar* 'narrow', *göyer-* 'become green' ← *gök* 'green'. The suffix *-(I)ldA* is used for onomatopoeic derivations such as *dağilda-* 'ring' (of a bell).

Verbs can also be derived analytically from nouns and adjectives by means of the auxiliary verbs *əle-*, *ət-*, *göl-* 'do' and *ol-* 'be, become', e.g. *niyyet əle-* 'intend' ← *niyyet* 'intention'.

As for deverbal verb derivation, passive, reflexive and middle forms are marked with the suffixes *-(I)l* and *-(I)n*, e.g. *soy-* 'undress' → *soyun-* 'undress

oneself', *aç-* 'open (transitive)' → *açıl-* 'open oneself, get open', *yaz-* 'write' → *yazıl-* 'be written', *al-* 'take' → *alın-* 'be taken'. Due to a phonotactic constraint, *-(I)l* does not combine with stems ending in *-l*.

Cooperative-reciprocals are marked with *-(I)ş*, e.g. *yaz-* 'write' → *yazış-* 'write to one another', *kaç-* 'flee' → *kaçış-* 'flee together'.

Causatives are marked with several suffixes, some of which go back to combinations of simple causative suffixes: *-dlr*, *-t*, *-Ar(t)*, *-(I)r(t)*, *-Iz(dlr)*, etc., e.g. *al-* 'take' → *aldır-* 'cause to take', *yat-* 'lie down, lie, sleep' → *yatır(t)-* 'lie down', *yatızdır-* 'make sleep', *otur-* 'sit down, sit' → *oturt-* 'cause to sit (down)', *görç-* 'be afraid' → *görçüt-* 'frighten'.

Negation is marked with the suffix *-mA*, e.g. *gətmək* 'go', *gətməmək* 'not go'. Possibility of performing an action is expressed with *-(y)A bil-*, e.g. *Görə bildim* 'I could see', *Ala bilmemişəm* 'I could not take'.

Finite Verb Forms

A conjugated verb form normally consists of a verbal stem, a thematic suffix and a personal marker. There are two sets of personal markers, of possessive and pronominal origin. The former is used in the simple past (Table 14.4) and the conditional (Table 14.6), the latter in other paradigms (Table 14.3). The imperative lacks a common thematic suffix and has its own set of personal markers (Table 14.5).

A present tense is formed with *-(y)Ir*. The present in *-mAKdA* is described as a more focal present (see pp. 43, 114–15, 214). The aorist in *-(y)Ar*, the former present, signals disposition and may be interpreted in terms of habituality and future reference. The relations between these items are rather similar to those holding between their Turkish counterparts *-(I)yor*, *-mAKtA* and *-(V)r*. There is simple past in *-DI*; see Table 14.4. The finite item in *-mİş* differs from its Turkish counterpart in *-miş* by displaying more perfect-like functions. The variant *-(y)İb*, which is used in the second and third person, can take on the suffix *-dlr* in the third person, e.g. *Alıbdır* '(S)he has taken', *Bilmeyib(dır)lar* 'They have not learned'.

The prospective (future) in *-(y)AjAK*, often expressing other modal

Table 14.3 Present and aorist (*al-* 'take')

	Present		Aorist	
	affirmative	negative	affirmative	negative
1p.sg.	alırım	almıram	alarım	almarım
2p.sg.	alırsan	almırsan	alarsın	almazsın
3p.sg.	alır	almır	alar	almaz
1p.pl.	alırıg	almırıg	alarıg	almazıg
2p.pl.	alırıñız	almırıñız	alarıñız	almazıñız
3p.pl.	alirlar	almirlar	alarlar	almazlar

Table 14.4 Simple past (*at-* 'throw', *öl-* 'die')

	<i>at-</i> 'throw'	<i>öl-</i> 'die'
1p.sg.	atdım	öldüm
2p.sg.	atdın	öldün
3p.sg.	atdı	öldü
1p.pl.	atdıg	öldük
2p.pl.	atdıñız	öldüñüz
3p.pl.	atdılar	öldüler

nuances, may take on *-dlr* in the third person, e.g. *Alajağdır* '(S)he will, shall take'.

The imperative paradigm is heterogeneous; see Table 14.5. The second-person singular may assume suffixes expressing a higher degree of emotionality, e.g. *-kinann* ~ *-kinen-* ~ *-kunann* ~ *-künen*, *-kilan* ~ *-külan*, *-sAnA(n)*.

The optative paradigm is shown in Table 14.6. The first-person forms originally consist of *-(y)A* plus personal markers, but are contracted to *-(I)m* and *-(I)K*, e.g. *Alam* 'I will take' < **alayam*, *Alağ* 'We will take' < **alayag*.

A necessitative is formed with *-mAlI*. Its personal endings in the first person are *-yAm* and *-yIK*, e.g. *Almalıyam* 'I must (should, ought to) take', *Almalıyig* 'We must (should, ought to) take'.

A conditional is formed with *-sA*; see Table 14.7. Additional finite items

Table 14.5 Imperative

	<i>böl-</i> 'divide'	<i>başla-</i> 'begin'
1p.sg.	bölüm	başlayım
2p.sg.	böl	başla
3p.sg.	bölsün	başlasın
1p.pl.	bölek	başlayağ
2p.pl.	bölün	başlayın
3p.pl.	bölsünler	başlasınlar

Table 14.6 Optative

	<i>al-</i> 'take'	<i>gör-</i> 'see'
1p.sg.	alam	görem
2p.sg.	alasan	göresen
3p.sg.	ala	göre
1p.pl.	alağ	görek
2p.pl.	alasıñız	göresinüz
3p.pl.	alalar	göreler

Table 14.7 Conditional

	<i>at-</i> 'throw'	<i>öl-</i> 'die'
1p.sg.	atsam	ölsem
2p.sg.	atsan	ölsen
3p.sg.	atsa	ölse
1p.pl.	atsağ	ölsek
2p.pl.	atsanız	ölseniz
3p.pl.	atsalar	ölseler

may be formed with the copula particle *idi*. Imperfects in *-(y)Ir idi*, *-mAKdA idi* and *-(y)Ar idi* are formed from the present and aorist stems. A pluperfect is formed with *-mlš + idi*. The prospective and the optative in *-(y)A* combine with *idi* to express obligation in the past, counterfactual wishes, etc., e.g. *Almayajağ idim* 'I should not take', *Ala idim* 'If I only took, had I taken!' The conditional stem + *idi* expresses counterfactual conditions. The copula particle can be suffixed, the initial *i* being dropped after consonants and transformed into *y* after vowels, e.g. *-(y)Ir idi > -(y)IrdI*, *-mlš idi > -mlšdI*, *-(y)A idi > -(y)Aydl*.

Combinations with *imiš* are always accompanied by indirective nuances, e.g. *Atam onu yaxşı tanıyar imiš* 'My father apparently knew him well'. Certain thematic stems also combine with the dynamic copula verb *ol-* 'become', e.g. the necessitative in *Biz çıxmalı olduğ* 'We had to go out'.

Non-finite Verb Forms

The verbal nouns include an infinitive in *-mAK*, e.g. *almag* 'take, taking', *ölmek* 'die, dying'. It can take on plural, possessive and case suffixes, and combine with postpositions.

Participles are formed mainly with the temporally rather indifferent suffixes *-(y)An* and *-DIK*, the post-terminal (anterior) *-mlš*, and the prospective-modal *-(y)AJAK*.

Some converbs are formed with case forms of participles, e.g. *-dIKJA* (equative), *-(y)AndA*, *-dIKdA* (locative). Some consist of verbal nouns + case marker or postposition, e.g. *-mAGA*, *-mAK üçün* '(in order) to'. Other converb suffixes are *-(I)nJA*, *-jAK* 'as soon as', *-(y)AlI* 'since', *-(y)A*, *-(y)AraK* '-ing', *-(y)Ib* 'and ...', *-mAdAn* 'without -ing', *-ken* 'while' < *iken*.

Adverbs

Adverbs can be derived from adjectives by means of suffixes such as *-JA*, *-dAn*, *-lIKdA*, *-lIKdAn*, *-lIKIA*, e.g. *yavaş* 'slow' → *yavaşJA*, *yavaşdan* 'slowly', *ayrı* 'separate' → *ayrılığda* 'separately', *téz* 'quick' → *tézlikle* 'fast'. Time adverbs may be formed with suffixes, e.g. *dün* 'night' → *dünen* 'yesterday', or by means of reduplication, e.g. *addim-addim* 'step by step'.

There are also adverbs of space and direction such *ireli* 'ahead', *içeri* 'inside', *aşağı* 'down'.

The adverbs *béle* and *éle* 'such' are derived from the demonstrative pronouns *bu* and *o*, respectively. Interrogative adverbs derived from *ne* 'what' include *néje*, *ne jür*, *ne sayag*, *ne teher* 'how', *néçe*, *ne geder* 'how much', *ne vaxt*, *ne zaman* 'when', *ne üçün*, *niye* 'why' etc. Other interrogative adverbs are *hara* 'to where', *har(a)da* 'where', *har(a)dan* 'from where', *haçan* 'when'.

Other Morphological Categories

Azerbaijani postpositions can be classified according to the cases they govern. Thus, *üçün* 'for, to', *ile* with the suffixed form *-lA* 'with', *kimi*, *tek(i)* 'like, as', *géder* 'as much as' etc., govern the nominative of nouns and the genitive of singular personal and demonstrative pronouns, e.g. *həykel kimi* 'like a monument', *senin ile* 'with you'. Postpositions governing the dative include *göre* 'for, with respect to', *garşı* 'against', *nisbeten* 'compared with', *doyru*, *sarı*, *teref* 'towards', *dair* 'about', *dek* 'until'. The postpositions *ğabağ*, *evvel* 'before', *sonra* 'after', *beri* 'since', *savayı*, *başğa*, *ğeyri* 'except, different from' govern the ablative.

Certain postpositions are based on nouns and normally contain possessive and case suffixes, e.g. *ğabayında* 'in front of' ← *ğabağ* 'front side'. Elements such as *hağ* or *bare* 'for, about' are used with the locative and often, but not always, with possessive suffixes, e.g. *Ğiz barede fikirləşdiyini hiss etdi* '(S)he felt that he thought of the girl'.

Azerbaijani has numerous conjunctions, mainly of Arabo-Persian origin, e.g. *ve* 'and', *hem* 'also, as well', *ya* 'or', *eger* 'if', *amma* 'but', *ne ... ne* 'neither ... nor'. Native conjunctions include *dA* (postposed) 'and, too', *yoxsa, olmaya?* 'or, is it not, could it be?'

There are numerous particles such as *beli*, *he* 'yes', *açı(r)* 'you see'. The particle *mI* is used to mark sentence questions; *meger* 'really' indicates that a negative answer is expected.

Syntax

Adjectives can be used as preposed attributes without any morphological marking, *gözel qız* 'beautiful girl'. Nouns normally cannot be used this way, except if they refer to materials, e.g. *gümüš saat* 'silver watch'.

As for verbal phrases, there are numerous compositions such as *heyat sür-* 'live' ← *heyat* 'life' + *sür-* 'draw', *ad gøy-* 'give a name, to call' ← *ad* 'name' + *gøy-* 'put', *başa düš-* 'understand' ← *baş* 'head' + dative + *düş-* 'come down'. Many of these constructions are copies of Persian ones.

Two verbs may be combined by means of the converb in *-(y)Ib* to form a new complex lexical item, e.g. *yazıb yarat-* 'write (as an author)' ← *yaz-* 'write' + *yarat-* 'create'. Actional modifications may be expressed in the same way (see p. 42). The semantically modified verb appears as a converb in

-(y)İb, -(y)A or -(y)ArAK, followed by the modifying auxiliary verb.

Adverbials of time may be formed with nouns combined with case markers and other suffixes, sometimes also with unmarked nouns such as *sabah* 'morning, in the morning'. Adverbials of place and direction can be formed with locative, dative and ablative suffixes. Words expressing measure such as *az* 'little' and *çox* 'much' modify nouns, adjectives and verbs.

The Sentence

Constituent Order

The constituent order in sentences resembles that of Turkish, the unmarked order being subject, object, predicate. Place and time adverbials are often placed sentence-initially, before the subject. The basic order may be changed according to rules of sentence perspective. Constituents may thus also be placed after the predicate, predominantly in spoken varieties.

Non-finite Clauses

The participles in *-DİK*, *-mİş*, *-(y)An* and *-(y)AjAK*, the verbal nouns in *-(y)İŞ* and *-mAK*, and a variety of converbs play important roles in forming non-finite clauses.

For the construction of complement clauses, the verbal noun in *-mA* plays an important role. To form relative clauses, participles are used as attributes of nouns, e.g. *uçan güş* 'the bird that flies', *açılmış gül* 'the rose that has opened'. Such clauses can, like other non-finite clauses, contain a subject and other complements.

If its first actant, whether realised as a subject or not, is coreferential with the head, the participles in *-(y)An*, *-mİş* and *-(y)AjAK* are used, e.g. *uçan güş* 'the bird that flies, flew', etc. If it is not coreferential with the head, mainly *-DİK* and *-(y)AjAK* are used, e.g. *oxuduğum kitab* 'the book I read'. However, *-(y)An* can be used if the head has categorial reference, e.g. *Eli bişiren pilovu həç kes bişire bilmez* 'Nobody can cook the kind of pilau Ali cooks', *su olan yər* 'a place where there is water'. *-DİK* and *-(y)AjAK* take possessive suffixes marking the first actant of the relative clause. The subject may appear in the genitive, sometimes in the nominative, e.g. *atamın yazdığı mektub* 'the letter my father wrote', *men yazacağım mektub* 'the letter I shall write'.

Converbs are used to form adverbial clauses. Thus, the converbs in *-(y)AndA*, *-dİkDA* and *-(y)AjAKdA* express temporal and sometimes conditional relations, e.g. *Müharibe başladığında herbi siyasi kursa aldılar* 'When the war started they took him to military-politics lessons', *Çoban yox olanda goyunu gurd yeyer* 'When (if) the shepherd is absent, the wolf eats the sheep'. Items such as *-mİşdAn evvel* or *-mİşdAn gadağ* express 'before having done', e.g. *Men şairle tanış olmamışdan evvel şər yazmaqla meşgul olurdu* 'Before I had met any poets, I was busy writing poems'. *-(I)nJA* and *-jAK* express 'as soon as' or 'if only', e.g. *Onu görünje hamisi yadından çıxıb* 'As soon as (s)he saw him/her, (s)he forgot it all', *Onun adını eşitjek gülümsedi*

'(S)he smiled if (s)he only heard his/her name'. *-(y)All* expresses 'since', e.g. *Bu mektub alalı üç gün olub* 'Since I got this letter three days have passed'. The converb in *-ken* signals 'when, while, during', e.g. *Biz bayıra çıxmışken yağış yaymaya başladı* 'When we had got out, it began to rain', and the one in *-mAdAn* expresses 'without ... ing', e.g. *Düşünmeden söyleme* 'Do not speak without thinking'. The form *-mAmİş* is sometimes treated as a converb suffix meaning 'not yet', but combinations such as *Yay girmemiş, havalar istileşir* 'The summer has not come yet, (but) the air is getting warm' may better be analysed as coordinations.

Verbal nouns marked with the dative or with postpositions such as *ötrü*, *üçün* and *dəyə* may express purpose, e.g. *Oylumu Bakıya oxumaya göndermişəm* 'I have sent my son to Baku to study', *Meşud atasının yanına görüşə kətdi* 'Mas'ud went to his father to welcome him', *Trest müdiri danışmaq üçün Gülşene söz verdi* 'The trust manager called upon Gülshen to speak'. Constructions with *ötrü*, *üçün*, *dəyə* and *köre* also express causal meanings, e.g. *Bir bax, şəhere çatırlar dəyə uşaqlar ne sevinirler* 'Look how happy the children are because [literally: 'saying'] they are going to town'.

The intraterminal converb in *-(y)ArAK* is used for introspection into the course of an event. The interpretation of how it modifies the head clause depends on the context, e.g. temporally as in *Kerim ayağ üstə duraraq yerindən javab verdi* 'Karim answered from his place, (while) standing up', *O mene baxaraq gülümsədi* 'Looking at me, (s)he smiled', or causally as in *Ayaz gərçərəg gəri çekildi* 'Ayaz withdrew out of fear [literally 'fearing']'. The converb in *-(y)A* is only used in verbal compositions and in reduplicated form, e.g. *öyrene-öyrene* 'permanently learning'. Though the converb in *-(y)İb* is syntactically subordinative, it connects clauses in a semantically rather open way similar to the use of English 'and', e.g. *Gelib söyledi* '(S)he came and said it', *Hesen çay içib çantani alıb bayıra çıxdı* 'Hasan drank his tea, took the bag and went out'.

The Use of the Particle ki

The particle *ki*, copied from Persian, plays an important role in Azerbaijani syntax. It often precedes clauses with consecutive, temporal, local, and purposive readings, e.g. *Çöl ələ garanlığ idi ki, bir addım irelini seçmək mümkün deyildi* 'The steppe was so dark that it was impossible to discern anything one step ahead', *Henife sözünü gurtarmamışdı ki, bayırdan maşın sesi geldi* 'Hanifa had just finished the sentence, when the sound of a car came from outside', *Bütün günü, gəjeləri gezdi ki, özünə bir iş tapsın* 'Day and night (s)he walked around to find her-/himself a job'. The optative is often used as a kind of subjunctive mood. The function of *ki* may also be specified by combinations with other elements, e.g. *onda ki* 'when', *harada ki* 'where', *bəle ki* 'so that'.

The clause following *ki* may also specify a complement of the preceding clause, in which case the complement is often expressed by an anticipating

pronoun, e.g. *Xoşuma gelen odur ki, çox tevazökdür* 'What I like is that (s)he is very modest', *Men buna inanmıram ki, bu vaxtadək meni aldatmışlar* 'I do not believe they have cheated me so far'.

Relative clauses preceded by *ki* are very common. They can follow directly upon the nominal phrase they determine, e.g. *Bu xasiyyət ki, mənim oylumda var, hara gətse gərləyəcəg* 'This characteristic [which is present] in my son will damage him wherever he goes'. They can also be placed after the predicate, e.g. *Adam var ki, atın sayı ilə solunu tanımir* 'There are people who cannot distinguish the right and the left side of a horse'.

Lexicon

Azerbaijani exhibits numerous Persian loanwords, particularly in varieties spoken in Iran. The Arabic lexical layer, denoting phenomena of the Islamic cultural sphere, has mostly entered Azerbaijani via Persian and not directly from Arabic. Extensive borrowing from Russian began in the nineteenth century, as west European loanwords began to invade Ottoman Turkish. There has not been any radical Azerbaijani language reform comparable to the Turkish one (see Chapter 11). A good deal of the *Öztürkçe* (pure Turkish) words of Turkish are not intelligible to native speakers of Azerbaijani.

Dialects

The classification of Azerbaijani dialects is still under discussion. Spoken Azerbaijani may be divided into three main groups: (a) northern Azerbaijani, spoken in the Republic of Azerbaijan, (b) southern Azerbaijani, spoken in northwest Iran, and (c) east Anatolian dialects of Turkey. For Kashkay and Aynallu, spoken in southwest Iran, see Chapter 16.

Further Reading

- Azizbekov, X. A. (1965) *Azerbajdžansko-russkij slovar*, Baku: Azerbajdžanskoe gosudarstvennoe izdatel'stvo.
- Doerfer, G. (1959) 'Das Aserbaidŝanische', in J. Deny *et al.* (eds) *Philologiae Turcicae Fundamenta* 1, Wiesbaden: Steiner, 280–307.
- Gadžieva, N. Z. (1966) 'Azerbajdžanskij jazyk', in *Jazyki narodov SSSR 2. Tjurkskie jazyki*, Moscow: Nauka, 66–90.
- Şiraliyev, M. S. and E. V. Sevortjan (eds) (1971) *Grammatika azerbajdžanskogo jazyka*, Baku: Elm.

15 Turkmen

Claus Schönig

Introduction

Turkmen belongs to the Oghuz branch of Turkic. The beginnings of a Turkmen literary language written in Arabic script are observed from the eighteenth century onwards, when elements of spoken language began to enter the written language of the courts. The development into a modern Turkmen literary language started in the 1920s under Soviet rule.

The Arabic script was in use up to 1929, when it was replaced by a Latin alphabet. The latter was replaced in 1940 by a Cyrillic script based on the Russian one. In the early 1990s, the new Turkmen republic decided on the transition to a Latin alphabet, the final version of which was adopted in 1993.

Sound System

Phonetics and Phonology

Turkmen exhibits the short vowel phonemes /a/, /e/, /i/, /ı/, /o/, /ö/, /u/, /ü/ and the long vowel phonemes /ā/, /ā̄/, /ī/, /ī̄/, /ō/, /ō̄/, /ū/, /ū̄/. A short *æ* variant occurs in a few words, and a long variant *ē* occurs as a contraction product in the words *gēr* 'comes' < *geler* and *bēr* 'gives' < *berer*. The vowel length in Turkmen corresponds to Old Turkic vowel length in a rather consistent way. Of the high long vowels, only *ū̄* in words of Turkic origin is expressed in the Cyrillic orthography, namely by the combination *Үӱ üy*, e.g. *сүӱт θū̄t* 'milk', but *мүдир müdür* < Arabic 'director'. The pronunciation of high long vowels tends to be diphthongoid.

The consonant phonemes of native Turkic words include the labial stops /b/ and /p/, the dental stops /d/ and /t/ and the velar stops /g/ and /k/. In back environments, /g/ is realised as a back stop *ğ* or as a back fricative *ɣ*. In front environments, /g/ is realised as *g*, but may also have a fricative front variant *ɣ̟*. In back environments, /k/ is realised as back-velar *q*, in front environments as mid-velar *k*. Of the affricates, only /tʃ/ and /dʒ/ occur in words of Turkic origin, whereas the dental affricate /tʃ/ only occurs in Russian loanwords. The sibilants /s/ and /z/ have an interdental pronunciation *θ* and *ð*, whereas /ʃ/ and /ʒ/ have preserved their normal Turkic pronunciation. Other fricative consonants are /h/, /x/, /y/ and the bilabial glide /w/. The consonants *f* and *v* are

not common to all spoken varieties of Turkmen. Of the liquids /r/ and /l/, the latter has a front variant *l* and a back variant *ɫ*. The nasals are /m/, /n/ and /ŋ/.

In the present survey, the difference between *l* and *ɫ* and assimilations not reflected in the official orthography will not be indicated in the transcription. The same goes, as a rule, for vowel length. The signs *a*, *e*, *ī*, *i*, *o*, *ö*, *u*, *ü* will represent both short and long vowels, and *æ* will stand for long and short low ones [æ].

The word accent of Turkmen is rather similar to that of Turkish. There is normally a pitch accent on the last syllable and often an expiratory accent on the first syllable. With some suffixes, e.g. the personal endings of the pronominal type and the negation suffix, the pitch moves to the preceding syllable.

Phonotactics

The consonants *z*, *ž*, *c*, *r*, *l*, *f*, *v*, *n* and *x* do not occur word-initially except in loanwords. In spoken language, *f* is often replaced by *p*. The consonants *b*, *d*, *ž* and *ʃ* are devoiced word-finally. The consonant *b* becomes *w* between two vowels or between a liquid and a vowel.

There are numerous consonant assimilations, but most of them are not reflected in the orthography, such as the devoicing of suffix-initial *b*, *d* and *g* after stem-final voiceless sounds, e.g. *Aşgabatda* 'in Ashgabad'.

Consonant clusters are allowed in word-final position if they consist of a sonorant and a voiceless obstruent, e.g. *θūŋk* 'bone', a voiced and a voiceless obstruent, e.g. *bağt* 'luck', or a voiceless fricative and a voiceless stop, e.g. *doθt* 'friend'. Clusters of a sonorant and a voiced obstruent are rare and restricted to loanwords, e.g. *qarð* 'debt'. Other clusters are dissolved by means of epenthetic vowels, e.g. *pikir* 'thought', which ultimately goes back to Arabic *fikr*. There are no word-initial consonant clusters. In copies of loanwords such clusters are dissolved by means of prothetic or epithetic vowels. These vowels are not reflected orthographically in recent loanwords from Russian, e.g. *клуб* 'club'.

Morphophonology

There are shifts of final lenes (see pp. 32, 98) in polysyllabic stems before suffixes beginning with a vowel, *p* → *b*, *t* → *d*, *k* → *ɣ*, *q* → *ɣ*, *č* → *ʃ*, e.g. *mekdep* 'school' → *mekdebi* 'his school', *ayač* 'tree' → *ayaji* 'his tree', *gelmek* 'come, coming' → *gelmeýi* 'his act of coming'. In monosyllabic stems, this change occurs after long vowels, e.g. *āt* 'name' → *ādi* 'its name'.

In the second syllable of many Turkic words, high vowels are dropped when suffixes with an initial vowel are added to them, e.g. *ayıð* 'mouth' → *ayđi* 'his mouth', *uruş* 'struggle' → *urşi* 'his struggle'.

Sound harmony is well developed. Front vs. back harmony is reflected by the Cyrillic orthography. Loanwords from Persian and Arabic often do not obey the rules of intersyllabic harmony, e.g. *meydan* 'square' < Persian. Recent Russian loans are normally written according to the original Russian

orthography, but are also pronounced according to the harmony rules, at least in lower sociolects.

Rounded vs. unrounded harmony is common in the spoken language and even includes suffixes with low vowels. In the Cyrillic orthography, it is only expressed in high vowels and not beyond the second syllable. We thus find graphic forms such as *гаравуллык* for *garawulluq* 'guard' and *огланлар* for *oylonnor* 'boys'. In word-final position, high suffix vowels of suffixes are always written as unrounded, e.g. *сүрүжү* *θürüjü* 'driver'.

Morphology

Nouns and Adjectives

Word Formation

The most important suffixes with which nouns are derived from nominal stems are the following. Actor nouns are formed with *-čI*, e.g. *awčI* 'hunter' ← *aw* 'hunting'. The polyfunctional suffix *-IJK* creates abstract nouns, e.g. *doθtluq* 'friendship' ← *doθt* 'friend', place nouns, e.g. *daşliq* 'stony place' ← *daş* 'stone', and collective nouns of numerals, e.g. *bæşlik* 'unit of five'. The combination *-čIIJK* denotes professions, e.g. *čopančiliq* 'herdsmanship' ← *čopan* 'herdsman', and *-dAş* expresses companionship, e.g. *tapaqdaş* 'classmate'. Diminutives are formed with *-čAčIK* and *-jIK*, e.g. *kitapčaçiq* 'booklet' ← *kitap* 'book', *köprüjik* 'little bridge' ← *köprü* 'bridge'. There are also other, less frequent diminutive suffixes.

The most frequent denominal suffixe-deriving adjectives are *-li*, e.g. *θowatli* 'with document' ← *θowat* 'document', and the privative suffix *-θIð*, e.g. *θuwθuð* 'waterless' ← *θuw* 'water'. *-KI* derives relational adjectives such as *ağşamqi* 'pertaining to the evening' ← *ağşam* 'evening', also from genitives and locatives, e.g. *Amaniŋqi* 'belonging to Aman' ← *Aman* + genitive, *ayayındaqi tufli* 'the slippers on his feet' ← *ayaq* + possessive + locative. Like possessive suffixes and demonstrative pronouns, *-KI* takes on a 'pronominal *n*' in oblique cases, e.g. ablative *ičkinden* 'from the inner one'. A suffix *-IJK*, homophonous with the noun-deriving suffix mentioned above, forms adjectives meaning 'intended or suitable for', e.g. *donluq mata* 'material for clothing'.

Nominal stems are derived from verbal stems by means of a set of suffixes of which only the most important ones can be cited here. An agent noun is derived with *-IjI* (after vowels *-yji*), e.g. *aliŋi* 'one who buys, customer' ← *al-* 'take, buy', *yađiŋi* 'writer' ← *yađ-* 'write'. The verbal noun in *-(y)Iş* denotes processes or results of actions, e.g. *yayış* 'rain(ing)'.

Comparison

The comparative of adjectives is formed with *-rAK*, e.g. *kičiræk* 'smaller' ← *kiči* 'small', *belendrek* 'higher' ← *belend* 'high'. Superlatives and elatives are

formed with particles such as *öræn*, *iň*, *igğæn*, *şar* and *çim*, e.g. *öræn edepli* 'very educated', *iň edepli* 'most educated'. Intensive forms are formed by full or partial reduplication, e.g. *belent-belent* 'very high', *göm-gök* 'completely green' ← *gök* 'green'. Detensive forms of adjectives are derived with the suffixes *-(I)mtll*, *-(I)mtlK*, *-jA(K)*, *-jIK*, e.g. *ayimtil* or *aqja* 'whitish' ← *aq* 'white', *θarimtiq* 'yellowish' ← *θari* 'yellow', *tædeje* 'rather new' ← *tæde* 'new'.

Declension

The plural of nouns is expressed with *-lar*, e.g. *kitaplar* 'books', *atlar* 'horses'. Turkmen exhibits the Turkic standard set of six cases shown in Table 15.1, and the possessive suffixes shown in Table 15.2. The dative form *-Ā* occurring after vowels is a result of fusion of the stem-final vowel with the suffix vowel (see p. xxi). The plural suffix is not used to express plurality of possessors in the third person. The oblique forms of the third-person suffixes take on a final 'pronominal *n*'. In other respects, possessive forms assume suffixes like normal consonant stems, e.g. *atama* 'to my father (dative)', *ataθini* 'his father (accusative)'.

Table 15.1 Case suffixes

	After vowels	After consonants
nom.	∅	∅
gen.	-nIn	-In
dat.	-(∅)Ā	-A
acc.	-nI	-I
loc.	-dA	-dA
abl.	-dAn	-dAn

Table 15.2 Possessive suffixes

1p.sg.	-(I)m
2p.sg.	-(I)ŋ
3p.	-(s)I
1p.pl.	-(I)mIδ
2p.pl.	-(I)ŋIδ

Pronouns

The personal pronouns *men* 'I', *θen* 'thou', *biδ* 'we', *θiδ* 'you' are declined like nouns. The only exceptions are the back vowels of the first- and second-person dative forms *maŋa* and *θaŋa*, an anomaly found in most Turkic languages. The demonstrative pronouns *ol* and *olar* are used in the third person.

The demonstrative pronouns *bu*, *şu*, *şol*, *ol*, *hol* express different degrees of distance to the denoted entity. Their oblique stems are *mun-*, *şun-*, *şon-*,

on- and *hon-*, to which the case suffixes are added. The reflexive pronouns are combinations of *öd* 'self, own' and possessive suffixes, e.g. *ödümüδ* 'ourselves'. The interrogative pronouns include *kim* 'who', *næme* 'what', *næhili*, *nætiyθli* 'which kind of', *hayθi*, *næçenji* 'which'. Indefinite pronouns are *bütün* 'all, whole', *hemme*, *æhli*, *bari* 'all', *birnæçe*, *birentek*, *ençe* 'some', *her*, *her bir*, *her næçæ* 'all, every', *her kim*, *her hayθi* 'everybody', *kimdir*, *kim bolθa* 'whoever' etc. Negative pronouns are formed with the word *hiç*, e.g. *hiç kim* 'nobody', *hiç næme* 'nothing'. They always occur with negative predicates.

Numerals

The cardinal numbers of the first decade are *bir* 'one', *iki* 'two', *üç* 'three', *dört* 'four', *bæş* 'five', *alti* 'six', *yedi* 'seven', *θekiδ* 'eight', *doquδ* 'nine', *on* 'ten'. The tens are *yigrimi* 'twenty', *otud* 'thirty', *qirq* 'forty', *elli* 'fifty', *altmiş* 'sixty', *yetmiş* 'seventy', *θegθen* 'eighty', *doğθan* 'ninety'. 'Hundred' and 'thousand' are expressed by *yüδ* and *müŋ*, from which the expressions for hundreds and thousands are derived in a multiplicative way, e.g. '400' *dört yüδ*, '2,000' *iki müŋ*.

Ordinal numbers are derived with *-(I)nji*, e.g. *ikinji* 'second', *üçünji* 'third'. There are no special suffixes for distributive or collective numerals. Instead, *bolup* may be used, or ablative and possessive suffixes may be added to the cardinal number, e.g. *üç bolup* 'as a group of three', *ikiden* 'two by two', *üçiniđ* 'you three (together)'. Approximative expressions are formed with juxtaposed numerals, e.g. *üç-dört* 'three or four'. The suffixes *-dIr* and *-lAn* may also be added to the numeral, e.g. *üçdür-dördür çaya* 'three or four children'. To express indefinite plurality of tens, hundreds and thousands, the suffix *-larçA* is used, e.g. *yüδlerçe θudent* 'hundreds of students'.

Copulas

The present copula is formed with *-dIrln*, *-(dIr)θIŋ*, *-dIr*, *-(dIr)(I)θ*, *-(dIr)θIŋIδ*, *-dIrlAr*, e.g. *Muyallim(dIr)θiŋiđ* 'You are teachers'. The past copula particle *-dl* is derived from Old Turkic *ærti* 'was' and takes on personal markers of the possessive type. To negate the present and past copulas the items *dæl* 'is not' and *dældi* 'was not' are used. Existence is expressed by *bar*, non-existence by *yoq*.

Indirectivity (see p. 45) is expressed by the copula particles *-mIş* and *eken*, e.g. *Işçimişin* 'You are obviously a worker', *Muyallim ekenθiŋ* 'I understand you are a teacher'.

Verbs

Deverbal Verb Derivation

Verbs may be derived synthetically from nominal stems with suffixes such as *-lA*, *-lAn*, *-lAt*, *-lAş*, *-dA*, *-rA*, *-θIrA*, e.g. *eyerle-* 'saddle' ← *eyer* 'saddle',

atlan- 'mount a horse' ← *at* 'horse', *bulutlaş-* 'become clouded' ← *bulut* 'cloud', *çayθira-* 'want to have tea' ← *çay* 'tea'. The suffix *-(A)l* is used to derive intransitive verbs from adjectives, e.g. *ğaral-* 'become black' ← *ğara* 'black', *gıθyal-* 'become short' ← *gıθya* 'short'. The suffixes *-(I)ldA* and *-(I)rdA* are mainly used for onomatopoeic derivations.

One type of analytic derivations consist of a noun or an adjective plus an auxiliary verb such as *et-* 'make', *bol-* 'become', *ber-* 'give', *çek-* 'draw', *gör-* 'see', *at-* 'throw, shoot' and *tut-* 'take, keep', e.g. *tayarliq gör-* 'prepare' ← *tayarliq* 'preparation', *θalam ber-* 'greet' ← *θalam* 'greeting'. Derivations with *bol-* and *edil-* (passive of *et-*) are often synonymous, e.g. *tamam bol-* = *tamam edil-* 'finish, fulfil'. Some derivations with *et-* and with the suffix *-lA* are synonymous, e.g. *harç et-* = *harçla-* 'spend'.

Deverbal Verb Derivation

Negation is marked with *-mA*, e.g. *al-* 'take', *alma-* 'not take'. Frequentativity of an action is marked with *-(A)lA*, *-GIlA*, *-KAlA*, *-(A)KlA*, *-mAIA*, *-jA*, *-jAKlA* and *-İşdIr*, e.g. *baθyila-* 'suppress' ← *baθ-* 'press'. Detensive forms are derived with *-(I)mjIrA*, *-(I)mθIrA* etc., e.g. *gülümθire-* 'smile' ← *gül-* 'laugh'.

Reflexive and middle forms are marked with *-(I)n*, e.g. *yuwun-* 'wash oneself' ← *yuw-* 'wash'. Passives are marked with *-Il* or (after stems ending in vowels or *-I*) *-(I)n*, e.g. *böl-* 'divide', *bölün-* 'be divided'. The cooperative-reciprocal voice is marked with *-(I)ş*. The causative voice is marked with several suffixes e.g. the highly productive *-dIr* and the variant *-t*, used after stems ending in vowels, *-l* and *-r*. Less frequent and unproductive variants are *-Ir*, *-lð*, *-Ar*, *-dAr*, *-It*, only in *aqıt-* 'let flow' ← 'flow' *aq-* 'flow', and *-keð*, only in *görkeð-* 'show' ← *gör-* 'see'.

One type of analytic derivation consists of combinations of two verbs. The first verb takes on the converb ending *-(I)p*, the suffix *-mAGA* (dative of the verbal noun in *-mAK*) or the same ending as the second one. The beginning of an action is expressed by the combinations *-mAGA başla-*, e.g. *ğaynamaya başla-* 'start to boil', *-(I)p uyra-*, e.g. *adat edip uyra-* 'start to get used to', *-mAGA dur-*, etc. Many combinations express modifications in the field of actionality. Continuous actions may be expressed by combinations with the verbs *dur-* 'stand', *otur-* 'sit', *yör-* 'move' or *gel-* 'come', e.g. *Diğer durar* '(S)he speaks continuously'. The verbs *bar-* 'go' and *git-* 'go away' only combine with the *-(I)p* converb, and *ber-* 'give' only with the *-A* converb. Combinations of *-(I)p* with *otur-*, *yör-* and *dur-* may specify the non-transformative actional character of the lexical verb. Transformativity is expressed by *-(I)p* plus the auxiliaries *bol-* 'be', *gutar-* 'finish', *goyber-* 'put, send', *git-* 'go away', *gal-* 'stay', *geç-* 'pass', *çiq-* 'go out', *goy-* 'put'.

Possibility to perform an action is expressed by *-(I)p bol-*, *-mAK bol-*, *-(I)p bil-*, *-θA bol-* (infrequent), e.g. *Oqap bilmerθiñ* 'You cannot read'.

Table 15.3 Present and aorist (*al-* 'take')

	Present		Aorist	
	affirmative	negative	affirmative	negative
1p.sg.	alyarın	almayarın	aların	almarım
2p.sg.	alyarsıñ	almayarsıñ	alarsıñ	almarsıñ
3p.sg.	alyar	almayar	alar	almað
1p.pl.	alyarıθ	almayarıθ	alarıθ	almarıθ
2p.pl.	alyarsıñıð	almayarsıñıð	alarsıñıð	almarsıñıð
3p.pl.	alyarlar	almayarlar	alarlar	almaðlar

Table 15.4 Simple past (*at-* 'throw' and *öl-* 'die')

1p.sg.	atdım	öldüm
2p.sg.	atdıñ	öldüñ
3p.sg.	atdı	öldü
1p.pl.	atdıq	öldük
2p.pl.	atdıñıð	öldüñüð
3p.pl.	atdılar	öldüler

Finite Verb Forms

A conjugated verb form consists of a verbal stem, a thematic marker and a personal marker. There are two sets of personal markers, one of possessive and one of pronominal origin. The possessive type is used in the simple past (Table 15.4) and the conditional, and the pronominal type elsewhere (Table 15.3). The optative-imperative has its own personal markers and no common thematic suffix (Table 15.5).

A present tense is formed with *-yAr*, negated *-mAyAr*, plus pronominal personal markers, e.g. *İşleyær* '(S)he works', *Yadmayar* 'Does not write'. The so-called aorist in *-Ar* (fusion with stem-final vowel: *-(θ)Ār*), negated *-mAð*, has predominantly modal meanings, e.g. *Geler* '(S)he comes, will come', *Oqar* '(S)he reads, will read', *Oqamad* '(S)he does not read, will not read'. A more focal present for events going on at the moment of speaking (see pp. 43, 115) is formed with the *-(I)p* converb plus the contracted aorists *dūr*, *otūr*, *yatūr* of the verbs *dur-* 'stand', *yat-* 'lie' and *otur-* 'sit'.

The simple past suffix *-dl* takes on personal markers of the possessive type. A negative 'present-past' is derived from the participle in *-An* + pronominal endings + *yoq* 'non-existing', e.g. *Yadamoq < yazanım yoq* 'I do not/did not write (at all)'.

The prospective (future) in *-jAK*, sometimes with modal nuances, is negated with *dæl*, e.g. *Men aytjaq dæl* 'I will not say it'. A conditional is formed with *-θA* plus possessive personal markers, e.g. *gelθek* 'if we come'.

The optative–imperative paradigm is as follows:

Table 15.5 Optative–imperative (*gel-* ‘come’)

1p.sg.	geleyin
2p.sg.	gel gelθin (urging) gelθene (polite)
3p.sg.	gelθin
1p.pl.	geleliň
2p.pl.	geliň gelθeňið-le(ň) (polite)
3p.pl.	gelθinler

Necessitative meanings are expressed with *-mAll*, which does not take on personal markers and is negated with *dæl* or with the privative suffix *-θið* instead of *-li*, e.g. *Gelmeli* ‘(S)he must come’, *Gelmeli dæl*, *Gelmeθið* ‘(S)he does not need to come’. Intention to carry out an action is expressed by *-mAKçi*, without personal markers, e.g. *Men yaðmaqçi* ‘I intend to write’.

Additional finite forms can be derived by means of the past copula *-di*. A pluperfect is formed with *-(I)pdI*, negated *-mAndI*. Imperfects are formed with *-yArdI* and *-ArdI*. The item *-jAKdi*, negated *-jAK dældi*, signals a wish or necessity to carry out an action, and is often used to signal irrealis. The conditional + past copula, which forms a counterfactual past, may have personal markers after both elements, e.g. 1p.sg. *-θAmdIm*.

Turkmen has several indirective finite forms. A constative past with subjective nuances is formed with *-(I)pdIr*, developed from *-(I)p + durur* ‘stands’, negated *-mAndIr*, taking on pronominal personal markers, e.g. *gidipdir* ‘has (apparently) gone, went (once)’. A presumptive present is formed with *-yAndIr*, e.g. *Bilyændir* ‘(S)he probably knows’. The indirective copula particle *-mİš*, suggesting second-hand information (‘reportedly’ etc.), combines with items such as *-(I)p*, *-An*, *-yAn*, *-mAll* and *-jAK*, e.g. *Gidipmišin* ‘They say it has gone’. The indirective particle *eken* implies ‘it turns out that’ (see p. 45).

Non-finite Verb Forms

An important verbal noun is the infinitive in *-mAK*, negated *-mAdIK*, e.g. *almaq* ‘take, taking’, *ölmek* ‘die, dying’. It may take declension suffixes and combine with postpositions. Turkmen has a post-terminal (past) participle in *-An* (fusion with stem-final vowel: *-(θ)Ān*), negated *-mAn*, *-mAdIK*, an intraterminal participle in *-yAn*, an aorist participle in *-Ar*, a future participle in *-jAK* and a necessitative participle in *-mAll*.

Participles referring to the first or second actant are formed with *-mA*, e.g. *göçme halq* ‘nomads’ ← *göç-* ‘nomadise’, *örme θaç* ‘plait’ ← *ör-* ‘weave, plait’.

There are numerous converbs such as the ones in *-(I)p*, negated *-mAn* or *-mAdĀn*, e.g. *alıp* ‘taking’, *gödläp* ‘looking’, *-(y)A*, and the abtemporal one in *-(y)All*. Many case-marked participles function as converbs and form adverbial clauses.

Adverbs

Adverbs may be derived from nominals by means of dative, locative and ablative suffixes or markers such as *-čA*, *-lAy(In)*, *-lAp*, *-(I)n*, e.g. *ruθča* ‘in Russian’ ← *ruθ* ‘Russian’, *wağtlayin* ‘for a while’ ← *wağt* ‘time’, *günläp* ‘for x day(s)’ ← *gün* ‘day’, e.g. *On bæš günläp gelmedi* ‘(S)he did not return for fifteen days’. A limited number of words only function adverbially, e.g. *ir* ‘early’, *gič* ‘late’, *tið*, *čalt*, *derrew* ‘quickly’, *indi* ‘now’, *hemişe* ‘always’, *eyyæm* ‘already’ etc. If they denote qualities, they also take on comparative, elative and superlative markers, e.g. *irräk* ‘earlier’, *öræn čalt*, *čalt-čalt* ‘very quickly’, *iň tið* ‘most quickly’. Interrogative adverbs include *ničik*, *neneň* ‘how’, *næče* ‘how much’, *hayθi*, *næčenji* ‘which’, *nire* ‘what place’, *hačan* ‘when’, *næme üčün* ‘why’ etc.

Other Morphological Categories

The postpositions *üčün* ‘for, to’, *bile(n)* ‘with’, *yali* ‘like, as’, *θebæpli*, *diyın* ‘because of’ and others are used with the nominative case of nouns, e.g. *Aman bile(n)* ‘with Aman’, *Aman üčün* ‘for Aman’ etc., whereas singular personal and demonstrative pronouns are in the genitive, e.g. *θeniň yali* ‘like you’, *şuniň bile(n)* ‘with this’. The postpositions *kibi*, *kimin*, *dey* and *dek* ‘like’ govern the nominative of both nouns and pronouns.

Postpositions governing the dative include *göræ* ‘according to’, *garanda* ‘in comparison with’, *baqan ~ baqa* ‘towards’, *θari* ‘towards’, *garşi* ‘against’, *čenli* ‘up to, until, as far as’, *derek* ‘instead of’, etc. The ablative is governed by postpositions such as *başya*, *gayri*, *ödye* ‘except’, *θoň* ‘after’, *bæri* ‘since’, *öň*, *öňinčæ*, *owal*, *odal* ‘before, in front of, prior to’, e.g. *θayät on ikiden θoň* ‘after twelve o’clock’.

Like other Turkic languages Turkmen has a set of postpositions based on auxiliary nouns designing correlations in space or time, e.g. *ič* ‘inside’, *aθt* ‘under part’, *ara* ‘interval’. They appear with possessive suffixes, stand mainly in the dative, locative or ablative, and require the nominative or genitive of preceding nouns and the genitive of preceding pronouns, e.g. *oba ičine* ‘into the village’, *meniň üθtümde* ‘about me’, *onuň baraθında ~ doyrubında ~ haqında* ‘about him’ or *öyleriň araθından* ‘from between the houses’. The alternative forms *haqda*, *barada* and *doyruda* occur with the nominative forms of pronouns as well.

The Turkmen conjunctions are mainly of Arabo-Persian origin, e.g. *we* ‘and’, *hem* ‘also’, *æmma* ‘but’, *eýer* ‘if’. The frequent postposed element *-dA* ‘and, too’ is of Turkic origin.

Among the numerous particles, *-mI* marks sentence questions, e.g.

Gelerθigmi? 'Will you come?' Many particles express the attitude of the speaker, e.g. *-A* (questions with a nuance of regret), *-IA* ('really'), *-KA* ('I wonder').

Syntax

Nominal Phrases

As in other Turkic languages adjectives mostly function as adverbials without any special marking, e.g. *yayşî* 'good, well', e.g. *yayşî kitap* 'the good book', *Yayşî oqayar* '(S)he reads well'. Nominal compounds are created by means of the third-person singular possessive suffix, e.g. *türkmen θöðlüýi* 'Turkmen dictionary' ← *türkmen* 'Turkmen', *θöðlük* 'dictionary' + third-person singular possessive suffix.

Another method of compounding is to combine two nouns of similar meaning, *çat-θowat* 'written things' ← *çat* 'letter' + *θowat* 'document', *aliş-beriş* 'trade' ← *aliş* 'taking, buying' + *beriş* 'giving, selling'.

Comparative meaning can be expressed without comparative markers in syntactic constructions of the standard type of comparison + ablative (+ particle *dA*) + adjective, e.g. *θenden-de güçli* 'stronger than you'.

The Sentence

Constituent Order

The constituent order is similar to that of Turkish. The unmarked order is subject–object–predicate. Adverbs of place and time are often placed sentence-initially. The unmarked order can be changed to modify the sentence perspective. Thus, other constituents may be placed after the predicate, particularly in spoken language.

Non-finite Clauses

Participles and verbal nouns such as the infinitive in *-mAK* play important roles in forming constituent clauses.

Participles are also used to form relative clauses, e.g. *giden adam* 'the man who went away', *bilyæn adam* 'the man who knows', *bolmað iş* 'a thing that will not, cannot happen', *uyrajaq poezd* 'station to be reached'. If the head of the relative clause is not coreferential with its subject, the latter can be referred to by a possessive suffix on the head, e.g. *barmali waytım* 'the time I must go'. It is thus not always clear whether or not the possessive suffix signals possession, e.g. *Artıq θüren yerlerine göð geðdirdi* 'Artık looked around over his fields ~ the fields he had ploughed'. Only the participle in *-dIK* can take personal markers, e.g. *iydiýiň nan* 'the bread you eat'. If the subject of the relative clause has no specific referent, unmarked constructions are possible, e.g. *işläp duran θayat* 'the hour one is working'. Relative

clauses can also occur without heads, e.g. *Oturanlar gülüşdiler* 'Those who were sitting laughed together'.

The converb in *-(I)p*, besides its use in verbal composition, is mainly used to connect clauses in a neutral way comparable to 'and', but also to modify the event expressed by the matrix clause, e.g. *Elleri titräp otirdi* '(S)he was sitting, his/her hands quivering'. Doubled *-(I)p* can express duration of the modifying action. The converb in *-(y)A* has only survived in double forms of identical or antonymic verbs such as in *dura-dura* 'standing all the time', *otura-tura* 'sitting down and standing up'. The converbs in *-(I)p* and *-(y)A* can even take on the comparative suffix *-rAK*, when used as adverbial modifiers, e.g. *Telpeýini baðaraq geyip, iyip başladi* '(S)he pulled his/her cap further over his/her face and started to eat'.

The abtemporal converb in *-(y)AlI*, e.g. *Olar gideli, üç gün geçdi* 'Three days have passed since they left', can also be used with the postposition *bæri*, e.g. *yað çiqali bæri* 'since summer began'.

Temporal clauses expressing simultaneity may be formed with participles in the locative, e.g. *gelemdede* 'when I came'. The construction *-mAK* + possessive suffix + *bile(n)* can be used in a similar sense, e.g. *obaniň içine girmeyimið bilen* 'as we entered the village'. Clauses expressing anteriority may be formed with participles + ablative + *θoŋ*, e.g. *gidenden θoŋ* 'after going'. Purpose clauses can be formed with infinitives in the dative case, e.g. *Aşyabada oqamaya geldim* 'I came to Ashghabad to study'. The aim of an action can also be expressed with *-mAK için*, *-Ar yali* or *-mAK ~ -mAðLIK niyeti bile(n)*, e.g. *içki duyguθini aňlatmaðliq niyeti bilen* 'with the aim not to show his inner feelings'. Causal clauses may be constructed with *-mAK ~ -AnLIK* + possessive suffix + *θebæpli ~ jæhtli*, e.g. *yaqın bolanliýi θebæpli* 'because (s)he was near'.

Lexicon

Besides old loanword layers common to most Turkic languages, Turkmen has numerous lexical elements of Arabic–Persian and Russian origin. The former denote phenomena belonging to the sphere of Islamic civilisation, whereas the Russian loanwords, introduced from the nineteenth century onwards, stand for phenomena of modern life.

Dialects

Spoken Turkmen may be divided into two main dialect groups. The dialects of Yomud, Teke, Göklenj, Salır, Sarıq and others are relatively close to the standard language. The dialects spoken in the valleys of the Kopet Dag mountains and on the Amu Daryā river are more distant from standard Turkmen.

Further Reading

- Azimov, P., Amansaryev, Dž. and Saryev, K. (1966) 'Turkmenskij jazyk', in *Jazyki narodov SSSR 2, Tjurkskie jazyki*, Moscow: Nauka, 91–111.
- Baskakov, N. A., Karryev, V. A. and Xamzaev, M. Ja. (eds) (1968) *Turkmensko-russkij slovar'*, Moscow: Sovetskaja Enciklopedija.
- Baskakov, N. A., Xamzaev M. Ja. and Čaryjarov, B. (1970) *Grammatika turkmenskogo jazyka 1, Fonetika i morfologija*, Ašxabad: Ylym.
- Clark, L. (1998) *Turkmen Manual*, Turcologica, Wiesbaden: Harrassowitz.
- Dulling, G. K. (1960) *An Introduction to the Turkmen Language*, Oxford: Central Asian Research Centre in Association with St Antony's College.

16 Turkic Languages of Iran

Gerhard Doerfer

Introduction

Iran was one of the last regions of the world to be explored linguistically. This is especially true of the languages of the Turkic peoples there. Only the language of Azerbaijan in the northwest and that of the Kashkay (Qaşqā'i) including the Aynallu region in the southwest were fairly well-known. However, even the Azerbaijani of Iran was only sparsely researched compared with the Azerbaijani spoken in the Soviet Union. Although English and Russian scholars had visited Khorasan, they had not recorded any Turkic texts or linguistic forms. The few recordings of Khalaj, spoken in central Iran, had either been noted in corrupted form, i.e. mixed with Oghuz forms, or disregarded.

According to research findings valid up to the 1960s, Azerbaijani was distributed over an area extending from the 44th degree of longitude up to the outskirts of Teheran and from the Soviet border in the north to a line connecting Hamadān and Qom, c.34 degrees north. In the southeast, the Kashkay–Aynallu region extended from about 50 to 56 degrees tapering off towards the southeast and reportedly northwards from 31 to 28 degrees north. Turkmens were reported in and beyond Khorasan from about 54 to 60 degrees. East and north of the 36th Parallel, Khalaj had not been accounted for at all.

These older findings will now be compared with more recent research, resulting from the Göttingen expeditions between 1968 and 1976.

Central and South Oghuz

The founder of Iranian Turcology, Karl Foy (1903–1904), was the first to investigate Iranian Azerbaijani more scientifically. He made a special study of the dialect of Tabriz, which enjoys wide recognition as a standard language in Iranian Azerbaijan, although it has never developed into an official written language. Foy also made some smaller investigations of the dialect of Urmia. Furthermore, he realised that the language spoken in Erzurum in eastern

Anatolia was not Turkish but Azerbaijani.

Subsequently, H. Ritter (1921, 1935) and H. S. Szapszal (1935) made important contributions to the studies of Iranian Azerbaijani. The eastern limits of this language seemed relatively well-established, but the Göttingen expedition of 1973 discovered additional dialects. First, the Galūgāh dialect, spoken on the southeastern edge of the Caspian Sea, was discovered. For some time this seemed to be the easternmost Azerbaijani dialect. Later, however, more focused research showed that there were Azerbaijani enclaves in Northern Khorasan as well, namely Dara-Gaz and Loṭf-ābād.

Of even greater importance were investigations leading into the southern regions. Not only was the Kashkay region itself studied in more detail from 1968 onwards, but also villages situated far to the north of the known area, e.g. Paradonba. This provided a completely new picture of the distribution of the Oghuz dialects of western Iran. Kashkay and related dialects will be referred to here as South Oghuz. Many transitional varieties between Central and South Oghuz appeared. The dialect of Zanjān can still be classified as Central Oghuz. The dialect of Sonqor and two neighbouring villages (c.35 degrees north, 48 degrees east) deviates more strongly. The dialects north of Khalajistan, Pougerd, Āšteyān (c.34 degrees north, 50 degrees east) are even closer to Kashkay and mostly correspond to the Afshar variety spoken in Kabul in Afghanistan. The dialects of Qazwīn, northeast of Teheran, and Solaymān-ābād, southwest of Hamadān, are structurally very close to Kashkay.

The examination of materials collected in Paradonba, Šahrak and Ali-Qūrčī has repositioned the area of pure South Oghuz to within about 70 kilometres of the transitional area of Pougerd. The following table of personal suffixes may illustrate the transitions. The first-person singular suffix is *-Am* throughout, which makes South Oghuz closer to Azerbaijani and the main body of Khorasan Turkic, but clearly separate from Turkish and Turkmen.

Khorasan Turkic

The discovery of the previously unknown Turkic dialects of Khorasan is an even more significant result of the expeditions.

A careful distinction must be drawn here. Turkmen dialects are only spoken in a region reaching from east of the Caspian Sea to about the 57th

Table 16.1 Personal suffixes

	<i>1p.sg.</i>	<i>2p.sg.</i>	<i>1p.pl.</i>	<i>2p.pl.</i>
Azerbaijani	-Am	-sAn	-IK/UK	-siz/sUz
Sonqor etc.	-Am	-sA	-Aχ	-sIz
Pougerd etc.	-Am	-Aŋ	-IK	-Iz
Qazwin etc.	-Am	-Aŋ	-AK	-Iz
Kashkay etc.	-Am	-Aŋ	-AK	-AŋIz

degree of longitude. Though they have hardly been researched, they clearly constitute extensions of the dialects of Turkmenistan, especially Yomud, Teke and Gökleŋ. In the past, Turcologists mistakenly designated North Khorasan as Turkmen. This gave Turkmen nationalists cause to claim Southern Turkmenistan, which extends about 200,000 square kilometres, from Bābol on the Caspian Sea, running along the 36th Parallel and including large cities such as Mašhad, an area containing roughly three times as many Iranians as Turks. This claim runs contrary to ethnic and historical facts.

The Turkic dialects spoken in Northern Khorasan north of the 37th Parallel, interspersed with Persian and Kurdish settlements, are not Turkmen, but Khorasan Turkic or East Oghuz dialects. They can roughly be classified by means of the following features:

- Present tense suffixes: (a) *-IyA*, (b) *-Ir*, (c) *-A*.
- Personal suffixes: (a) *-(A)m*, *-(A)ŋ* as in Kashkay (b) similar, but *l* and *2p.pl. -IK*, *-IGIs*, (c) *-mAn*, *-sAn* as in Oghuz Uzbek (*1p.pl. -mIz*), (d) *-mAn*, *-sAn*, but *1p.pl. -bIz*. There is thus a clear distinction from Turkmen, which has the personal suffixes *-In*, *-(s)Iŋ*, *-Is*, *-IK*, *-(s)IŋIz*.
- First-person forms of the imperative: (a) *-Im*, *-AK*, as in Azerbaijani, (b) *-Am*, *-AK* (same vowel in singular and plural), (c) *-Im*, *-IK* (same vowel in singular and plural), (d) *-Im*, *-Āyli* or similar special forms

Along these lines, six dialects can be distinguished: Northwest = Bojnūrd etc., North = Qūčān etc., Northeast = Gūjgī etc., South = Solṭān-ābād etc., Southeast = Xarw-e 'Olyā etc., and Langar. Khorasan Turkic is of special significance since it solves the riddle of the so-called 'olya bolya language', the previously undefinable character of certain Anatolian texts from the thirteenth and fourteenth centuries. These texts actually represent North and Northeast varieties of East Oghuz. The old mystery was not solved until the Göttingen expedition to Khorasan.

Turkmen, which is reminiscent of Oghuz Uzbek and has also been influenced from the east, displays lexical items such as *tašla* 'throw' and *köp* 'much, many', whereas genuine Khorasan Turkic has *atar-* and *čoχ* respectively.

Table 16.2 Khorasan Turkic dialect features

	<i>Present</i>	<i>Predicate</i>	<i>Imperative</i>	<i>Distribution of suffix variants</i>
NW	-iyA	-m	-Im, -AK	Aa Bb Cc
N	-A-	-mAn, -mIz	-Im, -AK	Ac Bc Ca
NE	-A-	-mAn, -bIz	-Im, -Ayll	Ac Bc Ca
Langar	-A-	-mAn, -bIz	-Im, -AK	Ac Bd Ca
S	-Ir-	-Am, -IK	-Am, -IK	Ab Ba Cb
SE	-Ir-	-Am, -IK	-Im, -IK	Ab Bb Cc

Khalaj

While all dialects listed above belong to the Oghuz branch of the Turkic family, the language of the central Iranian Khalaj strongly diverges from Oghuz. It is spoken 100 miles south of Teheran, between Qom and Arāk, covering an area of about 3000 square kilometres. Language recordings were made there in the course of four expeditions between 1968 and 1973. Seven dialects can be identified. The western dialect is so different from the others that some Khalaj do not acknowledge it as belonging to their language. We have examined Khalaj in 5 books and 36 articles. Research into this autonomous language has altered the classification of the Turkic languages.

Historical and Demographic Facts

The historical and demographic background poses certain difficulties. The history of many tribes, e.g. those whose varieties show transitional features, has not been sufficiently researched. In addition, virtually all (male) Turks in Iran are bilingual. It is generally safe to assume that someone who speaks Turkic as well as the predominant Persian of Iran is a Turk. This distinction is, however, not always easily made. A passer-by, whom I asked if he was a Turk or a Persian, replied, 'If you speak Turkic to me, I am a Turk, and if you speak Persian, I am a Persian'. Demographic information is also problematic because the Iranian population increases very rapidly, albeit differently by region. While according to official claims, all inhabitants of Iran are Iranians, the following distribution is more accurate: 41 per cent Persians, 29 per cent other Iranians, i.e. Kurds, Gilanians etc., 4 per cent Arabs, and 26 per cent Turks.

Azerbaijani constitutes the largest Turkic group, while South Oghuz, the transitional dialects of South Oghuz and Turkmen each comprise a few hundred thousand speakers. The Khorasan Turkic population may currently number 2 million. It is a typical characteristic of Iran that such a large population should remain undiscovered for so long, being subsumed under the name *torkī*. In contrast, the number of Khalaj is extremely small, amounting perhaps to roughly 28,000. The Khalaj are the politically most insignificant but linguistically the most important Turkic people of Iran.

The odds for survival seem promising for Azerbaijani, for the Kashkay portion of South Oghuz as well as for Turkmen. They are substantially worse for a large part of South Oghuz and for Khorasan Turkic. In all probability Khalaj will completely disappear during the first half of the twenty-first century. Along with progressing modernisation many local dialects, not only Turkic ones, are becoming extinct. Therefore last-minute linguistic recordings are of utmost scientific importance because increasing education and communication entail increasing Persianisation, as Persian is the only official language of Iran.

By far the largest group of Iranian Turks are descendants of those Oghuz who fled south with Seljuk (Selçük) in the tenth century in order to escape impending bondage and, in the eleventh century, founded an empire fanning out from the middle course of the Syr-Darya. From there, one group conquered first Khorasan and then Anatolia. Another group, however, stayed in Khorasan. Moreover, several Turkic tribes began pouring back into Azerbaijan from 1502 onwards. The Oghuz remaining in the old Turkmen region east of the Caspian Sea later conquered the eastern portion of present-day Turkmenistan. This has left an impact on the eastern Turkmen dialects.

The historical-linguistic situation can be summarised as follows: the Turkmens are descendants of those Oghuz who did not take part in the Seljuk migration and eventually conquered the eastern part of present-day Turkmenistan. The Khorasan Turks are the Turks who, after having conquered Khorasan, remained there (plus subsequent immigrants). The groups speaking Azerbaijani are descendants of those Oghuz who went further west, staying in Azerbaijan or returning there from Anatolia. Some of them again proceeded south (South Oghuz), and smaller numbers later went east (Galūgāh, North Khorasan).

The predecessors of the Khalaj, the Arghu, are first mentioned in a central Asian Manichaean source of 759-780. Between the eighth and eleventh centuries the Arghu lived in the vicinity of the Oghuz *Xalač*, from whom the designation Khalaj was adopted (like the French adopted the name of the Franks). From the ninth century onwards, Cisoxania and Khorasan became home to most of them. Probably fleeing from the 1219 Mongolian incursion, the Khalaj moved to central Iran, where they are first mentioned in 1370. Some of them settled in the region of the Kashkay, where they were subsequently assimilated.

Linguistic Features

Phonetically, morphologically and lexically, all Oghuz varieties demonstrate typical Oghuz characteristics, e.g. the sound changes *t* > (mostly) *d*-, *k* > (mostly) *g*-, loss of *-gl-γ*. Therefore one finds everywhere instances of Old Turkic *til* 'tongue' > *dil*, *kün* 'day' > *gün*, *sariγ* 'yellow' > *sari*. Khalaj is entirely different, having preserved the ancient state: *ti-l*, *kü-n*, *sāruγ*. The Oghuz sound shifts *t* > *d*- and *k* > *g*- show many exceptions. Some of them are the result of assimilation, e.g. usually *tāpā* 'hill', rarely *dāpā* etc., while others have semantic reasons: words which stand out semantically, such as *kim* 'who', or have a strong emotional effect, e.g. standing for repulsive things such as *kir* 'dirt', *toz* 'dust', prefer the voiceless sounds.

All Oghuz dialects show a number of common features or tendencies that contrast with Turkish. The accusative suffix after a vowel is generally *-ni* (as

opposed to Turkish *-(y)l*). This forms a transition to the neighbouring non-Oghuz Turkic languages, where *-nl* obtains after consonants as well.

Persian influence has had an especially profound effect. The old Turkic sound harmony has for the most part been lost. There is an intra-language struggle apparent here: partial maintaining of Old Turkic conditions, partial effects of Persian in a largely bilingual population. A few features may be mentioned:

- 1 Loss of sound harmony in suffixes, e.g. Azerbaijani *gäl-maç* 'come', mostly similar in South Oghuz and Khorasan Turkic, but not in Kashkay or Turkmen, which are hardly Persianised.
- 2 Occasional shifts *ö > e* and *ü > i* in South Oghuz (rarely in Kashkay), e.g. *gin* 'day', *ger-* 'see' = Azerbaijani and Turkmen *gün, gör-*. The shifts are also observed in northwest and southern varieties of Khorasan Turkic and the isolated Azerbaijani Galügāh dialect. Langar displays the shifts *ö > o* and *ü > u*.
- 3 The Persian influence on the vocabulary is stronger than in Turkish. This is especially true of the southern varieties of Khorasan Turkic, e.g. in cities north of Torbat-e Heydarīya. Here, 'navel' is not **kōpāk* or **kindik*, but *nāf*, as in Persian.

Table 16.3 illustrates the increasing divergence from Turkish. It contains (1) the post-vocalic accusative suffix, (2) the opposition *e : é (< *ē)*, (3) the development of **b-* in the words *bār* 'there is/are', *bar-* 'go', *bēr-* 'give', and *bōl-* 'become', which frequently occur in sandhi, (4) the personal suffixes, (5) the development of the Old Turkic long vowels.

Differences between Oghuz Dialects

Apart from the marked divergence from Turkish, there are also pronounced differences between the Oghuz dialects of Iran. Table 16.3 has already demonstrated the loss of the short vs. long opposition of the vowels which progresses from east to west and which is closely linked to the shift Proto-Turkic **ē > ē > é ~ e*.

As for the development of Old Turkic **bōl-*, most Oghuz sub-branches show *ol-*. Turkmen, in contrast, has *bol-*, only in sandhi *vol-* or *wol-*. Khorasan Turkic displays a typical transition feature: the northeast varieties and Langar have *bol-*, in sandhi *vol-*, while the southern and northern varieties differentiate between *ol-* 'become' and *bol-* 'become ready', 'mature', thus a separation on semantic grounds. Thus the shift *b- > v-*, which has taken place only gradually in Turkish, has not been completed further east.

Due to Persian influence, Turkic dialects of Iran have adopted *-lAn* as the plural suffix on the verb according to the colloquial Persian pattern *āma-dan* 'they came'. Exceptions are of the Azerbaijani dialects of Dara-Gaz, Loṭf-

Table 16.3 Divergence from Turkish

Turkish	Azerbaijani	South Oghuz	Khorasan Turkic	Turkmen
(1) <i>-yl</i>	<i>-nl</i> , rarely <i>-yl</i>	Like Azerbaijani	<i>-nl</i>	<i>-nl</i>
(2) <i>e</i>	<i>e : é</i>	Like Azerbaijani	Like Azerbaijani N/NW/Langar <i>e</i>	<i>e</i>
(3) <i>v-</i>	Like Turkish	Like Turkish	<i>b-</i> (e.g. <i>bar</i> , in sandhi <i>v-</i> , always <i>v-</i> in Čaram, Langar, Rūhābād Zourom)	<i>b-</i> (in sandhi <i>v-</i> , <i>-w-</i>)
(4) <i>-Im/-sIn</i> <i>-Iz/-sInIz</i>	<i>-Am/-aAn</i> <i>-UK, -sIz</i>	See Table 16.1	See Table 16.1	<i>-In/-(s)Iŋ</i> <i>-Iz-IK/-(s)Iglz</i>
(5) Short (*ē > e)	Short (but *ē > é)	Paradonba, Kabul like Azerbaijani, in others, long: <i>ō, õ, ū, ũ</i>	Like in South Oghuz	Long throughout

ābād and the dialect of Langar (strongly reminiscent of Oghuz Uzbek) where *-lAr* prevails.

Quite common throughout is the use of copies of the Persian comparative suffix *-tar* in place of native *-rAK*, e.g. in the Azerbaijani dialects of Dara-Gaz, Loṭf-ābād, South Oghuz Farhād-Xān, Qorwa and Paradonba as well as in the East Oghuz dialects of Čaram, Langar and Mārešk. This distribution is not surprising given the extreme isolation of many Turkic dialects and the overpowering influence of their Persian surroundings.

This Western influence is faced directly by a strong impact from the east, and this opposition is especially evident in Khorasan. Only in Turkmen and Azerbaijani enclaves of Khorasan is the word for 'big' **bādük*. Otherwise it is *kata*, which ultimately goes back to Indian *kattā*. The word for 'find' is **tap-* almost everywhere, in Azerbaijani, South Oghuz and Turkmen. Only some areas of the Southern dialects of Khorasan Turkic have *bul-* (like Khalaj, Turkish and Yakut). While 'much, many' predominates in the genuinely Oghuz form *čoç*, it is *köp* in Northeast Khorasan Turkic, Langar and Čaram (as in Turkmen, Oghuz Uzbek and Uzbek).

In three places in Southeast Khorasan Turkic we find the Uzbek or Oghuz Uzbek dative in *-GA* after vowels, in Rūh-ābād even after consonants, alternating with *-A*, e.g. *çatunğa* 'to the woman', *dārviša* 'to the dervish'. The un-Oghuz Uzbek future suffix in *-Gay* has entered some areas, as has the southern Uzbek personal suffix of the first-person plural in *-blz* instead of *-mlz*, both occurring in Northeast Khorasan Turkic and Langar.

Khorasan is truly an eclectic transitional region, and it is difficult to summarise the Northwest, South-Southeast and North-Northeast-Langar dialects. If it were not for the transitional phenomena linking them, they might rather be characterised as three separate Oghuz languages.

The reciprocal influences may be summarised as follows: Persian has influenced Azerbaijanian, South Oghuz, Turkmen, especially Khorasan Turkic and Khalaj. Azerbaijanian has influenced Khorasan Turkic, and also Persian (slightly). South Oghuz and Khorasan Turkic have had very little impact. In former times Khorasan Turkic influenced Uzbek. Khorasan Turkic and South Oghuz have perhaps also influenced Khalaj. Oghuz Uzbek or Uzbek has influenced North Khorasan Turkic (slightly), Northeast Khorasan Turkic and Langar. In former times it also influenced Turkmen. Turkmen has virtually had no impact on these dialects.

The variation discussed is characteristic of un-normed dialects which have not produced an independent written language and are influenced by a dominant foreign language, Persian in this case. Interestingly, speakers often exhibit a 'double grammar'. Besides their native dialect they may, in their narratives, employ foreign Turkic elements which are difficult to classify as Turkish, Azerbaijanian, Uzbek, or archaisms. Thus the verbal plural suffix *-lAn* in Kalāt (near Esferāyen) appears as *-lAr* in narratives.

Features of Khalaj

With Khalaj we enter a completely new territory. The language has a special character due to its highly archaic qualities and its contacts with Khorasan Turkic, and secondarily with South Oghuz. Khalaj displays extreme characteristics in two ways: a huge foreign influence on the one hand and strong archaism and autonomy on the other.

The vowel system has undergone extreme changes under Iranian influence: *ö* > *e*, *ü* > *i*, and *ī* appearing only as an allophone of *i*. There are often double forms such as *köz* ~ *kez* 'eye'. Only the northern dialects mostly preserve *ö* and *ü*, the Oghuz neighbourhood serving as a trigger. Sound harmony has been deeply affected; see instances such as *Vardiq* 'We went', *Hačūldi* 'It was opened'.

Morphology and syntax are largely Iranian. Thus, *-ī* appears as an indefinite article beside *bī* and *bī* ... *-ī*, corresponding exactly to Persian *yek*, *-ī*, *yek* ... *-ī*, e.g. *Pādišah-ī vār* 'There is a padishah'. *Bī hat vārum* is equivalent to Turkish *Atım var* 'I have a horse'. Here, a personal suffix is attached to *vār* analogously to Persian (*Yek*) *asb(ī) dāram*.

The lexicon is strongly influenced by Persian. Even many numerals are copied from Persian, namely those from 60 upwards and in money and time constructions, e.g. *Mēndīčā häštād-u-čōr tumān pul va* 'I have 84 tumans of money'.

Tātī too has strongly influenced Khalaj, especially the vocabulary, e.g. *diryāšim* 'burn mark' ← *Āštiyān dirišom*, cf. Middle Persian *drōšm*. The

peculiar conjugation system of Khalaj, which has an imperative stem that sometimes employs a prefix – a completely unusual phenomenon in Turkic – is mainly reminiscent of Tātī forms. The Tātī verb for 'chew' has the infinitive *zuwān*, the preterite *bīšzuwā*, and the imperative *bazun*; the verb for 'go' has the corresponding forms *ziyan*, *basse*, *bassa*. Compare these forms with Khalaj *čäynamäk*, *čäynadi*, *čäynäyi* (not just *čäynä*); *varmaq*, *vardi*, *yovar-* (therefore in some dialects also the simple past *yovardi*).

Furthermore, there has been influence from Oghuz, e.g. in words such as *dām* 'roof' and *gayin-bāba* 'father-in-law' beside genuine Khalaj *gādun-bāba*, with preserved long vowel and *-d-*.

Examples of the extreme archaicism of Khalaj are the following. The Khalaj vowel system displays three-fold quantity, namely short, e.g. *hat* 'horse', half-long, e.g. *sa-t-* 'sell', long, e.g. *āt* 'name' (compare Turkmen *at*, *sat-*, *āt*).

Khalaj is almost the only Turkic language which consistently has Old Turkic *h-*. This old element is still observable in an old Tibetan source: *ud ha-dak-leg* 'cow-footed'. There are also clear traces of *h-* in the South Oghuz of Kābul. The feature has interesting extra-Turkic connections. The Khalaj word for 'dung' is *harq*, which is the equivalent of Mongolian *harǰal* and Manju *fajan*. The three forms clearly point to a Proto-Turkic **parka* (cf. p. 96).

Old Turkic *-δ-* has become *-y-* in most Turkic languages, e.g. in Turkish *ayak* 'foot'. Yakut has *-t-*, Tuvan *-d-*, Khakas and Fu-yü *-z-*. Compare these forms with Khalaj *hadaq*. There is, however, no close connection between Khalaj and Tuvan.

Final and medial *g* and *γ* have been preserved, e.g. *hatluy* 'horseman' (cf. Oghuz *atlu* > *atli*).

Another remarkable fact is that Khalaj is the only modern Turkic language to have kept the dative in *-KA*, e.g. *hävkä* 'to the house' corresponding to Turkish *eve*. Oghuz mostly has the suffix *-(y)A*, other Turkic languages *-GA*.

Furthermore the vocabulary contains such ancient words as *yōr-* 'sit' and *hū-* 'bring out', of which *hūn-* 'come out' is the reflexive.

Let us summarise: through early isolation, Khalaj has retained many archaic features. Through the move into a new – Iranian and Oghuz – environment, it has adopted many new elements.

Further Research

Research on the Turkic varieties of Iran continues. Recent publications include a number of important studies on dialects, analyses of contact phenomena, dictionaries and text editions. For further references consult the bibliographies included in the works cited below.

Although many gaps have been filled in recent years, much research remains to be done. For Turcologists Iran is still a land of future discoveries.

References and Further Reading

- Behzādī, B. (1991) *Farhang-e āzārbāyjanī-fārsī*, Teheran.
- Bozkurt, M. F. (1975) *Untersuchungen zum Bojnurd-Dialekt des Chorasantürkischen*, Dr. Phil. thesis, University of Göttingen.
- Doerfer, G. (1970) 'Irano-Altaistica: Turkish and Mongolian languages of Persia and Afghanistan', in T. A. Sebeok (ed.) *Current Trends in Linguistics 6*, The Hague, Paris: Mouton, 217–34.
- (1976) 'Das Vorosmanische', *Türk Dili Araştırmaları Yıllığı Belleten*, Ankara: Türk Dil Kurumu, 81–131.
- (1988) *Grammatik des Chaladsch*, Turcologica 4, Wiesbaden: Harrassowitz.
- (1993) 'Chaladschica extragottingensia', *Central Asiatic Journal* 37: 33–81.
- and Heschke W., (1989) *Südghusische Materialien aus Afghanistan und Iran*, Wiesbaden: Harrassowitz.
- and ——— (1993) *Chorasantürkisch*, Turcologica 16, Wiesbaden: Harrassowitz.
- , ——— and Ravanyar J. (1990) *Oghusica aus Iran*, Wiesbaden: Harrassowitz.
- and Tezcan S. (1994) *Folklore-Texte der Chaladsch*, Turcologica 19, Wiesbaden: Harrassowitz.
- Foy, K. (1903–4) 'Azerbajganische Studien mit einer Charakteristik des Südtürkischen', *Mitteilungen des Seminars für Orientalische Sprachen*, 2. *Westasiatische Studien*, Berlin, 6.126–93, 7.197–265.
- Johanson, L. (1998) 'Code-copying in Irano-Turkic', *Language Sciences* 20.
- Kıral, F. (1997) *Syntaktische Einflüsse des Persischen auf das gesprochene Aserbaidschanisch von Iran*, Dr. Phil. thesis, University of Mainz.
- Peyfun, M. (1983) *Azərbaycanja farsça lüyāt*, Teheran: Danya.
- Ritter, H. (1921, 1935) 'Azerbeidschanische Texte zur nordpersischen Volkskunde', *Der Islam* 11: 181–212, 25: 234–68.
- Szapszal, H. S. (1935) *Próby literatury ludowej Turków Azerbajdzanu perskiego*, Cracow: Polska Akademia Umijetności, Prace Komisji Orjentalistycznej 18.
- Tulu, S. (1989) *Chorasantürkische Materialien aus Kalāt bei Esfarāyen*, Islamkundliche Untersuchungen 128, Berlin: Schwarz.

17 Tatar and Bashkir

Árpád Berta

Introduction

Tatar and Bashkir are closely related Volga Kipchak languages, belonging to the northwestern or Kipchak branch of Turkic languages. Tatar has been a written language since the second half of the nineteenth century, developed from the so-called Volga Turki language. The phonetic basis of modern literary Tatar is Kazan Tatar, whereas its morphology is based on the Mishar dialect. The modern Bashkir written language emerged after 1920 and is based on the so-called Mountain dialect.

Sound Systems

Phonetics and Phonology

Both Tatar and Bashkir have nine vowel phonemes, the back vowels /ɨ/, /a/, /u/, /ɔ/ and the front ones /i/, /ɛ/, /e/, /ü/, /ö/. The Tatar combination /ɨ/ + /y/ is sometimes taken to be a phoneme /i/, a back equivalent of /i/, but /y/ is a consonant segment. In modern loanwords from Russian, /o/, /e/ and /i/ are also found. The vowels of native words can be divided into fully articulated low and high vowels, /a/, /e/, /i/, /u/, /ü/, and reduced mid vowels, /ɨ/, /ɛ/, /ɔ/, /ö/. The quantitative difference is non-phonemic.

The phoneme /a/ has a slightly rounded variant *â* occurring in first syllables and in a syllable after *â*, e.g. *bâlâ* 'child', and is otherwise unrounded, e.g. *uram* 'street'. Unstressed reduced vowels alternate with zero in colloquial speech, e.g. *kššë* ~ *kšë* 'person'.

The Tatar consonant phonemes are /b/, /w/, /g/, /d/, /z/, /y/, /k/, /l/, /m/, /n/, /p/, /t/, /s/, /t/, /f/, /ç/, /š/, /j/, /x/, /ŋ/. Loanwords in literary Tatar also contain /v/, /ž/, /ç/, /h/. Though /f/ is of foreign origin, it is also found in native words, e.g. *yafraq* 'leaf'. The Bashkir consonant phonemes are /b/, /w/, /g/, /d/, /ɔ/, /y/, /k/, /l/, /m/, /n/, /ŋ/, /p/, /t/, /s/, /θ/, /t/, /f/, /x/, /h/, /š/. In loanwords, /v/, /ž/, /z/, /ç/, /ç/ are also attested. In both languages, /šç/ is only found in borrowings from Russian.

In native words, /k/ is realised as a velar *k* in front syllables, and as a postvelar *q* in back syllables. The corresponding front variant of /g/ is a velar voiced stop *g*, whereas the back variant is a postvelar voiced fricative *ɣ* or stop

ğ, both henceforth transcribed as ɣ. The variants of Tatar /č/ and /j/ stretch from the palatalised affricates [tʃʃ] and [dʒʒ] to the palatalised fricatives [ʃ] and [ʒ].

The phonetic differences between Tatar and Bashkir mainly concern their consonant systems. Tatar *s-* and *-s-* correspond to Bashkir *h-*, *-h-*, e.g. *saban* vs. *haban* 'plough', *yasa-* vs. *yaha-* 'make'. Tatar *č* corresponds to Bashkir *s*, e.g. *čiq-* vs. *siq-* 'go out', *ač-* vs. *as-* 'open', *baskič* vs. *baθqış* 'stairs'. Tatar *-s* corresponds to Bashkir *-θ*, e.g. *bas-* vs. *baθ-* 'press', *tös* vs. *töθ* 'colour', *as* vs. *aθ* 'lower part'. Tatar *z* corresponds to Bashkir *δ*, e.g. *zur* vs. *dur* 'big, great', *özön* vs. *öδön* 'long', *qız* vs. *qıδ* 'girl, daughter', *süz* vs. *hüδ* 'word'. The Tatar variants *y-* and *j-* correspond to Bashkir *y-*.

In the following chapter, corresponding Tatar and Bashkir forms will be written with the equals sign between them, e.g. *qız = qıδ*.

Though the word accent usually rests on the last syllable, some exceptions are found. Second-person imperatives have an accent on the first syllable, e.g. Tatar *Bäşla* 'Begin!', Bashkir *Ültirñyđ* 'Sit down!' Copula suffixes such as first-person singular *-mın* are unaccented. The negative suffix *-ma* is unaccented, and the preceding syllable accented, e.g. *Bárma* 'Do not go!' Postpositions and enclitic particles are unaccented, e.g. Tatar *baltá-bēlen* 'with the axe', Bashkir *táw-östönde* 'on the mountain'. The first component of compounds is accented, e.g. Tatar *çatın-qız* 'women', Bashkir *börön-böröndan* 'formerly'.

Phonotactics and Morphophonology

In certain bisyllabic stems, the second vowel is deleted if suffixes with initial vowel are added, e.g. *börön* 'nose', *börnö* 'its nose'. Stem-final /p/ and /k/ are weakened to /b/ and /g/ when a suffix with an initial vowel is added, e.g. *tap-* 'find' → *tabıl-* 'be found', *ayaq* 'foot' → *ayayı* 'his foot', *aq-* 'flow' → *ayıδ-* 'let flow'.

Native words exhibit intersyllabic front vs. back harmony, e.g. *öy-ler* 'houses', *at-lar* 'horses'. Rounded vs. unrounded harmony is weakly developed. However, *ö* and *ȫ* are the only reduced vowels occurring after an *ö* or *ȫ* of the stem syllable, and do not otherwise appear in non-first syllables.

Final vowels of Tatar verbal stems undergo changes when certain suffixes, e.g. those based on the *-A* converb, are attached. From a practical point of view, the initial vowel of these suffixes, written as *-(θ)l(y)*, may be said to 'replace' the stem-final vowel, e.g. *ěšli(y)*, present of *ěšle-* 'work'.

The quality of suffix-initial consonants is determined by the stem-final segment. Bashkir suffixes often have more allomorphs than their Tatar counterparts. Many Tatar suffix-initial consonants have two variants, voiced after stems ending in vowels and voiced consonants, e.g. *başla-dı* 'began', *awıl-ya* 'to the village', *taw-da* 'on the mountain', and voiceless after other stems, e.g. *qayt-tı* 'returned', *ayaq-ta* 'on foot'. Ablative and plural suffixes

have initial *n* after the nasals *n*, *m*, *ŋ*, e.g. *uram-nar* 'streets', *urman-nan* 'from the forest'.

Bashkir alternations are similar, but partly more complicated. The plural suffix *-LAR* is realised as *-lAr* after vowels, *-tAr* after voiceless consonants, *-δAr* after *w*, *y*, *δ*, *r*, and *-dAr* after other voiced consonants, e.g. *bülmeler* 'rooms', *attar* 'horses', *tawdar* 'mountains', *yěrδer* 'places', *külδer* 'lakes', *yěŋder* 'sleeves'. The ablative suffix *-DAn* is realised as *-nAn* after vowels, *-tAn* after voiceless consonants, *-δAr* after *w*, *y*, *δ*, *r*, and *-dAr* after other voiced consonants, e.g. *bülme-nen* 'out of the room', *at-tan* 'from the horse', *taw-dan* 'from the mountain', *yěr-δen* 'from the place', *yalan-dan* 'from the steppe'. The simple past suffix *-DE* has a similar distribution of allomorphs, e.g. *yaha-nım* 'I made', *kit-tēm* 'I went away', *yaδ-đım* 'I wrote', *al-dım* 'I took'. The locative suffix *-DA*, however, has *-lA* after vowels, e.g. *bülme-le* 'in the room', *qala-la* 'in the city', but the usual variants after consonants, e.g. *at-ta* 'on the horse', *taw-δa* 'on the mountain', *yalan-da* 'in the steppe'.

Native words display six types of syllable structure, V, VC, CV, VCC, CVC, CVCC. The orthographic representation of recent loanwords also suggest types such as CCV, CCVC, CCCVC.

Orthography

The orthographic systems of Tatar and Bashkir are somewhat different, Bashkir spelling being closer to pronunciation in some cases. Rounded reduced vowels in non-first syllables are written as rounded in Bashkir, but as unrounded in Tatar, e.g. Bashkir *озон*, Tatar *озын* *özön* 'long', Bashkir *көмөш*, Tatar *көмеш* *kömöš* 'silver'. In syllable-final position, the Tatar signs *и*, *у* and *ү* represent *ĩ*, *ũluw*, *ü/üw*, respectively. Bashkir orthography reflects the phonetic values by using *ей*, *уу*, *еү*.

Unlike Tatar, Bashkir uses different graphemes to distinguish velars and postvelars, *к* for *k*, *г* for *g*, *к* for *q* and *ф* for *ɣ*. The alternation of word-initial *y-* ~ *j-* in Tatar colloquial speech is not reflected in the orthography, but standardised according to the quality of the following vowel. In native words, the consonant is written as *j-* before *i*, mostly as *y-* before *a*, *e*, *ö*, *ȫ*, *u*, *ü*, and as *y-* or *j-* before *ě* and *ĩ*. Syllable-initial *yě* and *yı* are rendered with *e* in Tatar; Bashkir *yı* is rendered with *йы*. Tatar *ya* and *ye* are rendered with *я*; Bashkir *ye* is rendered with *йә*.

In some cases where the value of the Tatar graphemes *e*, *я*, *ю* is ambiguous, *ь* is used to mark frontness of the vowel, e.g. *яшь yeš* 'young', *юнь yün* 'cheap', whereas absence of *ь* implies backness, e.g. *ял yal* 'rest', *як yak* 'side'. In Bashkir, *ь* has the same function with the grapheme *ю*, e.g. *юнь yün* 'cheap'.

The assimilations *n > ŋ* before /g/ and /k/, and *n > m* before /b/ and /m/ are not reflected in the orthography.

Literary Tatar used the Arabic script up to 1927, subsequently systems based on Latin script, and, from 1939 onwards, a modified Cyrillic alphabet. The small group of Christian Tatars already used a modified Cyrillic alphabet

at the end of the nineteenth century. Literary Bashkir was written in Arabic script up to 1928, in a modified Latin script from 1928 to 1940, and finally in a Cyrillic script that is rather different from the Tatar one.

Morphology

Nouns and Adjectives

Word Formation

Derivational suffixes forming nominal stems from nominal stems include -AK, e.g. Tatar *başaq* 'ear of corn', Bashkir *kiθek* 'piece', -čAK = -sAk, e.g. Bashkir *uyñsaq* 'little toy', Tatar *irənček* 'lazy, sluggish', -čĖ = -sĖ, e.g. *ěščĖ* = *ěšsĖ* 'worker', *balıqçı* = *balıqsı* 'fisherman', -čĖK = sĖK, e.g. *öyčök* = *öysök* 'little house', *taščik* = *taşşıq* 'little stone', -čĖl = -sĖl, e.g. Tatar *üpkęčĖl* 'touchy, insulted', Bashkir *dinsĖl* 'religious', -DAş, e.g. Tatar *awıldaş* 'fellow-villager', Bashkir *mektepeş* 'schoolmate', -GĖ, e.g. *qışqı* = *qışqı* 'wintry', *alyı* 'frontal', *öskö* = *öθkö* 'upper', -KAy, e.g. Tatar *balaqay* 'dear little baby', *esekey* 'mummy'. The stem-final vowel is omitted in words such as Tatar *etkey* 'daddy' ← *eti* 'father' and *enkey* 'mummy' ← *eni* 'mother', -lĖ, e.g. *aqıllı* 'clever', *köçlö* = *köslö* 'strong'.

-lĖK, e.g. *süzlĖk* = *hüdlĖk* 'dictionary', *yĖğĖtlĖk* 'bravery', *küplĖk* 'multitude', -sĖz = -hĖđ, e.g. *ayaqsız* = *ayaqhıđ* 'footless', *ěšsĖz* = *ěšhĖđ* 'jobless'.

Nominal stems are derived from verbal stems with the following suffixes, some of which are not productive any more, e.g. -(Ė)č = -(Ė)s, e.g. *ışanıč* = *ışanıš* 'confidence', *söyönöč* = *höyönös* 'joy', -(Ė)K, e.g. *ilek* 'sieve', *taraq* 'comb', *bülĖk* 'section', -(Ė)m, e.g. *ayım* 'current', *bĖlĖm* 'knowledge', -(Ė)mtA, e.g. *quşımta* 'addition', *sözömtę* = *höđömtę* 'extract, conclusion', -(Ė)ş, e.g. *çıyış* = *çıyış* 'exit', *kilĖş* 'arrival', -GAK, e.g. *törgek* 'wrapping', *işkek* 'oar', *batqaq* 'mud, swamp', -GĖ, e.g. *čalyı* = *salyı* 'scythe', *kıyĖrgĖ* = *kĖyĖrgĖ* 'frame for embroidering', -GĖč (after consonants), -wĖč (after vowels), e.g. *ačqıč* = *asqış* 'key', *qaplawıč* = *qaplawış* 'lid', -GĖn, e.g. *azgın* = *ađyın* 'immoral', Bashkir *kiθkĖn* 'sharp', -GĖr, e.g. *sizgĖr* = *hiđgĖr* 'perceptive', *tapqır* 'clever, ingenious', -KA, e.g. *töŧqa* 'handle', *qışqa* = *qıθqa* 'short', -mA, e.g. *basma* = *baθma* 'stair, step', *bülme* 'room', -mAK, e.g. *uymaq* 'thimble', *ĖlmeK* 'knot, loop'.

Comparison

Comparatives are formed with -rAK, e.g. *özönraq* = *öđönraq* 'longer', *uqıyanraq* 'more learned'. In a comparative construction, the basic form may also be used, e.g. *yulbarıştan köčlö* 'stronger than the tiger'. Superlatives are formed with the preposed particle *iĖ*, e.g. *iĖ matur* 'most beautiful'.

Certain adjectives have a reduplicative intensive form, e.g. Tatar *ap-aq* 'snow-white', *yem-yeşĖl* 'very, very green'. Colour names may take detensive suffixes, e.g. *qıżılsu* = *qıđılhıw* 'reddish', *kügĖljĖm* = *kügĖlyĖm* 'bluish'.

Declension

The plural suffix is -LAR. Nouns are usually not marked with this suffix after numerals and counting words, e.g. Tatar *biş tölkö* 'five foxes', *küp kön* 'many days', except in some expressions denoting time, e.g. *ikĖ atmalar* 'two weeks'. The suffixless singular may also be interpreted as referring to several entities in phrases such as *at köti* - 'let a horse/horses graze'.

Tatar and Bashkir have six grammatical cases: suffixless nominative, genitive in -nĖĖ, accusative in -nĖ, dative in -GA, locative in -DA and ablative in -DAn.

The nominative is typically used for subjects. The accusative is used to mark specific direct objects, e.g. *Kitapnı aldım* 'I took the book'. Non-specific direct objects mostly lack the accusative suffix, e.g. Tatar *Kitap aldım* 'I took a book/books'. The genitive forms possessive attributes, e.g. Tatar *kitapnıĖ işı* 'the cover of the book'. The dative marks expressions of direction, aim, goal, duration, etc. The locative marks expression denoting the place and time of events, etc. The ablative denotes the starting point, the material of which something is made, etc.

Possessive suffixes, typically used to express the person and number of a possessor, have the following forms:

Table 17.1 Possessive suffixes

	Singular	Plural
1p.	-(Ė)m	-(Ė)bĖz = -(Ė)bĖđ
2p.	-(Ė)Ė	-(Ė)GĖz = -(Ė)GĖđ
3p.	-(s)Ė = -(h)Ė	-(s)Ė = -(h)Ė

The order of suffixes added to nominal stems is number, possessive, case, e.g. Tatar *dus-lar-ım-nan* 'from my friends', Bashkir *qala-lar-ıbıđ-đan* 'from our cities'. The following tables show the nominal paradigms.

Table 17.2 Case suffixes in Tatar

		at 'horse'	eti 'father'	urman 'forest'
nom.	θ	at	eti	urman
gen.	-nĖĖ	atnıĖ	etinĖĖ	urmannıĖ
acc.	-nĖ	atnı	etinĖ	urmannı
dat.	-GA	atqa	etige	urmanıya
loc.	-DA	atta	etide	urmanda
abl.	-DAn	attan	etiden	urmannan

Table 17.3 Case suffixes in Bashkir

		<i>at</i> 'horse'	<i>bala</i> 'child'	<i>yër</i> 'earth'	<i>urman</i> 'forest'
nom.	∅	<i>at</i>	<i>bala</i>	<i>yër</i>	<i>urman</i>
gen.	-nĕŋ	<i>atıñ</i>	<i>balanıñ</i>	<i>yërdëñ</i>	<i>urmandıñ</i>
acc.	-nĕ	<i>atı</i>	<i>balanı</i>	<i>yërdë</i>	<i>urmandı</i>
dat.	-GA	<i>atqa</i>	<i>balaya</i>	<i>yërge</i>	<i>urmanya</i>
loc.	-DA	<i>atta</i>	<i>balala</i>	<i>yërde</i>	<i>urmanda</i>
abl.	-DAn	<i>attan</i>	<i>balanan</i>	<i>yërden</i>	<i>urmandan</i>

Table 17.4 Declension with possessive suffixes in Tatar

	<i>1sg.</i>	<i>2sg.</i>	<i>1pl.</i>	<i>2pl.</i>	<i>3sg./pl.</i>
nom.	<i>añım</i> 'my horse'	<i>añıñ</i> 'your horse'	<i>añbız</i> 'our horse'	<i>añyız</i> 'your horse'	<i>añ</i> 'its/their horse'
gen.	<i>añımñıñ</i>	<i>añıññıñ</i>	<i>añbızñıñ</i>	<i>añyızñıñ</i>	<i>añıñ</i>
acc.	<i>añımñı</i>	<i>añıññı</i>	<i>añbızñı</i>	<i>añyızñı</i>	<i>añ</i>
dat.	<i>añıma</i>	<i>añıña</i>	<i>añbızğa</i>	<i>añyızğa</i>	<i>añıma</i>
abl.	<i>añımnan</i>	<i>añıñnan</i>	<i>añbızdan</i>	<i>añyızdan</i>	<i>añınnan</i>
loc.	<i>añımda</i>	<i>añıñda</i>	<i>añbızda</i>	<i>añyızda</i>	<i>añında</i>
nom.	<i>etım</i> 'my father'	<i>etiñ</i> 'your father'	<i>etiböz</i> 'our father'	<i>etigöz</i> 'your father'	<i>etisë</i> 'its/their father'
gen.	<i>etımñëñ</i>	<i>etiññëñ</i>	<i>etibözñëñ</i>	<i>etigözñëñ</i>	<i>etisëñëñ</i>
acc.	<i>etımñë</i>	<i>etiññë</i>	<i>etibözñë</i>	<i>etigözñë</i>	<i>etisëñ</i>
dat.	<i>etıme</i>	<i>etiñe</i>	<i>etibözge</i>	<i>etigözge</i>	<i>etisëne</i>
loc.	<i>etımda</i>	<i>etiñda</i>	<i>etibözde</i>	<i>etigözde</i>	<i>etisënda</i>
abl.	<i>etımnen</i>	<i>etiñnen</i>	<i>etibözden</i>	<i>etigözden</i>	<i>etisënnen</i>

Table 17.5 Declension with possessive suffixes in Bashkir

	<i>1sg.</i>	<i>2sg.</i>	<i>1pl.</i>	<i>2pl.</i>	<i>3sg./pl.</i>
nom.	<i>añım</i> 'my horse'	<i>añıñ</i> 'your horse'	<i>añbıd</i> 'our horse'	<i>añyıd</i> 'your horse'	<i>añ</i> 'its/their horse'
gen.	<i>añımđıñ</i>	<i>añıñđıñ</i>	<i>añbıdđıñ</i>	<i>añyıdđıñ</i>	<i>añıñ</i>
acc.	<i>añımđı</i>	<i>añıñđı</i>	<i>añbıdđı</i>	<i>añyıdđı</i>	<i>añ</i>
dat.	<i>añıma</i>	<i>añıña</i>	<i>añbıdğa</i>	<i>añyıdğa</i>	<i>añıma</i>
abl.	<i>añımдан</i>	<i>añıñдан</i>	<i>añbıdдан</i>	<i>añyıdдан</i>	<i>añınан</i>
loc.	<i>añımda</i>	<i>añıñda</i>	<i>añbıdda</i>	<i>añyıdda</i>	<i>añında</i>
nom.	<i>balam</i> 'my child'	<i>balanıñ</i> 'your child'	<i>balabıd</i> 'our child'	<i>balayıd</i> 'your child'	<i>balanıñ</i> 'its/their child'
gen.	<i>balamđıñ</i>	<i>balanıñđıñ</i>	<i>balabıdđıñ</i>	<i>balayıdđıñ</i>	<i>balanıñıñ</i>
acc.	<i>balamđı</i>	<i>balanıñđı</i>	<i>balabıdđı</i>	<i>balayıdđı</i>	<i>balanıñ</i>
dat.	<i>balama</i>	<i>balanıña</i>	<i>balabıdğa</i>	<i>balayıdğa</i>	<i>balanıña</i>
loc.	<i>balamda</i>	<i>balanıñda</i>	<i>balabıdda</i>	<i>balayıdda</i>	<i>balanıñda</i>
abl.	<i>balamдан</i>	<i>balanıñдан</i>	<i>balabıdдан</i>	<i>balayıdдан</i>	<i>balanıñан</i>

Pronouns

Tatar and Bashkir possess personal, demonstrative, reflexive, interrogative and indefinite pronouns.

The paradigm of singular personal pronouns differs slightly from that of nouns. The oblique stem of the third-person singular pronoun *ul* is *an-* in Tatar and *un-* in Bashkir. The segment *n* is dropped in the Bashkir dative form: *uya*. Tatar exhibits the demonstrative pronouns *bu* 'this', *şuşı* 'this here', *şul* 'that', *tögë* 'that', *ul* 'that' with the plurals *bölar* 'these', *şuşılar* 'these here', *şular* 'those', *tögëler* 'those', *alar* 'those'. The Bashkir singular forms are *bıl* 'this', *bınaw* 'this', *öşö* 'this here', *şul* 'that', *anaw* 'that', *tögë* 'that', *ul* 'that'.

The reflexive pronoun *üz* = *üd* takes possessive suffixes and is inflected as a noun. The corresponding Bashkir nominative forms are *üdëm* 'myself',

Table 17.6 Declension of personal pronouns

	'I'		'you'		'he, she, it'	
nom.	min =	min	sin =	hin	ul =	ul
gen.	minëm =	minëñ	sinëñ =	hinëñ	anıñ =	unıñ
acc.	minë =	minë	sinë =	hinë	anı =	unı
dat.	miña =	miñe	siña =	hiñe	aña(r)/añar(ya) =	uğa
loc.	minde =	minde	sinde =	hinde	anda/añarda =	unda
abl.	minnen =	minen	sinnen =	hinen	annan/añardan =	unan
	'we'		'you'		'they'	
nom.	böz =	böd	söz =	höd	alar =	ular
gen.	bözñëñ =	bödđëñ	sözñëñ =	hödđëñ	alarmıñ =	ularđıñ
acc.	bözñë =	bödđë	sözñë =	hödđë	alarmı =	ularđı
dat.	bözge =	bödđge	sözge =	hödđge	alarya =	ularya
loc.	bözde =	bödđde	sözde =	hödđde	alarda =	ularđa
abl.	bözden =	bödđden	sözden =	hödđden	alardan =	ularđan

Table 17.7 Declension of reflexive pronouns in Tatar

	'myself'	'thyself'	'himsel'	'ourselves'	'yourselves'	'themselves'
nom.	<i>üzëm</i>	<i>üzëñ</i>	<i>üzë</i>	<i>üzëböz</i>	<i>üzëgöz</i>	<i>üzlerë</i>
gen.	<i>üzëmnëñ</i>	<i>üzëññëñ</i>	<i>üzëñëñ</i>	<i>üzëbözñëñ</i>	<i>üzëgözñëñ</i>	<i>üzlerëñëñ</i>
acc.	<i>üzëmnë</i>	<i>üzëññë</i>	<i>üzëñ</i>	<i>üzëbözñë</i>	<i>üzëgözñë</i>	<i>üzlerëñ</i>
dat.	<i>üzëme</i>	<i>üzëñe</i>	<i>üzëne</i>	<i>üzëbözge</i>	<i>üzëgözge</i>	<i>üzlerëne</i>
loc.	<i>üzëmda</i>	<i>üzëñda</i>	<i>üzënda</i>	<i>üzëbözde</i>	<i>üzëgözde</i>	<i>üzlerënda</i>
abl.	<i>üzëmnen</i>	<i>üzëñnen</i>	<i>üzënnen</i>	<i>üzëbözden</i>	<i>üzëgözden</i>	<i>üzlerënnen</i>

üðeŋ 'thysself', *üðe* 'himself', *üðebëð* 'ourselves', *üðeğëð* 'yourselves', *üðlerë* 'themselves'.

Interrogative pronouns include Tatar *këm* 'who', plural *këmner*, *nerse* 'what', plural *nerseler*, *ni* 'what', plural *niler*, *ninëŋ* 'of what', *nige* 'to what', *nindi* 'what, what kind of', *niçe* 'how many', *qay*, *qaysı* 'what', Bashkir *këm* 'who', *neme* 'what', *ni* 'what', *qay*, *qayhı* 'what', *qaynı* 'which'. Most interrogative pronouns are declined as nouns.

Indefinite pronouns are mostly based on interrogative ones, e.g. *hiçkëm* = *his këm* 'nobody', *herkëm* = *her këm* 'everybody', *hiçnerse* = *his neme* 'nothing', *hernerse* = *her neme* 'everything'.

Numerals

The cardinal numerals from one to ten are *bër* 'one', *ikë* 'two', *öç* = *ös* 'three', *dürt* 'four', *biş* 'five', *altı* 'six', *jidë* = *yëtë* 'seven', *sigëz* = *higëð* 'eight', *tuyız* = *tuyıð* 'nine', *un* 'ten'. The tens are *yëğërmë* 'twenty', *utız* = *utıð* 'thirty', *qırıq* = *qırq* 'forty', *illë* 'fifty', *altmıış* 'sixty', *jitmëş* = *yëtmëş* 'seventy', *siksen* = *hikken* 'eighty', *tuqsan* = *tuqhan* 'ninety'. Higher cardinals are *yöz* = *yöd* 'hundred', *mëŋ* 'thousand', *tömen* 'ten thousand', *million* 'million'. Numerals are combined as in most other Turkic languages, e.g. *unbër* 'eleven', *mëŋ ëlli ikë* '1052', *biş yöz* '500', Tatar *ikë yöz mëŋ* '200,000'. The cardinals 101 to 119 and 1001 to 1099 optionally contain the particle *DA* 'and' after the first element, e.g. Tatar *yöz de öç* '103', *mëŋ de illë ikë* '1052'.

Tatar has a number of counting words preceding the word denoting the counted entity, e.g. *baş* 'piece', *töp* 'bottom', *kiyëm* 'pair'.

Ordinals are formed with *-(E)nčE* = *-(E)nsE* added to the last member of the cardinal construction, e.g. *bërenčë* = *bërensë* 'first', *ikëncë* = *ikënsë* 'second', *yëğërmë altınčı* = *yëğërmë altınsı* 'twenty-sixth'. The interrogative ordinal is *niçenčë* = *nisensë* 'which (in a series)'. The same suffix appears in the lexeme *urtančı* = *urtansı* 'middle'. Approximatives are formed with *-lap*, e.g. *unlap* 'about ten', *yözlep* = *yödlep* 'about a hundred'.

Verbs

Denominal Verb Derivation

Verbal stems are derived from nominal stems with suffixes such as *-A*, e.g. *aşa*- 'eat' ← *aş* 'food', *yeşe*- 'live' ← *yeş* 'age', *-(A)r*, e.g. *küger*- 'become blue' ← *kük* 'blue', *ayar*- 'become white' ← *aq* 'white', *yaşır*- = *yaşır*- 'become good' ← *yaşır* = *yaşır* 'good', *-Ay*, e.g. *azay*- = *aday*- 'decrease in number' ← *az* = *ad* 'few', *kübey*- 'increase in number' ← *küp* 'many', *kiñey*- 'become wide' ← *kiñ* 'wide', *-(E)K*, e.g. *bërëk*- 'unite' ← *bër* 'one', *yulıq*- 'meet' ← *yul* 'way', *-GAR*, e.g. *jilger*- = *yëlger*- 'winnow' ← *jil* = *yël* 'wind', *suyar*- = *hiuyar* 'water' ← *su* = *hiu*, *başqar* 'begin' ← *baş* 'head', *-lA*, e.g. *başla*- 'begin' ← *baş* 'head', *jırta*- = *yırta*- 'sing' ← *jır* = *yır* 'song', *ëşle*- 'work' ← *ëş* 'work'.

Compound verbal stems contain a nominal element and a verb, e.g. *küğël bul*- 'be satisfied' ← *küğël* 'spirit', *bul*- 'be, become', *küğël ač*- = *küğël as*- 'amuse oneself' ← *ač*- = *as*- 'open', *çëzmet it*- = *çëðmet it*- 'work' ← *çëzmet* = *çëðmet* 'job', *it*- 'do'.

Deverbal Verb Derivation

Passives are formed with *-(E)n* (mostly after *l* and *lA*) and *-(E)l*, e.g. *al*- 'take' → *alın*- 'be taken', *alda*- 'deceive' → *aldan*- 'be deceived', *başla*- 'begin' → *başlan*- 'be started', *yaz*- = *yad*- 'write' → *yazıl*- = *yadııl*- 'be written', *yasa*- = *yaha*- 'make' → *yasal*- = *yahal*- 'be made'.

Reflexive verbs are formed with *-(E)n* and may thus coincide with passives, e.g. *ki*- = *këy*- 'dress' → *kiyën*- = *këyën*- 'dress oneself', *bize*- = *bide*- 'adorn' → *bizen*- = *biden*- 'adorn oneself'.

Causatives are derived from intransitive and transitive verbs. The suffix *-(E)t* is added to polysyllabic stems ending in vowels, *r* and *y*, e.g. *aşa*- 'eat' → *aşat*- 'make eat', *qurq*- 'fear' (< **qorqı*-) → *qurqıt*- 'frighten', *nıyay*- = *nıyay*- 'become hard, strong' → *nıyayt*- = *nıyayt*- 'strengthen'. The suffix *-(E)r* is used after monosyllabic stems ending in *-ş*, *-t*, Tatar *-ç*, Bashkir *-s*, e.g. *pëş*- = *bëş*- 'cook' (intransitive) → *pëşër*- = *bëşër*- 'cook' (transitive), *küç*- = *küs*- 'move away' → *küçër*- = *küsër*- 'make move', *bët*- = *böt*- 'come to an end' → *bëtër*- = *bötör*- 'finish'. The suffixes *-Ar*, *-TËr*, *-Ëz*, etc. are added to stems ending in consonants, e.g. *qayt*- 'go back' → *qaytar*- 'let go back', *yan*- 'burn' → *yandıır*- 'make burn', *ač*- 'open' → *ačtır*- 'make open', *töt*- 'seize' → *töttör*- 'make seize', *tam*- 'drip' → *tamız*- = *tamıð*- 'make drip', *im*- 'suck' → *imëz*- = *imëð*- 'make suck'. There are also combinations of causative suffixes such as *-tTËr*, *-TËr-t*, *-GËr-t*.

Cooperative-reciprocal verbs are formed with the suffix *-(E)ş*, e.g. *söyle*- = *höyle*- 'talk' → *söyleş*- = *höyleş*- 'converse', *töt*- 'seize' → *tötöş*- 'seize each other', *yaz*- = *yad*- 'write' → *yazış*- = *yadıış*- 'write to each other'.

Frequentative verbs denoting periodically repeated actions are formed with *-GAlA* = *GËlA*, *-(E)ştËr* or combinations, e.g. Tatar *yaz*- 'write' → *yazyala*- 'write now and then', Bashkir *bar*- 'walk' → *baryıla*- 'walk often', *yasa*- = *yaha*- 'make' → *yasaştır*- = *yahaştır*- 'make from time to time'.

Verbs have negative stems in *-mA*, e.g. *kil*- 'come' → *kilme*- 'not come', *qaytar*- 'let come back' → *qaytarma*- 'not let come back'.

Possibility and impossibility can be expressed with various constructions. The *-A* converb and the auxiliary *al*- form a construction that occurs in all persons, e.g. Tatar *Yaza alam* 'I can write', *Yaza almıym* 'I cannot write', Bashkir *Uqıy aldıq* 'We could read', *Uqıy almanıq* 'We could not read'. The converb in *-(E)p* and the auxiliary *bul*- form an impersonal construction restricted to the third-person singular, e.g. Tatar *Yazıp bula* 'It is possible to write', *Yazıp bulmıy* 'It is impossible to write', Bashkir *Uqıp buldı* 'It was possible to read', *Uqıp bulmanı* 'It was impossible to read'.

A variety of constructions consisting of a converb of main verbs and a following auxiliary verb are used to express actional – descriptive, phasal, etc. – modifications, e.g. Tatar *kürəp al-* ‘catch sight of’ ← *kür-* ‘see’ and *al-* ‘take’ (see pp. 42, 113–14).

Finite Verb Forms

A present is formed with *-A* after consonants and with *-(θ)I(y) = -y* after vowels plus personal markers. An imperfect is formed with the *-A* converb + simple past of *i-* ‘be’, e.g. Tatar *Ekiyetler söyli idě* ‘(S)he was telling stories’ = *Ala ině* ‘(S)he was taking’.

Table 17.8 Present in Tatar

	Suffixes	<i>al-</i> ‘take’	<i>ěšle-</i> ‘work’
1sg.	-A-m(Ĕn), -(θ)I(y)-m(Ĕn)	al-a-m(ĭn)	ěšl-i(y)-m(ĕn)
2sg.	-A-sĔŋ, -(θ)I(y)-sĔŋ	al-a-sĭŋ	ěšl-i(y)-sĕŋ
3sg.	-A, -(θ)I(y)	al-a	ěšl-i(y)
1pl.	-A-bĔz, -(θ)I(y)-bĔz	al-a-bĭz	ěšl-i(y)-bĕz
2pl.	-A-sĔz, -(θ)I(y)-sĔz	al-a-sĭz	ěšl-i(y)-sĕz
3pl.	-A-lAr, -(θ)I(y)-lAr	al-a-lar	ěšl-i(y)-ler

Table 17.9 Present in Bashkir

	Suffixes	<i>al-</i> ‘take’	<i>ěšle-</i> ‘work’
1sg.	-A-m(Ĕn)/-y-m(Ĕn)	al-a-m(ĭn)	ěšle-y-m(ĕn)
2sg.	-A-hĔŋ, -y-hĔŋ	al-a-hĭŋ	ěšle-y-hĕŋ
3sg.	-A, -y	al-a	ěšle-y
1pl.	-A-bĔđ, -y-bĔđ	al-a-bĭđ	ěšle-y-bĕđ
2pl.	-A-hĔGĔđ, -y-hĔGĔđ or -y-hĔđ	al-a-hĭġĭđ	ěšle-y-hĕ(gĕ)đ
3pl.	-A-lAr, -y-đAr	al-a-lar	ěšle-y-đer

A simple past is formed with the suffix *-DĔ* plus personal markers:

Table 17.10 Simple past

	Suffixes	<i>al-</i> ‘take’	<i>ěšle-</i> ‘work’
1sg.	-DĔ-m	al-dĭ-m = al-dĭ-m	ěšle-dĕ-m = ěšle-nĕ-m
2sg.	-DĔ-ŋ	al-dĭ-ŋ = al-dĭ-ŋ	ěšle-dĕ-ŋ = ěšle-nĕ-ŋ
3sg.	-DĔ-θ	al-dĭ = al-dĭ	ěšle-dĕ = ěšle-nĕ
1pl.	-DĔ-k	al-dĭ-k = al-dĭ-g	ěšle-dĕ-k = ěšle-nĕ-k
2pl.	-DĔ-GĔz = -DĔ-GĔđ	al-dĭ-ġĭz = al-dĭ-ġĭđ	ěšle-dĕ-gĕz = ěšle-nĕ-gĕđ
3pl.	-DĔ-lAr	al-dĭ-lar = al-dĭ-lar	ěšle-dĕ-ler = ěšle-nĕ-ler

Table 17.11 *-GAn* past

	Suffixes	<i>al-</i> ‘take’	<i>ěšle-</i> ‘work’
1sg.	-GAn-mĔn	al-yan-mĭn = al-yan-mĭn	ěšle-gen-mĕn = ěšle-ym(ĕn)
2sg.	-GAn-sĔŋ = -GAn-hĔŋ	al-yan-sĭŋ = al-yan-hĭŋ	ěšle-gen-sĕŋ = ěšle-yhĕŋ
3sg.	-GAn-θ	al-yan = al-yan	ěšle-gen = ěšle-gen
1pl.	-GAn-bĔz = -GAn-bĔđ	al-yan-bĭz = al-yan-bĭđ	ěšl-gen-bĕz = ěšle-ybĕđ
2pl.	-GAn-sĔz = -GAn-hĔGĔđ	al-yan-sĭz = al-yan-hĭġĭđ	ěšle-gen-sĕz = ěšle-yhĕ(gĕ)đ
3pl.	-GAn-nAr = -GAn-dAr	al-yan-nar = al-yan-dar	ěšle-gen-ner = ěšle-gen-der

The *-GAn* past, consisting of the *-GAn* participle plus predicative suffixes, is a post-terminal with perfect functions, denoting past events with a connotation of present relevance. It may also express that the event is inferred from a state resulting from it. A pluperfect, denoting events that had already taken place at some point, is formed with the *-GAn* participle + simple past of *i-* ‘be’, e.g. Tatar *Min aŋa yazyan idĕm* ‘I had written to him’.

A habitual past is formed with the *-A* converb + *töryan* + past tense of *i-* ‘be’, e.g. Tatar *Ekiyetler söyli töryan idĕ* ‘(S)he used to tell stories’.

The aorist, formed with the aorist participle plus personal markers, denotes events tending or foreseen to take place. The suffix variant occurring after vowel stems is *-r*. The choice of the suffix vowel, *A* or *Ĕ*, after consonant stems is not phonologically conditioned. The negative stem is formed with *-mAs = -mAθ*, the final consonant of which is dropped in the first person. The first person singular has the personal marker *-m*, not *-mĔn*.

The aorist participle plus the simple past of *i-* ‘be’ may denote an event that

Table 17.12 Aorist

	<i>al-</i> ‘take’		<i>ěšle-</i> ‘work’	
	affirmative	negative	affirmative	negative
1sg.	al-ĭr-mĭn = al-ĭr-mĭn	al-ma-m = al-ma-m	ěšle-r-mĕn = ěšle-r-mĕn	ěšle-me-m = ěšle-me-m
2sg.	al-ĭr-sĭŋ = al-ĭr-hĭŋ	al-mas-sĭŋ = al-maθ-hĭŋ	ěšle-r-sĕŋ = ěšle-r-hĕŋ	ěšle-mes-sĕŋ = ěšle-meθ-hĕŋ
3sg.	al-ĭr = al-ĭr	al-mas = al-maθ	ěšle-r = ěšle-r	ěšle-mes = ěšle-meθ
1pl.	al-ĭr-bĭz = al-ĭr-bĭđ	al-ma-bĭz = al-ma-bĭđ	ěšle-r-bĕz = ěšle-r-bĕđ	ěšle-me-bĕz = ěšle-me-bĕđ
2pl.	al-ĭr-sĭz = al-ĭr-hĭġĭđ	al-ma-sĭz = al-maθ-hĭġĭđ	ěšle-r-sĕz = ěšle-r-hĕgĕđ	ěšle-me-sĕz = ěšle-meθ-hĕgĕđ
3pl.	al-ĭr-lar = al-ĭr-đar	al-mas-lar = al-maθ-tar	ěšle-r-ler = ěšle-r-đer	ěšle-mes-ler = ěšle-meθ-ter

Table 17.13 Prospective

	<i>al-</i> 'take'	<i>ěšle-</i> 'work'
1sg.	al-ačaq-mñ = al-asaq-mñ	ěšle-yeček-měn = ěšle-yesek-měn
2sg.	al-ačaq-sñ = al-asaq-hñ	ěšle-yeček-sěŋ = ěšle-yesek-hěŋ
3sg.	al-ačaq = al-asaq	ěšle-yeček = ěšle-yesek
1pl.	al-ačaq-bñ = al-asaq-bñ	ěšle-yeček-běz = ěšle-yesek-běđ
2pl.	al-ačaq-sız = al-asaq-hıyđ	ěšle-yeček-sěz = ěšle-yesek-hěgěđ
3pl.	al-ačaq-lar = al-asaq-tar	ěšle-yeček-ler = ěšle-yesek-ter

would occur, in the sense of a prospective in the past or counterfactually, e.g. Bashkir *Bññ ěšler inēm* 'I would make it'. *-GAN* + aorist participle of *bul-* 'become, be' + simple past of *i-* 'be' denotes an event that *would have* occurred, e.g. Tatar *uqıyan bulır idēm* 'I would have read'.

A prospective item is formed with the participle *-(y)AČAK = -(y)AsAK* plus personal markers. Its negative forms are regular: *-mAyAČAK = -mAyAsAK*. The combination with the simple past of *i-* 'be' denotes an event intended in the past, e.g. Tatar *Yazačaq iděk* 'We intended to write', Bashkir *Alasaq inē* '(S)he intended to take'.

The complex suffix *-mAKčĚ = -mAKsĚ* denotes intention, e.g. *almaqčġ = almaqsġ* 'intending to take', and also combines with the simple past of *i-* 'be', e.g. Bashkir *Almaqsġ inē* '(S)he intended to take'.

The optative-imperative paradigm is of heterogeneous nature. In the first person, Tatar has *-(θ)I(y) < *-Ay-Ě*, written as *ı* after stems with front vowels, and as *ıñ* after stems with back vowels. The addition of the past tense of *i-* 'be' may express meanings of desirability in the past, e.g. Tatar *Mektepke barsın idē* '(S)he should go to school'.

A conditional is formed with *-sA = -hA* plus personal markers. The conditional + simple past of *i-* 'be' denotes counterfactual conditions. The personal markers are attached to the main verb or the auxiliary, e.g. Tatar *kilsem idē* or *kilse idēm* 'if I had come'.

Table 17.14 Optative-imperative

	Suffixes		<i>al-</i> 'take'		<i>ěšle-</i> 'work'	
	singular	plural	singular	plural	singular	plural
1p.	<i>-(θ)I(y)m = -(θ)I(y)K</i>	<i>-(θ)I(y)K</i>	<i>al-ıym = al-ayım</i>	<i>al-ıyq = al-ayıq</i>	<i>ěšl-i(y)m = ěšl-i(y)k</i>	<i>ěšl-i(y)k = ěšl-i(y)k</i>
2p.	<i>-θ</i>	<i>-(Ě)gĚz = -(Ě)GĚđ</i>	<i>al = al</i>	<i>al-ıyız = al-ıyıđ</i>	<i>ěšle = ěšle</i>	<i>ěšle-gěz = ěšle-gěđ</i>
3p.	<i>-sĚn = -hĚn</i>	<i>-sĚnnAr = -hĚndAr</i>	<i>al-sñn = al-hñn</i>	<i>al-sñnnar = al-hñndar</i>	<i>ěšle-sěn = ěšle-hěn</i>	<i>ěšle-sěnnar = ěšle-hěnder</i>

Table 17.15 Conditional

	<i>al-</i> 'take'	<i>ěšle-</i> 'work'
1sg.	al-sa-m = al-ha-m	ěšle-se-m = ěšle-he-m
2sg.	al-sa-ŋ = al-ha-ŋ	ěšle-se-ŋ = ěšle-he-ŋ
3sg.	al-sa = al-ha	ěšle-se = ěšle-he
1pl.	al-sa-q = al-ha-q	ěšle-se-k = ěšle-he-k
2pl.	al-sa-yız = al-ha-yıđ	ěšle-se-gěz = ěšle-he-gěđ
3pl.	al-sa-lar = al-ha-lar	ěšle-se-ler = ěšle-he-ler

Non-finite Verb Forms

The most frequent verbal noun, used as the dictionary form, ends in *-(V)w*, e.g. *aluw = alıw* 'taking', *ěšlew* 'working'. The suffix is realised as *w* after stem-final vowels and as vowel + *w* after consonants. An infrequent verbal noun is formed with *-mAK*, e.g. *almaq* 'taking', *ěšlemek* 'working'.

The intraterminal (present) participle consists of an *-A* converb and the *-GAN* form of the auxiliary verb *tör-* 'stand', e.g. *ala törıyan* 'taking', *ěšli törıyan = ěšley törıyan* 'working'. The post-terminal ('past') participle ends in *-GAN*, e.g. *alyan* 'having taken', *ěšlegen* 'having worked'. The negative forms are regular, e.g. *almayan* 'not having taken', *ěšlemegen* 'not having worked'. A necessitative participle is formed with the complex suffix *-AsĚ/ -ysĚ = -AhĚ/ -yhĚ*, e.g. *alası = alahı* 'bound to take'. It also combines with the simple past of *i-* 'be', e.g. Bashkir *alahı idē* 'had to take'. As mentioned, there are also a prospective (future) participle in *-(y)AČAK = -(y)AsAK* and an aorist participle in *-(V)r*, negated *-mAs = -mAθ*. The form in *-(Ě)wčĚ = -(Ě)wsĚ* functions as a noun or as an adjective, e.g. Tatar *aluwčġ* 'one who takes', Bashkir *alıwsġ kěšē* 'the man who takes'.

The intraterminal *-A* converb, formed with *-A* after consonants and *-I(y)* = *-y* after vowels, denotes single or repeated events simultaneous with the event of the superordinate verb, e.g. *al-* 'take' → *ala*, *ěšle-* 'work' → *ěšli = ěšley*. Corresponding negative forms are Tatar *almıy*, *ěšlemi*, Bashkir *almay*, *ěšlemey*. The first actant of this converb, whether expressed as an overt subject or not, is coreferential with that of the main verb. The converb in *-(Ě)p* typically denotes an event beginning before the event expressed by the main verb, e.g. *alıp* 'takes and ...' etc. It is not formed from negative stems. The converb in *-GAč = -GAs* is used to refer to an event preceding the event of the main verb, e.g. *ěšlegeč = ěšleges* 'after having worked'. The converb in *-GANčĚ = -GAnsĚ* denotes an event taking place later than the event of the main verb, e.g. *ěšlegenčē = ěšlegensē* 'before/until working'. The so-called infinitive in *-(V)rGA*, e.g. *alırya* '(in order) to take', a dative form of *-(V)r*, negated *-mAsKA = -mAθKA*, may be used as a purposive converb, e.g. Tatar *uqırya* 'in order to study'.

Adverbs

Suffixes forming adverbs from other word classes include $-čA = -sA$, e.g. *kěšě* 'human being' → *kěšěče = kěšěse* 'in a human way', $-DAy$, e.g. *qoš* 'bird' → *qoštay* 'like a bird', $-lAtA$, e.g. *ikě* 'two' → *ikělete* 'two together', *akča = aqsa* 'money' → *akčalata = aqsalata* 'in cash', $-DAn$ (ablative), e.g. Bashkir *yañınan*, Tatar *yañadan* 'anew'. Some temporal adverbs go back to forms of the old instrumental suffix $-(Ė)n$, e.g. *qışın* 'in winter' → *qış* 'winter', *tönön* 'at night' → *tön* 'night'.

There are compound adverbs such as *arı-birě* 'back and forth', *könön-tönön* 'day and night'. Some contain a component that cannot be used independently, e.g. Tatar *epən-töpön* 'stumbling', Bashkir *qıyıq-mıyıq* 'aslant'.

Other Morphological Categories

The postpositions *bėlen = mėnen* 'with', *öčön = ösön* 'for', *kėbėk, kük = kėwėk* 'like', *aša* 'through', *qader = qėder* 'up to' govern the nominative of nouns and the genitive of certain pronouns. Postpositions governing the dative are *taba* 'towards', *qarşı* 'against' and *küre* 'in view of'. Those governing the ablative are *başqa* 'besides, except', *birlě* 'from, since', *söğ = huğ* 'after'.

There are also postpositions based on possessive forms of auxiliary nouns such as *yan* 'side', *art* 'back', *urta* 'middle', *ara* 'interval', e.g. Tatar *öy aldında* 'in front of the house', Bashkir *öy artınan* 'from behind the house'.

Conjunctions include *hem* 'and', *tayı* 'and, also', *lekin, e, emme* 'but, however', *belki* 'but', *tik* 'but, only', *ye, yeki* 'or', *yeyse* 'or, or else', *čönki = sönki* 'because', *güye* 'as if', *ki* 'that', *eger* 'if, when', *eger de = egeřde* 'if'.

Of the numerous interjections, some express joy and pleasure, e.g. *e, ay, hu, aba, ix, ihi*, others anger, e.g. *e-e, hey, aχ, uχ, abay, ay-hay*, or surprise, e.g. *ay-ay, ay-yay, ixı-xı, ix-ma*.

Enclitic particles, mostly obeying the rules of sound harmony, include the interrogative particle $-mĖ$, e.g. *Bu atmı?* 'Is this a horse?', *Anı kürdėğmė?* = *Unı kürdėğmė?* 'Did you see him?', and the extended form $mĖni(y) = mĖněy$, which adds a shade of doubt, e.g. *Ul kilgenmėni = Ul kilgenmėněy?* 'Has (s)he really come?' Some particles give emphasis to the preceding word or phrase, e.g. *DA* 'also, too', *GĖnA, KĖnA* 'only', *UK* 'exactly', etc. The particle $čĖ = sĖ$ attached to imperatives adds the meaning 'please', e.g. *Alčı = Al sı* 'Please take it!'

Syntax

Nominal Phrases

In nominal phrases headed by a noun, the attribute can be an adjective, a numeral, a pronoun or a participle, e.g. Tatar *al bayraq* 'red banner', *öç at* 'three horses', *uqıyan kėšě* 'a learned man', Bashkir *bıl defter* 'this notebook', *ütķer qılıs* 'sharp sword'. The attribute can also be a noun, a proper name, the

name of a profession, etc., e.g. *yartı ay* 'half a month', Bashkir *Azat ayay* 'Uncle Azat', Tatar *timėrčė Gali* 'the blacksmith Ali'. Attributes are usually placed in the following order: demonstrative pronoun, numeral, adjective, e.g. Tatar *bu öç qara kitap* 'these three black books'.

Adjectives used as nouns take the same suffixes as nouns, e.g. Tatar *qızıllar* 'the red ones'.

Genitive attributes denoting a possessor are followed by the head with a third-person possessive suffix, e.g. Tatar *bu kitapnıñ üşı* 'the cover of this book'. If the attribute is a first- or second-person pronoun, the head may or may not carry a possessive suffix that agrees with it, e.g. Tatar *minėm defter* or *minėm defterėm* 'my notebook'. Except for these possessive cases, there is no agreement morphology in nominal phrases.

The main type of compound noun is the one in which the second word carries a third-person possessive suffix, e.g. Tatar *kitap üşı* 'book cover', Bashkir *qala baqsahı* 'city park'. In such constructions, the first element may also be a proper name, e.g. Bashkir *Öfö qalahı* 'the city of Ufa', Tatar *Qazan šeherė* 'the city of Kazan'.

Compound nouns can also be formed with juxtaposition of semantically coordinated members as in *ata-ana = ata-ese* 'parents', Tatar *yört-jir* 'farm', *qaz-ürdek* 'poultry', Bashkir *taw-taş* 'rocky country', *duθ-iş* 'friends', or of one modifying and one modified member as in Tatar *almayač* 'apple-tree', *yulbasar* 'highwayman'.

Numerals may be juxtaposed coordinatively to express approximative numbers, e.g. *bėr-ikě* 'one or two'.

Adjectival and Other Phrases

Adjectival phrases have an adjective, a participle etc. as their head. The dependent may be an adverb, e.g. *bik zur = bik dur* 'very great, large', or a case-marked nominal, e.g. Tatar *küzge yayımlı* 'pleasant to the eye', Bashkir *nėftke bay* 'rich in oil'. Adjectives in comparative constructions are preceded by ablative-marked nominals, e.g. Bashkir *attan durıraq* 'bigger than a horse'.

Adjectival juxtapositions may be reduplicative, e.g. *özön-özön = ödön-ödön* 'long-long', *kėčė-kėčė = kėse-kėse* 'tiny little', or composed of rhyming words, e.g. Tatar *kekrė-bökrö* 'crooked', *arqılı-törqölö* 'crosswise'. The first adjective may modify the second one, e.g. *quyı qızıllı = quyı qıdıllı* 'dark-red'.

Pronominal phrases may consist of a personal and a reflexive pronoun, e.g. *min üzėm* 'I myself'.

Verbal Phrases

Verbal phrases may contain an accusative-marked or unmarked direct object, e.g. Tatar *xat yaz-* 'write a letter/letters', *bu xatnı yaz-* 'write this letter', or an ablative or locative complement, e.g. Bashkir *döşmandan qurq-* 'fear the/an enemy', *sağıyla yörö-* 'ski'. They may also contain other complements, such as a directive nominal in the dative or with a postposition, e.g. Tatar

yulya čik- 'start off on a journey', *awılyya qayt-* 'return to the village', Bashkir *urmanıya taban atla-* 'go towards the forest'.

Tatar and Bashkir have rich systems of constructions with converbs + auxiliary verbs, e.g. *barıp jıt-* = *barıp yět-* 'arrive' ← *bar-* 'go' and *jıt-* = *yět-* 'reach', Bashkir *aşıya bar-* 'go in a hurry' ← *aşıq-* 'haste' and *bar-* 'go', Tatar *yaqınlaşıp kil-* 'come near' ← *yaqınlaş-* 'approach' and *kil-* 'come'.

Adverbials

Adverbial complements may carry special markers, e.g. Bashkir *tatarsa yađ-* 'write in Tatar', Tatar *ač büređey* 'like a hungry wolf', but nominals may also function adverbially without any markers, e.g. Bashkir *Yaqşı uqıym* 'I read well', *Şul kěşě eşli aqıllı* 'That man works reasonably'. Adverbial phrases may have an adverb as dependent, e.g. *bik irte* 'very early' = *bik yaqşı* 'very well'.

In postpositional phrases, the postposition is preceded by a nominal category, e.g. Bashkir *balta mēnen* 'with the axe', Tatar *kičeden birle* 'since yesterday', *siněŋ bēlen* 'with you'.

The Sentence

Finite Clauses

If the predicate of a finite clause is verbal, it agrees with the subject in person, e.g. Bashkir *Ul yađa aldı* '(S)he was able to write', Tatar *Min kitapnı alıp kildēm* 'I brought the book'. If it is nominal, it can occur with copula suffixes, e.g. Bashkir *Min uqıwsımın* 'I am a pupil', *Min Qazannanmın* 'I am from Kazan'. Note that this construction is also possible without a copula, e.g. Bashkir *Min awıldan* 'I am from the village'. The third person usually has *-θ*, e.g. Tatar *Kitap siněkě de, miněkě de* 'The book is both yours and mine'. Nominal predicates may also occur with copula particles, e.g. *idě* = *ině* 'it was' and Tatar *-ise* 'if it was' (fossilised forms of the defective verb *i-*) or with copula verbs such as *bul-* 'be' or *tör-* 'stand, stay', e.g. Bashkir *Ul böğön bēđge kilmeksě bula* '(S)he intends to come to us today'.

Agreement in person between subject and predicate is usually obligatory when the predicate is a verb, e.g. Tatar *Min jıyılışqa qaldım* 'I remained at the meeting', Bashkir *Bēđ qalanan qaytıq* 'We returned from the city'. Agreement is usual, though not obligatory, when the predicate is a noun, e.g. Bashkir *Min yađıwsımın* 'I am a writer', Tatar *Běz kolxozçı* 'We are collective farmers'. There is mostly no number agreement between plural subjects and predicates. The predicate is often in the singular if the subject is marked for plural, e.g. Tatar *Alar qaytı* 'They returned', *Tawlar biyěk idě* 'The mountains were high', Bashkir *Ularđıŋ balaları bik matur* 'Their children are very pretty'. The plural is possible if the referents are conceived of as individuals, e.g. Bashkir *Ularđıŋ balaları bik maturdar* 'Their children are very pretty'. Agreement is also possible with more than one subject, e.g. Bashkir *Qarluyas hem törna qōšťar* 'The swallow and the crane are birds'.

Pronominal subjects may be dropped according to special rules, e.g. Tatar *Kildēm* 'I came' instead of *Min kildēm* 'I came'.

Constituent Order

The predicate usually occupies the final position, e.g. Bashkir *Bala yaqşı* 'The child is good', *Min bıl xeberdē išettēm* 'I have heard this news', Tatar *Dusım kildē* 'My friend came', but a verbal predicate may also precede the subject and/or another complement, e.g. Bashkir *Qaytı ul* '(S)he has come back', *išettēm min bıl xeberdē* 'I have heard this news'.

Negation

The predicate of a negative sentence can be a negative verb form, e.g. *Běz sině kürneděk* = *Bēđ hine kürmeněk* 'We did not see you', or a nominal form negated with *tügěl* 'not', e.g. Tatar *Bu taş tügěl* 'This is not a stone', *Ul kileček tügěl* '(S)he will not come', Bashkir *Bıl öy matur tügěl* 'This house is not pretty'. The negation of *bar* 'existing, there is' is *yuq* 'non-existing, there is not', e.g. Bashkir *Minde at yuq* 'I have no horse', Tatar *Anı hič kürgenēm yuq* 'I have never seen him'.

Interrogatives

Yes/no questions are expressed with rising intonation and with or without the interrogative particle *-mĔ*, e.g. Bashkir *Hin bıl kitapnı aldıŋ?* 'Did you take this book?', *Atıŋ šep tügelmä?* 'Is not your horse fast?', Tatar *Sěz anı bēlēsöz?* 'Do you know him?', *Ul kildēm?* 'Did (s)he come?' Questions are also formed with interrogative pronouns or adverbs plus falling intonation, e.g. *Kēm kildē?* 'Who came?', *Ni buldı?* 'What happened?' In questions such as Bashkir *Hin bıl defterdē kēmgē birdēŋ?* 'To whom did you give this notebook?', Tatar *Kiče sēzge kēm kildē?* 'Who came to you yesterday?', the intonation rises up to the accented syllable of the interrogative word, and then falls.

Coordination

Coordination of phrases, clauses and sentences may be marked with conjunctions or expressed by mere juxtaposition, e.g. Tatar *Jir yeşermes, göl açılmas* 'The earth does not become green, the rose does not unfold'.

Non-finite Clauses

Tatar and Bashkir non-finite clauses are formed with participles, verbal nouns and converbs according to the normal Turkic patterns (see pp. 59–65).

Lexicon

The basic lexical fields – numerals, kinship terminology, nouns denoting natural phenomena, parts of the body, animals, plants, and verbs denoting basic human activities – are dominated by words of Turkic origin. More or

less significant groups of borrowings are of Middle Mongolian, Arabic, Persian and Russian origin. Words of Finno-Ugric origin occur mainly in the dialects.

Dialects

Tatar can be divided into three main dialects: Kazan Tatar or Central dialect, the Mishar or Western dialect and the West-Siberian or Eastern dialect. These can be further divided into subdialects and local varieties. Bashkir has three dialects and numerous subdialects. The main dialects are the Mountain or Eastern dialect, the Steppe or Southwestern dialect and the Northwestern dialect.

Further Reading

- Berta, Á. (1989) *Lautgeschichte der tatarischen Dialekte*, Studia Uralo-Altaica 31, Szeged: Attila József University.
- Golovkina, O. V. (ed.) (1966) *Tatarsko-russkij slovar'*, Moscow: Sovetskaja Ėnciklopedija.
- Maksjutova, N. H. (1976) *Vostočnyj dialekt baškirkogo jazyka*, Moscow: Nauka.
- Maxmutova, L. T. (ed.) (1969) *Tatar tēlēnēj dialektologik süzlēgē*, Kazan: Tatarstan Kitap Näšriyatı.
- Miržanova, S. F. (1979) *Južnyj dialekt baškirkogo jazyka*, Moscow: Nauka.
- Poppe, N. (1963) *Tatar Manual. Descriptive Grammar and Texts with a Tatar-English Glossary*, Indiana University Publications, Uralic and Altaic Series 25, Bloomington: Indiana University, The Hague: Mouton.
- (1964) *Bashkir Manual. Descriptive Grammar and Texts with a Bashkir-English Glossary*, Indiana University Publications, Uralic and Altaic Series 36, Bloomington: Indiana University, The Hague: Mouton.
- Schönig, C. (1984) *Hilfsverben im Tatarischen: Untersuchungen zur Funktionsweise einiger Hilfsverbverbindungen*, Akademie der Wissenschaften und der Literatur Mainz, Veröffentlichungen der Orientalischen Kommission 35, Wiesbaden: Steiner.
- Uraqsin, Z. G. (ed.) (1996) *Baškortsa-russa hüdlēk*, Moscow: Digora.

18 West Kipchak Languages

Árpád Berta

Introduction

The modern West Kipchak languages are Kumyk, Karachay-Balkar, Crimean Tatar, and Karaim in its Trakai and Halič varieties.

Sound Systems

Phonetics and Phonology

The vowel systems generally distinguish the eight phonemes /a/, /e/, /i/, /ī/, /ö/, /o/, /u/, /ü/. Halič Karaim lacks the rounded front /ö/ and /ü/. Kumyk has a fronted variant of /a/, *a*, before syllables containing *i* or *ī*, and in the environment of *h* and *k* in words of foreign origin. The occurrence of *e* and *ē* as variants of /e/ is determined by position. Kumyk *e* occurs in non-first syllables, whereas *ē* occurs in primary stems and suffixes containing *y*, e.g. *Bèremen* 'I give', *G'èteles* 'They leave'. Crimean Tatar /i/ has two variants, *i* and a reduced *ĩ*. This difference is phonemic in the northern dialect, e.g. *kir* 'mud' vs. *kĩr* 'come in'. Rounded low /o/ and /ö/ are mostly restricted to first syllables. Kumyk initial /o/ is pronounced *u*^o, e.g. *u^onda* 'there', *u^ot* 'fire'. In all languages, /ö/ and /ü/ are somewhat backed to *ó* and *ü*.

Though old vowel length is not preserved, distinctions between half-long and short vowels are found in some monosyllabic Balkar and Kumyk words, e.g. Kumyk *bĩz* 'awl' vs. *biz* 'we'. Secondary lengthening is mostly due to contraction, e.g. Kumyk *Barādim* 'I went' < *bara edim*, *gelmē* 'to come' < *gelmege*, or it occurs in open syllables if the following vowel is reduced, e.g. Karachay-Balkar *qārĩn* 'stomach', *ōrũn* 'seat'.

The consonant phonemes generally are /b/, /v/, /w/, /g/, /d/, /j/, /ʒ/, /z/, /y/, /k/, /l/, /m/, /n/, /ŋ/, /p/, /r/, /s/, /t/, /ʃ/, /č/, /š/, /h/. Some languages have /ʃč/ in more recent Russian loanwords. Kumyk and Karachay-Balkar /f/, /v/, /c/ are found in words of foreign origin. The phoneme /h/ is present in Crimean Tatar dialects, but it is extremely rare in Kumyk, e.g. *hiž* 'line', and

absent in Karachay-Balkar. Under Slavic influence, Trakai Karaim consonants exhibit palatalised variants in front environments, e.g. *d'*, *g'*, *k'*, *l'*, *m'*, *n'*, *s'*, *z'*, *t'*, *š'*. Halič Karaim has, except in front of *i*, /dz/ instead of /j/, /c/ instead of /č/, and /z/ instead of /ž/, whereas /ŋ/ is missing. Kumyk dialects display the glottal stop /ʔ/ as a phoneme, mainly in Caucasian loanwords.

The front realisations of /k/ and /g/ are generally *k* and *g*, whereas the back realisations are *q* and *ɣ*, respectively. Trakai Karaim has *k'* and *g'* in front environments, and *k* and *g* in back environments. Karachay-Balkar and Crimean Tatar even offer evidence for the oppositions /k/ vs. /q/ and /g/ vs. /ɣ/. Kumyk /g/ has the variants *g*, *ɣ'*, *g̃*, *ɣ* and, before rounded vowels, *g^u*.

Kumyk /j/ has the variant *ž*, typically occurring between vowels. Under the influence of Ossetian and Caucasian languages, Karachay-Balkar voiceless obstruents exhibit aspirated and non-aspirated variants: *p^h* and *p*, *t^h* and *t*, *k^h* and *k*, *q^h* and *q*, *č^h* and *č*. The aspirated ones occur word-initially, in the second position of consonant clusters, between vowels, and word-finally if the following word begins with a vowel, e.g. *k^helt^hir-* 'bring', *č^hač* 'hair', *q^horq^hup* 'fearing', *q^hič^hir-* 'shout', *q^ha^hin* 'woman', *a^h az* 'the horses are few'. Kumyk has aspirated variants of voiceless stops such as *p^h*, *t^h*, *q^h*, the non-aspirated variants occurring after consonants and word-finally.

As for word accent, there are, as in other Turkic languages, unaccentable suffixes, e.g. copula suffixes, the negative suffix *-mA*, interrogative particles and the adverb suffix *-čA*.

Diachronic Changes

Karachay-Balkar **a* has become *e* in front of *y*, *j*, and *i*, e.g. Balkar *sewdi* 'milked', *bey* 'very'. Some other back vowels have also been fronted in the neighbourhood of *y*, e.g. *yiyin* 'army'. The shift *i* > *e* also occurs in the environment of Karachay-Balkar *č* and *j*, Balkar *z*, e.g. *bičaq* 'knife'. Fronting of *i* is also observed in Kumyk, especially in syllables after *a*, e.g. *sari* > *sari* 'yellow'. Due to regressive assimilation, word-final *iw* and *iw* have generally developed into *üw* and *uw*, e.g. Halič Karaim *acuw*, Trakai Karaim *ačuv*, in Karachay-Balkar dialects into the long vowels *ū* and *ü*, e.g. *ačū* 'anger'.

Kumyk shows cases of initial *j*- instead of *y*-, e.g. *jüy*- 'collect'. In Balkar dialects, *j*- is replaced by *z*-, *ž*-. Karaim *y*- is unstable before high vowels, e.g. *yig'it'* ~ *ig'it'* 'young', Halič *yillar* ~ *illar* 'years', Trakai *yür'äk'* ~ *ür'äk'* 'heart', *yüv* ~ *üv* 'house'. A similar phenomenon occurs in Karachay-Balkar before low vowels, e.g. *aman* 'bad' < *yaman*, *aχšü* 'good' < *yaχšü*, though Karachay-Balkar normally has *j*- instead of *y*-.

Unlike the normal situation in Kipchak, initial *d*- and *g*- are found relatively often, e.g. Crimean Tatar *dört* 'four', *digle-* 'listen', *gemi* 'ship', Balkar *doğuz* 'pig', *göz* 'eye', *geten* 'linen'. The replacement of *k*- by *g*- allows minimal pairs such as *gök* 'blue' vs. *kök* 'sky', *göp-* 'swell' vs. *köp* 'many'.

Phonotactics and Morphophonology

Word-final stops are voiceless. The consonants *p*, *k*, *χ* and sometimes *č* are voiced between vowels, e.g. Trakai Karaim *t'öš'äk* 'bed' → *t'öš'äg'im* 'my bed', Karachay *ešek* 'donkey' → *ešegim* 'my donkey', Crimean Tatar *qapaq* 'top' → *qapayı* 'its top', also in recent loanwords such as *spisok* 'list' → *spisogi* 'its list'. The high vowel in the second syllable of certain bisyllabic words disappears in front of a suffix beginning with a vowel, e.g. Crimean Tatar *burun* 'nose' → *burnu* 'its nose', *köküs* 'chest' → *köksü* 'its chest'. Initial high vowels may also disappear, e.g. Crimean Tatar *(i)šte* 'behold', *(u)sta* 'master'.

Front vs. back harmony is usually consistent in native words. In Trakai Karaim, front vowels occur with palatalised consonants and back vowels with non-palatalised consonants, e.g. *it'l'er'im'iz* 'our dogs', *atlarimiz* 'our horses'. There are various exceptions to the intersyllabic front vs. back harmony. Thus, the Crimean Tatar nominal suffixes *-Daš*, *-mač* and *-Day* have invariable vowels, and the same is true of some Halič Karaim copula suffixes, e.g. *Barasen* 'You go', *Cababiz* 'We run', *Kolasiz* 'You ask', *Kaytibiz* 'We will return', *Sindiriz* 'Break!' In Trakai Karaim the shift *-ay* > *-ey* may cause harmony violations, e.g. *atey* 'your father' ← *ata* 'father', *Barıeym* ~ *barıeymin* 'I should come'. In words such as Kumyk *babiš* 'duck', *bišlaq* 'cheese', Crimean Tatar *böjek* 'insect', *ütü* 'iron', *čabik* 'fast', the first-syllable vowel is relatively fronted: *á*, *ó*, *ü*.

Rounded vs. unrounded harmony affects high vowels of the second syllable, e.g. Karaim *kolum* 'my hand', Trakai *K'ör's'un* 'Let him see'. Halič Karaim provides fewer examples since *ö* and *ü* are lacking. Low suffix vowels are seldom affected, e.g. Karaim *kollar* 'hands', not **kollor*. The harmony is most consistently applied in Balkar, where forms such as *ölgön* 'dead', *üydö* 'in the house' occur.

Due to progressive assimilation, suffix-initial consonants are devoiced after voiceless sounds, e.g. Crimean Tatar *bašta* 'on the head', Karaim *atti* 'shot'. Kumyk and Karachay-Balkar orthography conceals the devoicing, e.g. Kumyk *ишден* 'from work'. Certain suffixes are invariable, e.g. Crimean Tatar *-čiq*, *-čan*.

Progressive nasal assimilation is found in Crimean Tatar: *nl* > *nn*, e.g. *čobannar* 'shepherds', *ŋl* > *ŋn*, e.g. *ağna-* 'understand'. In colloquial Kumyk and Karachay, the clusters *nd*, *ŋd*, *md* are replaced by *nn*, *ŋn*, *mn* respectively, e.g. Kumyk *sałamna* 'in the straw' < *sakumda*, Karachay *k^helinnen* 'from the bride' < *kelinden*, *t^hağnan* 'from dawn' < *tağdan*.

Some assimilations are less consistent: Kumyk *l + d* > *ll*, *n + g* > *ŋŋ*, *m + g* > *mŋ*, e.g. *awullaš* 'fellow villager', *qumŋan* 'tankard', Balkar *m + n* > *mm*, e.g. *anammi* 'of my mother', Karachay *-n + g*, *ɣ* > *nŋ*, e.g. *qatınğa* 'to the woman', *burunğu* 'early'. Not all assimilations are indicated orthographically.

Cases of non-standard regressive assimilation are *n + l* > *ll* in Kumyk and Karachay-Balkar, e.g. Kumyk *ulallar* 'boys' < *ulanlar*, *taballı* 'with soles' <

tabanlı. Kumyk aorist forms are also affected, e.g. *G'èteller* 'They will go'. Other cases are Kumyk $z + s > ss$, e.g. *yassa* 'if (s)he writes', $n > m$ before *b, m, p*, e.g. Karachay-Balkar *kelgemme* 'I came', Karachay-Balkar *ombir* 'eleven', Karachay-Balkar $l + \eta > \eta\eta$, e.g. *jaŋŋiz* 'only'. The written forms are Гетерлер, язса, келгенме, онбир, джанъыз, etc. Long *ll* is dissimilated to *ŋl* in Karaim, e.g. Trakai *uŋlu* 'big' < *ullu*, *koŋlar* 'hands' < *kollar*. The dissimilation of *čč* into *šč* in colloquial Kumyk is not marked in the orthography, e.g. Иччи *Išči* 'Drink up!'

Sandhi phenomena include the development of Kumyk word-initial *q* into *γ* after word-final voiced sounds, e.g. *qizit yan* 'red blood', *ati γot* 'six hands', $n + b > nm$, e.g. *munnan mašya* 'besides this' < *mundan bašya*. After word-final voiced sounds, Kumyk *g'*- shifts to *ʳy*- before unrounded vowels and to *ʳw* before rounded vowels, e.g. *uffi ʳyiš'i* 'big man', *ek'i ʳwül* 'two roses'. If *ʳy* or *ʳw* follows a word ending in *-n*, the stop fuses with the nasal, e.g. *barayaŋ yiši* 'going man' < *barayan ʳyiš'i*, *batyaŋ wün* 'setting sun' < *batyan ʳwün*. These orthoepic features are not reflected graphically.

Orthography

The Arabic script was used for writing Crimean Tatar until 1928. A Latin alphabet was then in use for almost ten years and finally replaced by a Cyrillic alphabet. Kumyk employed a Latin alphabet from 1929 to 1938, when a Cyrillic one was introduced. For Karachay-Balkar, a reformed Arabic alphabet was in use between 1916 and 1926, a Latin alphabet between 1927 and 1939, and a Cyrillic alphabet from 1940 onwards. The Karaims have no unified literary language. The Hebrew alphabet was used for older texts, and later on, printed materials appeared in Latin and Cyrillic scripts based on Polish, Russian and Ukrainian orthographies. Currently, efforts are being made to introduce an orthography based on Lithuanian.

As for *k* and *g* sounds, the Kumyk, Karachay and Crimean Tatar orthographies mark front variants with *κ* and *γ*, and back variants with *къ* and *гъ*. Kumyk *ж* indicates both *ž* and *j*. Karachay-Balkar *ж* indicates *ž* in Russian loanwords, while *j* is written *дж*. Kumyk *в* represents *v* in recent Russian loanwords and *w* in native words. Karachay distinguishes *v* and *w* by using the Cyrillic letters *в* and *у*, respectively. The sound *ŋ* is represented by *нγ* in Kumyk, by *нъ* in Karachay-Balkar and Crimean Tatar.

Kumyk *y* is written *й* before *i* and *ï*. Combinations with other vowels are written *e = ye*, *я = ya*, *ë = yo*, *yö*, *ю = yu*, *yü*. Elsewhere, *ö* and *ü* are written *оь* and *уь*, respectively. In Karachay and Crimean Tatar, labial front vowels are written *ë* and *ю*, whereas *я* represents *ya*. Kumyk *h* is written *гъ* and appears initially in words of foreign origin.

Morphology

Nouns and Adjectives

Word Formation

The following denominal nominal suffixes are widely used: *-lXK*, e.g. Halič Karaim *dostluk* 'friendship', Trakai *tašliχ* 'road', Kumyk *bičenlik* 'pasture', *göklük* 'blue', Crimean Tatar *tuzluq* 'saltcellar', Kumyk, Karachay *ašliq* 'crop, seed', *-čX*, e.g. Karachay *otunču* 'woodcutter', Trakai Karaim *jolču* 'traveller', Kumyk *qoyču* 'shepherd' (the Crimean Tatar suffix vowel is *l*, e.g. *odunji* 'woodcutter'), *-dAŠ*, e.g. Trakai Karaim *yoldaš* 'spouse, fellow passenger', Kumyk *gündeš* 'rival, competitor', Crimean Tatar *soydaš* 'relative', *-čXK*, e.g. Karachay *naratčiq* 'small pine', *jolčuq* 'tiny road', *künčük* 'little sun'. In Crimean Tatar, the vowel of the latter suffix is *l*, and the initial consonant is invariable, e.g. *odunčiq* 'small log'. Trakai Karaim *-čAK*, and Halič Karaim *-cAK*, e.g. *t'ilček*, *t'ilcek* 'little tongue', are historically distinct from *-čXk*.

The Karaim feminine suffixes *-KA* and *-ČA* are of Slavic origin, e.g. *oruska* 'Russian woman' ← *orus* 'Russian', Trakai Karaim *dostča* 'girlfriend' ← *dost* 'friend'.

The deverbal nominal suffixes *-mA*, *-mAK* and *-(X)w* may denote results or object of actions, e.g. Balkar, Kumyk *tüyme* 'button' ← *tüy-* 'tie a knot', Karachay *basma* 'printing press' ← *bas-* 'print', Trakai Karaim *talamaχ* 'robbery' ← *tala-* 'rob', Kumyk *ilmek* 'noose' ← *il-* 'hang', Karachay *batmaq* 'swamp' ← *bat-* 'sink', Kumyk *boyaw* 'paint' ← *boya-* 'paint'. The suffix vowel of *-(V)w* is *l* in Karachay e.g. *jaziw* 'writing' ← *jaz-* 'write', and sometimes *U* in Karaim, e.g. Trakai Karaim *satuw* 'trading' ← *sat-* 'sell'. The suffix *-GXČ* denotes instruments, e.g. Trakai Karaim *ačxič* 'key' ← *ač-* 'open', Kumyk *süzgüč* 'sieve' ← *süz-* 'sift', Karachay *bašxič* 'stairs' ← *bas-* 'step'. Numerous other suffixes are used with varying degrees of frequency.

Adjectives do not differ formally from nouns. Denominal adjective markers include: *-lX*, e.g. Karaim *kutlu* 'lucky' ← *kut* 'luck', Kumyk *erli* 'married' ← *er* 'husband', Crimean Tatar *köylü* 'peasant' ← *köy* 'village'. The Karachay suffix vowel is *U*, e.g. *t'awlu* 'mountainous' ← *taw* 'mountain'. Other adjective suffixes are *-lXk*, e.g. Kumyk *ayliq* 'monthly' ← *ay* 'month', Crimean Tatar *küzlük* 'autumnal' ← *küz* 'autumn', *-sXz*, e.g. Halič Karaim *ucsuz* 'endless' ← *uc* 'end', Kumyk *tilsiz* 'mute' ← *til* 'tongue', Karachay *üysüz* 'homeless' ← *üy* 'house'. Crimean Tatar has *-slz*, e.g. *sessiz* 'voiceless' ← *ses* 'voice'.

Suffixes that weaken the meaning of adjectives include *-muš*, e.g. Trakai Karaim *kizilmuš* 'reddish', *uzunmuš* 'longish', and Crimean Tatar *-ČA*, *-jArAk*, *-Xlilm*, e.g. *ajjija* 'somewhat bitter', *ajjijaraq* 'a little bitter', *kögültim* 'bluish'. The suffix *-KX* forms adjectives from adverbs etc., e.g. Karaim

bundayi 'local' ← *bunda* 'here', Kumyk *burunyu* 'previous' ← *burun* 'long ago'.

Deverbal adjective markers include $-(X)K$, e.g. Trakai Karaim *ariχ* 'thin' ← *ari-* 'get tired', *suvuχ* 'cold' ← *suvu-* 'cool off', Kumyk *ačiq* 'open' ← *ač-* 'open', and $-X$, e.g. Trakai Karaim *öl'ü*, Halič Karaim *eli* 'dead' ← *öl'-*, *el-* 'die'.

Comparison

The comparative marker is $-rAK$, e.g. Trakai Karaim *yaχširaχ* 'better' ← *yaχši* 'good', Halič Karaim *k'ičirek* 'smaller' ← *k'iči* 'small'. The Karachay ending is $-(X)rAk$, e.g. *uzunuraq* 'longer' ← *uzun* 'long'.

Superlatives are formed with particles, e.g. Trakai Karaim *eηk uηlu* 'biggest' ← *uηlu* 'big', Karachay *em* or *eη alamat* 'prettiest' ← *alamat* 'pretty', Kumyk *čink de* or *lap da yuvuq* 'nearest' ← *yuvuq* 'near'. The particles are mostly used with comparatives, e.g. Karaim Trakai *eηk k'ič'ir'ak* 'smallest' ← *k'ič'i* 'small'. Karaim also uses particles of Slavic origin, e.g. *nay yaχširaχ* 'best' ← Polish *najlepszy*, or *sami yaχširaχ* 'best' ← Russian *samyj lučšij*.

Adjectives have reduplicated intensive forms such as Karachay *qara-qara* 'jet-black', *sap-sari* 'very yellow', Crimean Tatar *ap-aydin* 'very light', Trakai Karaim *yap-yaχši* 'extremely good'.

Declension

The plural suffix is $-lAr$. Its final $-r$ is lost in Karachay-Balkar except before possessive suffixes, e.g. *jolla* 'roads', *atlarim* 'my horses'. In Kumyk, it is lost before case markers, e.g. *yollaya* 'to the roads', *yollar* 'roads'. Karaim refers to paired body parts with the plural form, e.g. Trakai *k'öz*, *k'öz'l'är*, Halič *kez*, *kezler* 'eye, eyes', and frequently uses the plural after cardinal numbers, e.g. Trakai *üč kızlar* 'three girls'. This phenomenon is attributed to Slavic influence.

The case endings are nominative $-Ø$, e.g. *at* 'horse', *ata* 'father', accusative Karaim, Kumyk, Karachay $-nX$, Crimean Tatar $-nI$, e.g. *atni*, *atani*, genitive Crimean Tatar $-nIη$, e.g. *atniη*, *ataniη*, Balkar dialects $-nXη$, $-nXn$, e.g. *atniñ*, *ataniñ*, Kumyk, Karachay $-nX$, e.g. *atni*, *atani*, dative $-GA$, e.g. Trakai Karaim *atχa*, Halič Karaim *atka*, Trakai and Halič Karaim *ataya*, Kumyk, Karachay *atya*, *ataya*, Crimean Tatar *atqa*, *ataya*, locative $-DA$, e.g. Trakai Karaim, Crimean Tatar *atta*, *atada*, Kumyk, Karachay *atda*, *atada*, ablative $-DAn$, e.g. Karaim, Crimean Tatar *attan*, *atadan*, Kumyk, Karachay *atdan*, *atadan*, Karachay-Balkar $-nAn$ after *m*, *n*, *η*. There are also instrumental markers such as Karaim $-bA$, Karachay $-bIA$, written *bla*, and equative markers such as Karachay $-lAy$, $-čA$ (cf. p. 314).

The possessive suffixes are 1p.sg. Karaim, Kumyk, Karachay $-(X)m$, Crimean Tatar $-(I)m$, 2p.sg. Trakai Karaim $-(X)y$, Halič Karaim $-(X)n$, Karachay $-(X)η$, Crimean Tatar $-(I)η$, 3p.sg. Karaim, Kumyk, Karachay $-(s)X$, 1p.pl. Karaim $-(X)mXz$, Crimean Tatar $-(I)mIz$, Kumyk, Karachay $-(X)bXz$,

2p.pl. Trakai Karaim $-(X)yXz$, Halič Karaim $-(X)nXz$, Kumyk, Karachay $-(X)gXz$, Balkar dialects, Crimean Tatar $-(I)ηIz$, 3p.pl. Karachay $-lArI$, Kumyk $-(s)X$, Karaim $-(s)I$, Crimean Tatar $-(s)I$ ~ $-lArI$.

Some case endings deviate when preceded by first- and second-person singular possessive suffixes. The dative is $-A$, e.g. Karaim, Kumyk, Karachay, Crimean Tatar *atima* 'to my horse', Kumyk, Karachay, Crimean Tatar *atiηa*, Trakai Karaim *atiya*, Halič Karaim *atina* 'to your horse'. Due to simplification of *mn* and *ηη*, the Karachay accusative and genitive markers in these persons are $-X$, e.g. *atimi* and *atiηi*. After third-person suffixes, the Karaim, Karachay and Crimean Tatar accusative marker is $-n$, e.g. *babasin* 'his father'.

Pronouns

The first- and second-person personal pronouns are *men* 'I', *sen* 'you', *biz* 'we', *siz* 'you'. The third-person pronouns are: Karaim, Karachay *ol*, Kumyk *o* or *ol*, Crimean Tatar *o* 'he, she, it', Karaim *alar*, Kumyk, Crimean Tatar *olar*, Karachay *ala* 'they'. The forms *bizler*, *sizler* occur in Crimean Tatar. As a polite pronoun of address, Karaim uses an element of Hebrew origin with possessive endings: Trakai singular *koduy*, plural *koduyuz*, Halič *kanuz*.

Possessive pronouns are derived from the personal ones, e.g. Karaim *m'en'im* 'mine', *s'en'in* 'yours', *aniñ* 'its', *b'iz'(n')in* 'ours', *s'iz'(n')in* 'yours', *alarniñ* 'theirs'. The marker $-kl$ may be added, e.g. Kumyk *meniki* 'mine', *seniki* 'yours', *onuki* 'his', *bizinki* 'ours', *sizinki* 'yours', *olaniki* 'theirs'.

Reflexive pronouns are Trakai Karaim *öz*' and Kumyk *öz*, Crimean Tatar *öz*, Halič Karaim *ez* 'self' + possessive endings, e.g. Halič Karaim *ezim* 'myself', Trakai Karaim *öz'üy* 'yourself', Crimean Tatar *özüme* 'to myself'. The Karachay pronoun is *kes* with forms such as *kesim* 'myself', *kesiη* 'yourself'.

Demonstrative pronouns of near reference ('this') are Kumyk, Karachay, Crimean Tatar *bu*, Karaim *bu*, *bul*, Trakai Karaim *ušpu*, Halič Karaim *uſpu*, Kumyk *šu*. Colloquial Karaim *etol* results from fusion of Slavic *eto* and Turkic *ol*. Pronouns of distant reference ('that') are Kumyk, Crimean Tatar *o*, Karaim *ol*, Karachay *ol* (before vowels), *o* (before consonants), Trakai Karaim *ošol*, Halič Karaim *osol* 'that', Kumyk *šo*, Crimean Tatar *šu*.

Interrogative pronouns are Karaim, Kumyk, Karachay, Crimean Tatar *kim* 'who', *ne* 'what', Karaim *kaysi* 'which', Kumyk, Karachay, Crimean Tatar *qaysi* 'what kind, which'. Many may be used as relative pronouns. Derivatives of *ne* include Kumyk *nečik* 'which', *neče* 'how much', *nege* 'why', Karachay *neči* 'what occupation', *nelik* 'what is it for', *neli* 'provided with what'.

Indefinite pronouns are formed with interrogative pronouns + particles, e.g. Karaim *k'im-es* 'someone', *n'e-es* 'something', Kumyk *kimdir* 'someone', *nedir* 'something'. Karachay has pronouns formed with the old conditionals *ese* and *bolsa* + *da*, e.g. *kim ese da* 'anyone'. Another type

contains *bir* 'one', e.g. Kumyk *bir neče* 'some', Crimean Tatar *biri(si)*, *birev(i)* 'someone', *bir kimse* 'someone'.

Collective pronouns are quantifiers such as Karaim *bar*, Kumyk *bari* 'all', Trakai Karaim *b'üt'ün*, Kumyk *bütün*, Crimean Tatar *bütün* 'whole, all'. Karaim *yar*, Kumyk *har*, Karachay *yar*, Crimean Tatar *er* 'every' are of Persian origin.

Negative pronouns are formed by adding interrogative or indefinite pronouns to forms of *hič* 'nothing', e.g. Kumyk *heč kim* 'nobody', *heč-ne* 'nothing', Crimean Tatar *ič bir şey* 'nothing', *ič kimse*, *ič birisi*, 'nobody'. Karaim pronouns can be formed with *-DA*, e.g. *n'em'e-d'e* 'nothing', Trakai Karaim *k'iš'i-d'e* 'nobody'.

Numerals

The digit numerals exhibit the expected forms, e.g. Crimean Tatar *bir*, *eki*, *üç*, *dört*, *beş*, *altı*, *yedi*, *sekiz*, *doquz*, Karachay *bir*, *eki*, *üç*, *tört*, *beş*, *altı*, *jeti*, *segiz*, *toyuz*. Decades are normally expressed in the regular Turkic way, e.g. Crimean Tatar *on*, *yigirmi*, *otuz*, *qırq*, *elli*, *altmış*, *yetmiş*, *seksen*, *doqsan*, Karachay *on*, *jüyirma*, *otuz*, *qırq*, *elli*, *altmış*, *jetmiş*, *seksan*, *toqsan*. Karachay earlier used a multiplicative method based on 'ten', e.g. *eki on* 'twenty' (two × ten), *üç on* 'thirty' (three × ten), and a multiplicative-additive method based on 'twenty', e.g. *jüyirma bla on* 'thirty (twenty + ten)', *eki jüyirma bla on* 'fifty (two × twenty + ten)'.

The higher numbers have the expected forms, e.g. Trakai Karaim *yüz*, Karachay *jüz*, Crimean Tatar *yüz*, Karaim, Halič Karaim *iz* 'hundred', Karaim *m'in*, Karachay *miñ*, Crimean Tatar *biñ* 'thousand', Trakai Karaim *t'üm'an*, Halič Karaim *timen* 'ten thousand'.

Ordinal numbers are derived with Kumyk, Karachay, Trakai Karaim *-(X)ncX*, Crimean Tatar *-(X)njI*, Halič Karaim *-(X)ncX*, e.g. Kumyk *dörtünçü* 'fourth', Crimean Tatar *altınjı* 'sixth', Karachay *jüz bla jüyirma altınçı* 'one hundred and twenty-sixth'.

Collective numbers are derived with Trakai Karaim *-Ow*, Kumyk, Karachay, Crimean Tatar *-Aw*, Halič Karaim *-ow*, *-ew*, e.g. Trakai *ek'öv*, Halič Karaim *ekew* 'two together', Kumyk *onaw* 'ten together', Crimean Tatar *üçew* 'three together'. One variant adds the suffix *-IAn*, e.g. Trakai Karaim *üç'öv'l'an*, Halič Karaim *icevlen* 'three together', Karachay *altawlan* 'six together'.

Distributive numbers are formed with Trakai Karaim, Kumyk, Karachay, Crimean Tatar *-(š)Ar*, Halič Karaim *-(s)Ar*, e.g. Kumyk *segizer* 'eight at a time', Karachay *ekişer* 'two and two'. The ending may be expanded with *-IAP*, e.g. Kumyk *yettişerlep* 'seven at a time'.

Fractions are expressed by lexemes such as Karaim *yarım*, *yartı* 'half', Crimean Tatar *buçuq* 'half', Trakai Karaim *ç'er'ik*, Halič Karaim *cerik* 'fourth', e.g. Karaim *b'ir yarım* 'one and a half', Crimean Tatar *dört buçuq* 'four and a half'. Fractions may also be expressed with the denominator in the

ablative, genitive or locative plus the numerator, mostly with the possessive suffix, e.g. Kumyk, Karachay *ekiden bir(i)* 'one half', Karachay *törtnü ekisi* 'two quarters', *beşte üç* 'three fifths'.

Approximative numbers may be expressed with a preposed *bir*, e.g. Halič Karaim *bir bes adam* 'about five men', or juxtaposed cardinal numbers, e.g. Trakai Karaim *üç-d'ör't' k'iš'i* 'some three or four people'.

Copulas

The copula suffixes are: 1p.sg. Trakai Karaim *-m(In)*, Halič Karaim *-m(en)*, Kumyk *-mAn*, Karachay *-mA*, Crimean Tatar *-(I)m* or *-mAn*, 2p.sg. Crimean Tatar *-sIn* or *-sIn*, Trakai Karaim *-s(In)*, Halič Karaim *-s(en)*, Kumyk *-sAn*, Karachay *-sA*, 3p.sg. Kumyk, Crimean Tatar *-Ø*, Karaim *-DXr*, Karachay *-dX*, 1p.pl. Karaim *-(IAr)-bXz*, Kumyk, Karachay *-bXz*, Crimean Tatar *-(I)mlz*, 2p.pl. Karaim *-(IAr)-sXz*, Kumyk, Karachay *-sXz*, Crimean Tatar *-sInlz* or *-sIz*, 3p.pl. Kumyk, Crimean Tatar *-IAr*, Karaim *-DXr-IAr*, Karachay *-dXIA*. Examples: Kumyk *Yazıvčuman* 'I am a writer', *Iščisen* 'You are a worker', *Güçlüsüz* 'You are strong'.

Negation is expressed with elements such as Trakai Karaim *t'üv'ül*, Halič Karaim *d'iwil*, Kumyk *tügül*, Crimean Tatar *degil* '(it) is not'.

Verbs

Denominal Verb Derivation

The most frequent suffixes forming verbs from nominals are *-IA*, e.g. Karaim *kusokla-* 'cut' ← *kusok* 'piece', Karachay *jirila-* 'sing' ← *jir* 'song', *-A*, e.g. Kumyk *aşa-* 'eat' ← *aş* 'bread', Karachay *ata-* 'name' ← *at* 'name', *-Ar*, e.g. Kumyk, Karachay *ayar-* 'turn white' ← *aq* 'white', *göger-* 'become blue' ← *gök* 'blue', *-IAn*, e.g. Karaim *γuyalan-* 'to nest' ← *γuya* 'nest', Kumyk *paydalan-* 'utilise' ← *payda* 'profit', Karachay *awuzlan-* 'eat' ← *awuz* 'mouth', *-IAt*, e.g. Halič Karaim *ekşizlet-* 'orphan someone' ← *ekşiz* 'orphan', *-Ay*, e.g. Karaim *usay-* 'become wiser' ← *us* 'mind', *-(X)K*, e.g. Kumyk *yoluq-* 'meet' ← *yol* 'road', *birik-* 'unite' ← *bir* 'one'.

Compound verbs are derived from nominals with auxiliary verbs, e.g. Karaim *er'k' b'er-* 'empower' ← *er'k* 'strength' + *b'er-* 'give', Kumyk *yoq et-* 'annihilate' ← *yoq* 'non-existing' + *et-* 'make', *ač bol-* 'starve' ← *ač* 'hungry' + *bol* 'be'.

Deverbal Verb Derivation

Passives are formed with *-(X)l* after consonant stems, *-(X)n* after stems ending in vowels or *l*, e.g. Karaim *b'er'il-* 'be given' ← *b'er-* 'give', Karachay *alin-* 'be taken' ← *al-* 'buy'. Reflexives are formed by adding *-(X)n* to transitive verbs, e.g. Kumyk *giyin-* 'get dressed' ← *giy-* 'dress', Karachay *juwun-* 'wash oneself' ← *juw-* 'wash', Crimean Tatar *gizlen-* 'hide oneself' ← *gizle-* 'hide'. Reflexive and passive suffixes also produce middle verbs, e.g. *qiril-* 'break (intransitive), get broken'.

Causatives are formed with $-(X)t$, e.g. Trakai Karaim *aɣlat-* 'make understand' ← *aɣla-* 'understand', Kumyk *aʃat-* 'feed' ← *aʃa-* 'eat'. A second variant is $-DXr$, e.g. Trakai Karaim *eʃ'it't'ir-* 'let hear' ← *eʃ'it-* 'hear', Kumyk *qaldir-* 'keep' ← *qal-* 'stay', Crimean Tatar *yedir-* 'feed' ← *ye-* 'eat'. A third variant is $-Xr$, e.g. Trakai Karaim *kačir-* 'let escape' ← *kač-* 'escape', Kumyk *učur-* 'let fly' ← *uč-* 'fly'. The suffixes $-Ar$, $-GXz$, $-(X)z$ are unproductive and rare.

Verb forms denoting cooperative or reciprocal action are formed with $-(X)ʃ$, e.g. Kumyk *öbüʃ-* 'kiss each other' ← *öp-* 'kiss', Karachay *oquʃ-* 'read together' ← *oqu-* 'read', Halič Karaim *k'il'is-* 'laugh together' ← *k'il-* 'laugh' (with $s < ʃ$).

Frequency and intensity of an action is expressed with $-KAIA$, e.g. Trakai Karaim *öp'k'al'ä-*, Halič Karaim *epkele-* 'kiss all over or frequently' ← *öp-*, *ep-* 'kiss'.

The suffix of negation is $-mA$, e.g. *kelme-* 'not come'. Possibility is mostly expressed by the $-A$ converb plus the auxiliary *al-* 'take', e.g. Karachay *köre al-*, *köral-* 'be able to see', Karaim *baral-* < *bara al-* 'be able to go' ← *bar-* 'go', *aʃayal-* 'be able to eat' ← *aʃa-* 'eat'. Karaim has also developed the auxiliary verb *bolal-* < **bol-a al-* 'be able'. Crimean Tatar tends to use *bil-* 'know', e.g. *ata bil-* 'be able to throw'. Impossibility is expressed with the negative stems of *al-*, *bil-* and *bol-* (Crimean Tatar *ol-*), e.g. Karaim *baralma-* 'be unable to go', Karachay *köralma-* 'be unable to see', Crimean Tatar *atalma-*, *atama-* 'be unable to throw'.

Constructions of the type converb + postverbs (p. 42) are frequent. The auxiliaries, developed from verbs such as *al-* 'take', *ber-* 'give', *qoy-* 'put', *tur-* 'stand', *kel-*, *gel-* 'come', *bar-* 'go', *ket-* 'go', *qara-* 'watch', *qal-* 'stay', *yiber-* 'send' and *baʃla-* 'begin', express various kinds of actional modification: completion, suddenness, permanence, continuity, e.g. Kumyk *qačib ket-* 'suddenly run away', *sölešib bar-* 'keep talking', Karachay *baqip tur-* 'look'.

Finite Verb Forms

Many finite verb forms assume personal markers of predominantly pronominal origin, namely 1p.sg. Trakai Karaim $-m(In)$, Halič Karaim $-m(en)$, Karachay $-mA$, Kumyk $-mAn$, Crimean Tatar $-(X)m$, 2p.sg. Trakai $-s(In)$, Halič Karaim $-s(en)$, Karachay $-sA$, Kumyk $-sAn$, Crimean Tatar $-sIn$, 3p.sg. Karaim $-D(Xr)$, Karachay $-dX$, Kumyk, Crimean Tatar $-\emptyset$, 1p.pl. Karaim, Karachay $-bXz$, Kumyk $-blz$, Crimean Tatar $-mlz$, 2p.pl. Karaim, Karachay $-sXz$, Kumyk $-slz$, Crimean Tatar $-sl(\eta)z$, 3p.pl. Karaim $-DXrAr$, Karachay $-dXIA$, Kumyk, Crimean Tatar $-lAr$.

Other finite forms take markers of possessive origin, namely 1p.sg. Karaim, Kumyk, Karachay, Crimean Tatar $-m$, 2p.sg. Trakai Karaim $-y$, Halič Karaim $-n$, Kumyk, Karachay, Crimean Tatar $-\eta$, 3p.sg. Karaim, Kumyk, Karachay, Crimean Tatar $-\emptyset$, 1p.pl. Karaim, Kumyk, Karachay, Crimean Tatar

$-K$, 2p.pl. Trakai Karaim $-yXz$, Halič Karaim $-nXz$, Kumyk, Karachay $-GXz$, Crimean Tatar $-\eta/z$, 3p.pl. Karaim, Kumyk $-lAr$, Karachay $-lA$, Crimean Tatar $-(lAr)$.

A general intraterminal present is formed with the $-A$ converb + pronominal personal markers, e.g. Karachay *Oquyma* 'I read', Kumyk *Oxuyusan* 'You read', Karaim *Olturam* 'I sit', Kumyk *Getebiz* 'We go', Crimean Tatar *Isteysi(ŋi)z* 'You want', Karachay *Baradila* 'They go'. It is not highly focal in the sense that it refers to events limited to the time of speech, e.g. Kumyk *Men madrasada oxuyman* 'I learn at school'. The interrogative $-mX$ or $-ml$ precedes the personal markers, but may follow them in colloquial speech, e.g. Kumyk *Baraman-mi?* 'Am I going?'

All languages except Karaim also use a more focal present to express events currently taking place. It is mostly formed with the $-A$ converb + the auxiliary verb *tur-* 'stand' + pronominal personal markers, e.g. Kumyk *Bu kitabni oxuy turaman* 'I am reading this book', Karachay *Işley turmayma* 'I am not working'. Crimean Tatar uses the $-A$ or $-(I)p$ converb + aorist of *yat-* 'lie' + pronominal personal markers, e.g. *Ala yatırım, Alıp yatırım* 'I am buying'.

The aorist expresses tendency or willingness and is thus also used prospectively to refer to possible future events, often expressing uncertainty about their accomplishment, e.g. Karachay *Alırma* 'I will (probably) take'. Its meaning in Kumyk and Crimean Tatar is not always distinct from that of the general present. It is formed with $-(V)r$ + pronominal personal markers, the suffix vowel being A , I or X , e.g. Kumyk *Bararman* 'I will go', *Oxurman* 'I will read', Crimean Tatar *Alırsıñ* 'You will take'. Karaim colloquial forms lose the r in non-third persons, e.g. *alim, alıs, alir, alibiz, alısız, alırlar* ← *al-* 'take'. Negation is formed with $-mAs$ in Karaim, Kumyk, Karachay, and with $-mAz$ in Crimean Tatar, e.g. Kumyk *Barmasman* 'I will not go'. Karaim $-mAs$ often loses its $-s$ in the first and second persons, e.g. *Almam* 'I will not take', *Almas* 'You will not take'. In Crimean Tatar, z disappears in the first-person singular, e.g. *Almam* 'I will not take'.

Kumyk and Crimean Tatar form prospective items in $-(A)žAK$ and $-AJAK$ + pronominal personal markers, e.g. Kumyk *Getežekmen* 'I will (definitely) go', *Oxužaqsan* 'You will read'. Crimean Tatar $-AJAK$ is added to consonant stems, $-yžAK$ to vowel stems, e.g. *Alajayım* 'I will take', *Dinleyjegim* 'I will listen'. Negation is formed regularly, e.g. Kumyk *Barmažaqsan* 'I will (definitely) not go'. The Karachay equivalent is formed with $-(V)rXK$ with the variants $-rXK$, $-nXK$, $-lXK$, e.g. *Qallıqsa* 'You will stay', negated with *tüyül* '(is) not', e.g. *Qallıq tüyülme* 'I will not stay'. Unlike the aorist, the prospective only refers to future events that are foreseen to occur.

The simple past is formed with $-DX$ + personal markers of possessive origin, e.g. Trakai Karaim *Aldiy*, 'You took', Kumyk *Oxuduŋ* 'You read', Karachay *Aytdım* 'I said'. The Crimean Tatar suffix vowel is I , e.g. *Tüşürdim* 'I dropped'. Negation is formed regularly, e.g. Kumyk *Oxumadıım* 'I did not

read'. The interrogative *-mX* or *-ml* follows the personal markers, e.g. Kumyk *Bardilarmi?* 'Did they go?', but there are exceptions in colloquial speech.

A post-terminal past with perfect functions is formed in Kumyk, Karachay and Crimean Tatar with *-GAn* + pronominal personal markers, e.g. Kumyk *Baryanman* 'I have gone', Karachay *Aytyansa* 'You have said', Crimean Tatar *Alyanmiz* 'We have taken it'. Negation is formed regularly, e.g. Kumyk *Barmayanman* 'I have not gone'. The interrogative particles follow the personal markers. This item is used to express past events of current relevance, often on the basis of its result or indirect evidence (indirective use).

The combination *-(V)r* + past copula particle is a non-focal imperfect that mostly describes habitual events typical of a certain period in the past in an intraterminal perspective, without information on their beginning or end, e.g. Trakai Karaim *Alir edim*, Halič Karaim *Alar ed'im*, Crimean Tatar *Alir edim* 'I used to take'. The item may also express counterfactuality. Contractions occur in colloquial speech, e.g. Kumyk *Barardim* < *barar edim*. The negation is formed with the negative aorist suffix, e.g. Kumyk *Barmas edim*, colloquial *Barmasidim* 'I did not use to go'.

A more focal imperfect, formed with the *-A* converb + past copula, is found in Kumyk and Crimean Tatar. It indicates single or repeated events envisaged intraterminally at some past point in time, e.g. Kumyk *Bara edim* 'I was going', Crimean Tatar *Ala edik* 'We were taking'. Contracted forms occur in colloquial speech, e.g. Kumyk *Baradim*. Negation is regular, e.g. Crimean Tatar *Almay edim* 'I was not buying'. A similar Crimean Tatar item is formed with the *-A* converb + aorist of *yat-* + past copula, e.g. *Ala yatir edim* 'I was taking'.

A pluperfect is formed with *-GAn* + past copula, e.g. Karaim, Karachay, Crimean Tatar *Alyan edim* 'I had taken', Kumyk *Tilegen edim* 'I had asked'. Negation is regular, e.g. Crimean Tatar *Almayan edim* 'I had not taken'. Crimean Tatar has another past tense reserved for relatively remote events, formed with *-DI* + possessive personal markers + past copula, e.g. *Aldiq edi* 'We took (long ago)'. Karachay-Balkar forms in *-XwčAn* are used to express habitual events, e.g. *Jaziwčanma* 'I often write', *Jaziwčan edim* 'I used to write'.

Optative-imperatives take on the personal markers 1p.sg. *-(A)yIm*, 2p.sg. *-θ* or *-KXn*, 3p.sg. Karaim, Kumyk, Karachay *-sXn*, Crimean Tatar *-sIn*, 1p.pl. *-(A)yIK*, 2p.pl. Trakai Karaim *-(X)yXz*, Halič Karaim *-(X)nXz*, Kumyk, Karachay *-(X)GXz*, Crimean Tatar *-(I)ηIz*, 3p.pl. Karaim, Kumyk *-sXnlAr*, Karachay *-sXnlA*, Crimean Tatar *-sInlAr*. Crimean Tatar examples are *alayim*, *al* or *alyin*, *alsin*, *alayiq*, *alijiz*, *alsinlar* ← *al-* 'take'. Particles expressing encouragement etc. may be added, e.g. Kumyk, Crimean Tatar *-čI*, *-sA*, *-sAnA*. Negation is formed regularly, e.g. Kumyk *Barmayim* 'I shall not go'. Interrogative markers are added to the personal ones, e.g. Kumyk *Alayimmi?* 'Shall I buy?'

A present optative is formed with *-Gay* + pronominal personal markers, e.g. Trakai Karaim *Baryeys(in)*, Halič Karaim *Baryays(en)* 'I wish you would go'. A past item is formed with the past copula particle, e.g. Trakai Karaim

Baryed'im 'I wish I had gone'. The Karachay equivalent may refer to both present and past, e.g. *Tabxay edim* 'I wish I would find/had found'. Contractions such as *alyay edim* > *alyaydim* are found in all languages.

Non-finite Verb Forms

Intraterminal (present) participles are formed in various ways, in Karaim with the *-A* converb + *doyon*, e.g. *baradoyon* 'going', in Kumyk with *-A* + *γAn*, e.g. *barayan* 'going', *yilayyan* 'crying', in Karachay with *-(X)wčX*, e.g. *tintiwčü* 'studying'. Post-terminal ('past') participles are formed with *-Gan*, e.g. Karaim *baryan* 'having gone', Halič Karaim *bergen* 'having given', Kumyk *gelgen* 'having come'.

Prospective (future) participles are formed with Kumyk and Crimean Tatar *-(A)jAK*, e.g. Kumyk *gelejek* 'coming, foreseen to come'. The aorist participle in *-(V)r* expresses qualities, often projected on the future, e.g. Crimean Tatar *aqar (suw)* 'running (water)', Kumyk, Karachay *geler* 'who will come', *görüür* 'who will see'. Karachay also uses *-(V)rXK*, e.g. *qarariq* 'who will watch'.

Among the converbs, an intraterminal converb is formed with *-A* after consonants and *-y* after vowels, e.g. Karaim, Karachay *ayta* 'saying', Kumyk *işley işley* 'working'. Karaim exhibits an expanded form, e.g. Trakai *k'ül'adoyoč*, Halič *k'iledoyoč* 'laughing'. Crimean Tatar has a similar item in *-(y)ArAK*, e.g. *külerek* 'laughing'. Converbs in *-(X)p* or *-(X)b* are present in all languages, e.g. Karachay *aytib* 'saying and ...', Kumyk *oxup* 'reading and ...'. A negative converb is formed with *-mAYIn* and *-mAY*, e.g. Karachay-Balkar *körmey* 'without seeing'. Karaim also has the extended form *-mAYInčA*, e.g. Trakai *aytmayinča*, Halič *aytmayinca* 'without saying'.

Kumyk and Karachay-Balkar *-GANll* expresses 'since', e.g. *eşitgenli* 'since ... heard', Karaim *-GAčok* 'as soon as', e.g. Trakai *čixkačox* 'as soon as ... left', Kumyk and Karaim *-GXnčA*, Karachay-Balkar *-GxnčX* 'until, before', e.g. Trakai Karaim *čixkinča* 'until ... comes out', Kumyk *görgünče* 'until ... sees it'.

Conditionals are formed with *-sA* + possessive personal markers, e.g. Karaim *aytsam* 'if I speak', Kumyk *oxusaŋ* 'if you read', Karachay *tabsaq* 'if we find/would find', Crimean Tatar *alsam* 'if I take/had taken'. Past conditionals are formed with *-sA* + past copula, e.g. Crimean Tatar *alsa edim* 'if I had bought'. Contractions such as *alsaydim* are common in spoken language. Karachay has *-sA* + personal marker + past copula, e.g. *tabsaq edi* 'if we had found', Kumyk *-GAn* + *bolsa* + personal markers, e.g. *baryan bolsam* 'if I had gone'.

Various verbal nouns called 'infinitives' are formed with *-mAK*, *-mA*, *-(X)w* and *-(V)rGA*, e.g. Kumyk *oxumaq* 'read'. Purposive meaning is often expressed by *-mA* < *-mAYA* and *-(V)rGA*, e.g. Halič Karaim *Kolabiz k'elmä* 'We ask ... to come', Kumyk *getme* 'in order to go', Karachay *qararya* 'to watch'.

Adverbs

Frequently used adverb suffixes are *-lay*, *-layIn* and *-CA*, e.g. Trakai Karaim *yalinley* 'like a flame', Kumyk *šulay*, *bulay*, *olay* 'like this', *issileyin* 'warmly', Crimean Tatar *aqşamlayin* 'in the evening', Trakai Karaim *karayče* 'in Karaim', Crimean Tatar *menje* 'according to me'.

Nominals with fossilised case endings are often found, e.g. in Trakai Karaim *kišin* 'in winter', Halič Karaim *ertenbe* 'in the morning', *artqari* 'backwards', Crimean Tatar *açıqtan* 'clearly'.

Other Morphological Categories

Postpositions governing the nominative include Trakai Karaim *üč'ün* 'for', Halič Karaim *icin*, Karaim *bila* 'with', *k'ib'ik* 'as, like', Crimean Tatar *kibi* 'as, like'. The dative is used with Trakai Karaim *k'ör'a*, Halič Karaim *k'ere*, Kumyk *göre* 'considering', Karaim *d'eyin*, Karachay *deri*, Crimean Tatar *qadar* 'up to', Karaim *utru* 'in front of', etc. The ablative is governed by Karaim *son*, Kumyk *soŋ* 'after', Kumyk *bašya* 'other, except'. Sets of auxiliary nouns, e.g. Karaim, Karachay *art* 'back', Karaim *ara* 'middle', Karachay *jan* 'side', are used to construct complex postpositions with possessive and case endings. Karaim has prepositions of Slavic origin such as *do* 'up to' and *po* 'on' used mostly in stereotyped expressions.

Conjunctions include connective, disjunctive, adversative, conditional, causal and purposive ones such as Karaim, Kumyk, Karachay *da* 'and', Karaim *γ'em* 'and, also', Karachay *em*, *emda* 'and', Kumyk *wa* 'and', Karachay *bla* 'and', Kumyk *bulan* 'and', Crimean Tatar *ya ... ya* 'either ... or', Halič Karaim *yemese* 'or', Kumyk *ne ... ne* 'neither ... nor', Karaim *t'ek* 'but', Trakai Karaim *ančeχ*, Halič Karaim *ancaχ* 'but', Kumyk *amma*, *tek* 'but', Crimean Tatar *läkin* 'but', Trakai Karaim *eg'är*, Kumyk, Crimean Tatar *eger* 'if', Karaim *ki*, *bunar ki* 'in order to', Crimean Tatar *čünki* 'because'. Colloquial Karaim has conjunctions of Slavic origin, e.g. *a*, *to*, *a to* 'but' < Russian, *vale* 'but' < Polish. Interrogative pronouns and derivatives such as Karachay *qačan* 'when', *qalay* 'how', *kim* 'who', *qallay* 'what kind' may also serve as conjunctions.

The particles are of various kinds: interrogative, e.g. Kumyk, Karachay *mX*, restrictive, e.g. Karaim *KInA* 'only', evidential, e.g. Kumyk *buyay*, *eken* 'apparently'. The colloquial Karaim *že* 'just' is of Slavic origin. A variety of interjections express a wide range of emotions.

Syntax

Nominal Phrases

In nominal phrases, attributes precede the heads, e.g. Karaim *s'eg'iz* 'adam' 'eight men', *alyan b'it'ik* 'the received letter', Karachay *taš köpür* 'stone bridge', *ullu adam* 'big man', Crimean Tatar *baqir tenjere* 'copper pan'. Karaim sometimes deviates from this common Turkic rule, e.g. *xiyar*

karayski 'Karaim cucumber'. It may even exhibit number agreement between attribute and head, e.g. *suklančılar baçurlar* 'wonderful young men'.

Compounds of the possessive type are common, e.g. Trakai Karaim *yaz baši* 'spring' ← *yaz* 'summer', *baş* 'beginning', Crimean Tatar *kireč kesegi* 'piece of lime'. The possessive suffix is absent in cases such as Trakai Karaim *üb'iy* 'host' ← *üw* 'house', *b'iy* 'lord', Kumyk *qolγap*, Karachay *qolqab* 'glove' ← *qol* 'hand', *qab* 'sack', Karachay *taw jol* 'mountain road'.

Nominal phrases may also consist of coordinative compounds, e.g. Kumyk *qatın-qız* 'women' ← *qatın* 'woman' + *qız* 'girl', Karachay *satiw-aliw* 'trade' ← *satiw* 'selling' + *aliw* 'purchase', *oq-tob* 'weaponry' ← *oq* 'bullet' + *tob* 'gun'.

Genitive constructions are formed according to normal Turkic rules, e.g. Crimean Tatar *adamniŋ degeri* 'dignity of man', Karachay *Kavkazni xawasi* 'the air of the Caucasus'.

Due to Slavic influence, Karaim often displays inverted word order, e.g. *yiri mamanin* 'mama's song'. More frequently than in other Turkic languages, the possessive suffix may be omitted if the possessor is a personal or demonstrative pronoun, e.g. *b'iz'n'in* 'at our horse'. This is not an exclusively Karaim tendency; for example, Crimean Tatar *siziŋ köy* 'your village' is also possible.

The Sentence

Sentences are formed according to the usual Turkic patterns, with the exception of some aberrant Karaim structures. A widely used Karaim word order is subject, predicate, object, e.g. Trakai Karaim *Ol aytıı bunı maya* '(S)he told me this'.

Words expressing negation are Karaim *yoxt* and *yo* '(there) is not', Kumyk *yoq* etc. Lexical elements expressing modality include Trakai Karaim *k'er'äk*, *t'iyiŝ'l'i*, Kumyk *gerek*, *tariq* 'necessary', Karaim *može* 'possible', *moget* 'can' (< Slavic).

Complex sentences are formed through coordination of clauses with or without conjunctions, e.g. Karachay *Biz Dnepr suwuna bardıq emda anda bizge bir taniŝ adam edi* 'We went to the River Dnieper and met an acquaintance there', *Taŋ atdı, iŝ baŝlandı* 'Morning came, the work began'. Certain clause combinations with or without conjunctions resemble Indo-European subordinations, e.g. Crimean Tatar *Endi ne olir, bilmem* 'I do not know now what will happen', Kumyk *Issilik olay güçlü çü, hatta adam suwdan çıqmaya süymey* 'The heat is so strong that one does not want to come out of the water'.

Fully embedded clauses are, however, formed with non-finite verb forms such as participles and converbs, e.g. Kumyk *Men qaytyanli beŝ ay bitdi* 'Five months have elapsed since I returned', *Kim köp oxuy busa ol köp bile* 'He who reads much knows much'. Participles may take possessive and case endings or postpositions, e.g. Trakai Karaim *k'el'g'an'im'd'a* 'when I came', *k'el'g'an'im'd'an son* 'after I arrived'.

Lexicon

The vocabulary of all West Kipchak languages has been strongly influenced by foreign languages. The most profound effects on Karaim vocabulary have been exerted by Hebrew, Polish, Ukrainian, Russian and Lithuanian. Kumyk has been influenced by Caucasian and Russian. In Karachay and Balkar, old Alan and more recent Ossetic and Russian influences can be traced. The strongest influence on Crimean Tatar comes from Oghuz Turkic, Persian, Arabic (through Persian mediation), and Russian.

Dialects

Certain differences between the Karaim dialects probably developed before the migration from the Crimea. The main differences between Trakai and Halič Karaim are of a phonetic nature, e.g. Trakai *ö, ü* = Halič *e, i*, Trakai *ä* = Halič *e*, Trakai *i* = Halič *é*, Trakai *š, ž, č, j* = Halič *s, z, c, dz*, Trakai *-χ* = Halič *-k*, Trakai *y < *ŋ* = Halič *n*. Only the Trakai dialect has the whole set of palatalised consonants.

Of four rather similar Balkar dialects, the Baχsan dialect exhibits **y- > j-*, preserved *č-*, *é > e* after *ö* and *ü*, plural in *-lar*, negative converb in *-map*, etc. The Čegem dialect has **y- > j-*, preserved *č-*, plural in *-la*, genitive in *-l* after 1p.sg. possessive suffixes, e.g. *atamni* 'my father's' > *atami*. The rather mixed Xulam-Bezinga dialect has **y- > j- ~ ž-*, **č- > c-*, plural in *-lar ~ -la*, fusion of the genitive marker with 1p.sg. possessive suffixes, e.g. *atamni* > *atamni*. The Malqar dialect has **y- > z- ~ j- ~ ž-*, **č- > c-*, plural in *-la*, and rounded vs. unrounded harmony of low suffix vowels.

Kumyk may be divided into the Buynaq, Xasavyurt and Qaytaq dialects, of which the first two form the basis of the written language.

Crimean Tatar has a northern 'steppe' dialect, a southern dialect and a central dialect, *orta yolaq* 'middle lane', the basis of the written language. One criterion is the distribution of initial **y-*, namely *y-* in the south, *j-* in the north, and *y ~ j* in the centre. Initial **t-* is represented as *d-* in the south, as *t-* in the north and as *d- ~ t-* in the central area. Tendencies towards rounded vs. unrounded harmony are strongest in the south, and weakest in the north. The dative forms of personal pronouns are *mana ~ mā, sana ~ sā, una ~ uya* in the south, *maya ~ mā, saya ~ sā, oya* in the north, whereas the central dialect has *maŋa, saŋa, oŋa*.

Further Reading

- Asanov, Š. A. *et al.* (eds) (1988) *Krymskotatarsko-russkij slovar'*, Kiev: Radjans'kaja škola.
- Bammatov, Z. Z. (ed.) (1969) *Kumyksko-russkij slovar'*, Moscow: Sovetskaja Ėnciklopedija.
- Baskakov, N. A., Zajončkovskij, A. and Šapšal, S. M. (eds) (1974) *Karaimsko-russko-pol'skij slovar'*, Moscow: Russkij jazyk.
- Benzing, J. (1959) 'Das Kumükische', in J. Deny *et al.* (eds) *Philologiae Turcicae*

- Fundamenta I*, Wiesbaden: Steiner, 391–406.
- Çeneli, I. and Gruber, E. A. (eds) (1980) *Krimtatarische Chrestomathie aus Gegenwartstexten*, Frankfurter Turkologische Arbeitsmittel 2, Wiesbaden: Harrassowitz.
- Csató, É. Á. (forthcoming) 'Some typological features of the viewpoint aspect and tense system in spoken North-Western Karaim', in Ö. Dahl (ed.) *Tense and Aspect in the Languages of Europe*, Berlin: Mouton De Gruyter.
- (forthcoming) *Spoken Karaim*, Turcologica, Wiesbaden: Harrassowitz.
- Doerfer, G. (1959) 'Das Krimtatarische', in J. Deny *et al.* (eds) *Philologiae Turcicae Fundamenta I*, Wiesbaden: Steiner, 369–90.
- Musaev, K. M. (1964) *Grammatika karaimskogo jazyka, Fonetika i morfologija*, Moscow: Nauka.
- (1966) 'Karaimskij jazyk', in N. A. Baskakov *et al.* (eds) *Jazyki narodov SSSR 2*, Moscow: Nauka, 260–79.
- (1975) *Leksika tjurkskix jazykov v sravnitel'nom osveščanii. Zapadno-kypčakskaja grupa*, Moscow: Nauka.
- Magomedov, A. G. (1966) 'Kumyjskij jazyk', in N. A. Baskakov *et al.* (eds) *Jazyki narodov SSSR 2*, Moscow: Nauka, 194–212.
- Pritsak, O. (1959a) 'Das Karaimische', in J. Deny *et al.* (eds) *Philologiae Turcicae Fundamenta I*, Wiesbaden: Steiner, 318–40.
- (1959b) 'Das Karatschaische und Balkarische', in J. Deny, *et al.* (eds) *Philologiae Turcicae Fundamenta I*, Wiesbaden: Steiner, 340–68.
- Sevortjan, Ė. V. (1966) 'Krymskotatarskij jazyk', in N. A. Baskakov *et al.* (eds) *Jazyki narodov SSSR 2*, Moscow: Nauka, 234–59.
- Tenišev, Ė. R. and Sujunčev, X. I. (eds) (1989) *Karačaevo-balkarsko-russkij slovar'*, Moscow: Russkij jazyk.
- Xabičev, M. A. (1966) 'Karačaevo-balkarskij jazyk', in N. A. Baskakov *et al.* (eds) *Jazyki narodov SSSR 2*, Moscow: Nauka, 213–33.

19 *Kazakh and Karakalpak*

Mark Kirchner

Introduction

Kazakh and Karakalpak belong to the South Kipchak group of Turkic languages. They are so closely related to each other that most of the characteristics to be mentioned in the following survey of Kazakh will also be valid for Karakalpak. Some distinctive features of Karakalpak will, however, be dealt with in the last section.

In the nineteenth and early twentieth centuries, there was some confusion about the name of the Kazakh language. It was called Kazakh-Kirghiz or Kirghis-Kaisak, whereas present-day Kirghiz was referred to as Kara-Kirghiz. The confusion partly originated in similarities between Kazakh and Kirghiz. Speakers of Kazakh and Kirghiz can communicate easily with each other in their mother tongues. This fact is, however, due to long-standing and intensive contacts rather than a very close genetic relationship.

Unlike other major Turkic languages such as Turkish, Uzbek and Tatar, Kazakh was not a written language in the pre-Russian period. As the Kazakh economy was mainly based on cattle breeding in the vast steppes, nomadism or semi-nomadism was the most appropriate way of life. Social norms and culture were preserved in oral form. Treaties, contracts and other documents were written in Tatar, Chaghatay or non-Turkic languages. In the nineteenth century, Kazakh was established as a written language under the influence of the Russian administration and its Tatar helpers. It was thus natural that the new literary language was based on the Kazakh dialects of the northeastern regions, where Russian and Tatar influence was strong.

Kazakh written in Arabic script was first used mainly for religious and pedagogical texts translated from oriental languages and for Kazakh supplements in some local newspapers published during the last decades of the nineteenth century. Although Abay Qunanbayuli (1845–1904) is considered to be the founder of the Kazakh literary language and the first writer among his people, his work was not published until the beginning of the twentieth century.

In 1989, more than 90 per cent of the Kazakhs living in the Soviet Union stated Kazakh to be their first language. Since then, Kazakh has been declared the official language of the republic. Nevertheless, Russian is still predominant in many spheres of life. Nearly 50 per cent of the Kazakhs do not have sufficient command of their mother tongue. On the other hand, only very few Russians and members of other minorities speak Kazakh.

The Russian influence upon spoken and written Kazakh is still very strong. Since the areas of Kazakh settlement are separated from the North Kipchak regions, the contacts with Tatar and Bashkir have been interrupted for a long time. The contacts with Turkic languages in the west, e.g. Turkmen and Azerbaijanian, are limited due to natural barriers. The contacts with Kirghiz are still maintained. The southern part of Kazakhstan constitutes a diffuse zone of contact with other Turkic languages, in which Kazakh and Uzbek settlements merge.

Sound System

Phonetics and Phonology

Kazakh has nine short vowel phonemes. Length is only found at a subphonemic level. As in most Turkic languages, there are four back phonemes /a/, /i/, /o/, /u/. Besides the corresponding front phonemes /e/, /i/, /ö/, /ü/, there is a fifth front phoneme /æ/, lower than the mid-high /e/. The opposition /æ/ vs. /e/ does not, however, reflect the supposed distinction between open and closed *e* in Old Turkic. Kazakh /æ/, which is restricted to the first syllable, is an innovation caused by secondary fronting of *a* in words copied from Arabic and Persian, e.g. *ædebiyet* 'literature'. Nowadays /æ/ is also found in native words, e.g. *bæri* 'all', and is thus an integral part of the vowel system. While /e/ occurs in all positions, it is higher in first syllables than in others. Word-initial /e/ often exhibits a prothetic *y*-, e.g. *yeki* 'two'. This phenomenon is not a result of Russian phonetic influence, since it is also attested in the Kazakh dialects of China, but it is reinforced in the speech of Russian–Kazakh bilinguals. Word-initial /ö/ and /o/ can also be realised with a prothetic bilabial glide. The vowel /i/ is usually shorter than its Turkish counterpart. Reduction of high vowels is very common. Though the realisations of /a/ are rather stable, it may be fronted in contact with /y/ or rounded before /l/.

There are the following consonants:

Table 19.1 Consonants

	Labial	Alveodental	Alveopalatal	Velar	Glottal
stop	b, p	d, t		g, k, q	
fricative	f	z, s	ʒ, ʃ	ɣ, χ	h
nasal	m	n		ŋ	
glide	w		y		
liquid		r, l, ʎ			

The Kazakh consonant system is more different from the Turkic norm than the vowel system. Some oppositions are only weakly established. The opposition /b/ vs. /p/ is neutralised in initial and final position, whereas the oppositions /d/ vs. /t/ and /g/ vs. /k/ are neutralised in final position. Between vowels, /b/ may be realised as a fricative. The voiceless fricative /f/ is restricted to loanwords, where it is often replaced by /p/, e.g. *aftobus* ~ *aptobus* 'bus'. As in most other Turkic languages, /g/ displays the variant *g* in front environments and *ɣ* in back ones. The distribution of *k* and *q* as variants of /k/ is analogous. The variant *χ* mainly occurs instead of *q* after /a/, e.g. *žaχsi* 'good', but is not reflected in the orthography. Of the phonemes /ʒ/ and /ɣ/, the former is almost restricted to word-initial positions, and the latter to non-initial positions, e.g. *žer* 'land', *ay* 'moon'. In certain positions there is a *č* occupying the former place of *š* (which has developed into *s*), e.g. *češe* 'grandmother', written *weure*. Though /l/ has the variants *l* and *ʎ* depending on the frontness and backness of the syllable, it is not realised as *ʎ* in direct contact with front consonants. Bilabial /w/, mostly developed from *ɣ*, fuses with a preceding /u/ or /ü/ into a long vowel.

As in most Turkic languages, accent is characterised by a rising tone that normally is on the last syllable. This pitch accent is different from the stress accent, which normally occurs on the first syllable. Exceptions include expressive words, onomatopoeias, imperatives and interrogative pronouns. Certain suffixes and enclitic particles cannot be accented, e.g. copulas, the negation suffix *-MA*, the adverbial marker *-ša*, the equative marker *-DAy*, the interrogative particle *-MA*, the instrumental *-Men(en)*, the affirmative *Koy*, the restrictive *Kana*, and the emphatic *-šI*. Accent may thus be distinctive, e.g. *Oquwšimiz* 'We are students', with an unaccentable copula, vs. *oquwšimiz* 'our student', with an accentable possessive suffix.

Phonotactics and Morphophonology

Kazakh has different types of sound harmony that affect native lexemes and most suffixes. The front vs. back harmony clearly affects consonants such as /k/, /g/ and /l/ to the effect that their articulation shifts between front and back positions, e.g. /karga/ *qarya* 'into the snow', /kölge/ *kölge* 'into the lake'. As for the rounded vs. unrounded harmony, the strength of rounding decreases with the distance from the first syllable. Both low

and high vowels are affected, but the range of rounding is larger with high vowels, e.g. *üydö* 'in the house', *üyümüzde* 'in our house'. In suffixes containing a low vowel such as *-DA*, the phonetic distribution of vowels is asymmetric, since *e*, *ö*, *a* are admitted, whereas *o* is not, e.g. *qolda* (not **qoldo*) 'in the hand'.

Kazakh orthography does not reflect these kinds of harmonic variation adequately, but offers an archiphonemic representation. Low suffix vowels are only represented by <a> and <e>, high suffix vowels only by <i> and <ü>. The rounded vs. unrounded harmony is thus not reflected at all, e.g. *үйміз* for *üyümüz* 'our house'.

In most Turkic languages, the distribution of suffix-initial consonants depends on whether the stem-final sound is voiced or voiceless. The combinability of consonants is more restricted in Kazakh. A suffix such as the ablative in *-DAn* also has an allomorph *-nAn* that occurs after stem-final nasals, e.g. *adamnan* 'from the man', probably due to the final *n* of the suffix.

There are several other progressive assimilations and dissimilations by which suffix-initial nasals and laterals are adapted to the preceding stem. The affected suffixes can be divided into three major classes:

Suffixes with initial *M = m, b, p* such as the negation suffix *-MA*, the enclitic interrogative particle *-MA* and the enclitic instrumental marker *-Men(en)*, e.g. *Keldi me?* 'Did (s)he come?', *Keldiñ be?* 'Did you come?', *Keldik pe?* 'Did we come?'

Suffixes with initial *L = l, d, t* such as the denominal verb suffix *-LA* and the plural marker *-Lar*, e.g. *balalar* 'children', *közder* 'eyes', *qazaqtar* 'Kazakhs'.

Suffixes with initial *N = n, d, t* such as the accusative and genitive markers *-NI* and *-NIñ*, e.g. *balanıñ* 'of the child', *közdiñ* 'of the eye', *qazaqtıñ* 'of the Kazakh'.

The diachronically primary variants *m, l, n* occur after stem-final vowels, and sometimes also after nasals, sonorants and glides. In the remaining cases, the corresponding stops *b* and *d* occur, assimilated to *p* and *t* after voiceless consonants.

Sandhi phenomena are widespread, though not reflected in orthography. The opposition /k/ vs. /g/ is neutralised word-initially and represented by /k/, but the voiced counterpart /g/ is used when the preceding word ends with a vowel, e.g. *eki gözü* 'his two eyes', written *eki көзі*. In sandhi, initial *ž-* may shift to *j-* after laterals and nasals, e.g. *eñ jaχsi* 'best'.

Orthography

A modified Arabic script was used up to the end of the 1920s, when it was replaced by the unified Latin alphabet, which was abandoned in 1940 in favour of the present-day Cyrillic alphabet. The Arabic script is still used by the Kazakhs of China. Recent Turkish attempts to promote a unified Roman script for all Turkic peoples have not been welcomed by the Kazakh government.

For the letters of the Cyrillic alphabet and their transliteration, see p. 454. Russian loanwords are written according to their original orthography. Cyrillic *и* and *у* represent diphthongoids which can be analysed phonologically as vowel + consonant, /iy/, /iy/ and /uw/, /üw/, respectively. The standard pronunciation of Cyrillic *у* is *ū* in back environments, e.g. алы = *alū* 'take', and *ū* in front environments, e.g. келу = *kelū* 'come'. Cyrillic *и*, however, is mostly pronounced as *ī*.

The consonants *g*, *γ*, *k* and *q* are all represented in the Kazakh alphabet, possibly due to Arabic graphic traditions. The Cyrillic letters *x* and *h* are marginal graphemes that do not reflect distinct sounds of the standard language.

Morphology

Nouns and Adjectives

Word Formation

Kazakh is similar to other Kipchak languages in the way it derives nominal stems from other nominal stems. The few differences are easily explained in terms of historical phonology and the rules of Kazakh morphophonology.

For example, *-šI*, corresponding to *-čI* in most other Turkic languages, denotes persons usually occupied with the denotee of the preceding stem, e.g. *žumisši* 'worker' ← *žumis* 'work', *-LIK*, added to nominal stems, forms nouns and adjectives, e.g. *qalaliq baq* 'municipal park' (copied from Russian *gorodskoj sad*), *-LI* produces adjectives from nouns, e.g. *muđdi* 'sad' ← *muğ* 'sadness', and *-LAs*, corresponding to *-DAš* in many other languages, denotes fellowship, e.g. *žerles* 'countryman' ← *žer* 'land'. The most common diminutive suffix is *-šA*, e.g. *kitapša* 'booklet' ← *kitap* 'book'.

Comparison

-(I)rAK can be regarded as a comparative suffix, e.g. *ülkenirek* 'bigger'. Superlative meaning can be expressed by several particles, primarily *eğ* and *asa*, e.g. *eğ kišisi* 'the youngest one'.

Declension

The plural marker is *-LAr*, e.g. *aqindar* 'poets' ← *aqin* 'poet'. Possessive suffixes (Table 19.2) hold the second position in the nominal morpheme chain.

The third position is occupied by case markers (Table 19.3). The paradigm is slightly different for nominals with possessive suffixes. The dative marker attached to first- and second-person possessive suffixes is *-A*. The accusative is marked with *-n* after third-person possessive suffixes. The locative and ablative take on a 'pronominal *n*': *-ndA* and *-nAn* < *-ndAn*.

Table 19.2 Possessive suffixes

	Singular	Plural
1p.	-(I)m	-(I)mlz
2p.	-(I)ŋ	-(I)ŋlz
3p.	-(s)l	-LArl

Table 19.3 Case markers

nom.	∅
gen.	-Nlŋ
acc.	-NI
dat.	-GA
loc.	-DA
abl.	-DAn

Pronouns

The personal pronouns are *men* 'I', *sen* 'you', *ol* 'he, she, it', *biz* 'we', *siz* ~ *sender* ~ *sizler* 'you (plural)', *olar* 'they'. The declension of certain pronouns differs slightly from the above case paradigm (Table 19.4).

Of the demonstrative pronouns, *bul*, *osī* and *mīna* are used for objects within the range of view, and *ol*, *sol*, *ana* for those out of sight or outside the conversational setting. Their dative suffix is *-Gan*. Pronouns ending in *l* replace this consonant by *n* in oblique cases (except in the dative and in some ablative forms). The initial consonant in *bul* varies with *m* in the genitive, accusative and locative. Some demonstratives have so-called emphatic forms, which can take on possessive suffixes, e.g. *mīnaw*, *anaw*, *sonaw*.

The reflexive pronoun in Kazakh is *öz*, used attributively in its bare form or as a noun with possessive suffixes, e.g. *öz ækem* 'my own father', *Özim bardim* 'I went myself'.

Numerals

The cardinal numerals for digits are *bir* 'one', *eki* 'two', *üš* 'three', *tört* 'four', *bes* 'five', *altı* 'six', *žeti* 'seven', *segiz* 'eight', *toyiz* 'nine', for decades *on*

Table 19.4 Personal pronouns

nom.	men	sen	ol
gen.	meniŋ	seniŋ	oniŋ
acc.	meni	seni	oni
dat.	mayan	sayan	oyan
loc.	mende	sende	onda
abl.	menen	senen	onan ~ odan

'ten', *žiyirma* 'twenty', *otiz* 'thirty', *qırq* 'forty', *elüw* 'fifty', *alpis* 'sixty', *žetpis* 'seventy', *seksen* 'eighty', *toqsan* 'ninety'. Higher numerals are *žüz*, often pronounced *düz*, 'hundred' and *miñ* 'thousand'. Complex numbers are formed by compounding, the highest decimal place being on the left. Hundreds, thousands etc. are expressed multiplicatively, e.g. *bir miñ bes žüz alpis eki* '1562'.

Ordinals are formed with *-(I)nšI*. Collectives are only formed from the numerals one to seven with *-(Ø)Aw* (with loss of final stem vowels), e.g. *birew*, *ekew*. The suffixes *-LAp* and *-LAGAn* form approximative and multiplicative numerals, e.g. *žüzdegen* 'hundreds'.

Copulas

Copulas in nominal sentences are formed with enclitic personal markers (Table 19.5). The same set of personal markers occurs in most finite verb forms. The second-person plural form *-slz*, expresses politeness when used to one addressee. The simple plural form is *-sIñdAr*, a combination of the second-person singular form and the plural suffix. The form *-slzdAr* is a particularly polite plural. The past copula particle is *edi* 'was'. The negated copulas are formed with *emes* 'is not'.

Table 19.5 Copulas

	Singular	Plural
1p.	-MIñ	-MIz
2p.	-sIñ	-slz
3p.	Ø	Ø

Verbs

Verb Derivation

The derivation of verbal stems is similar to that of other Kipchak languages. Word-formation suffixes stand closest to the primary stem, and the following slots are filled by voice suffixes, negation and thematic markers of aspect, mood and tense. If the verb form is finite, it is concluded by personal markers of pronominal or possessive origin.

The most common derivation suffix forming denominal verbs is *-LA*, e.g. *žerle-* 'bury' ← *žer* 'soil'. The rather productive suffixes *-sl* and *-sIn* form verbs with the meaning of considering oneself like the entity denoted by the stem, e.g. *žigitsin-* 'consider oneself a brave young man'.

Compound verbs may consist of a noun and a verb, e.g. *bas ur-* 'submit', literally 'hit the head', *at qoy-* 'name', literally 'put the name'. Numerous verbs consist of a native or foreign nominal element plus an auxiliary verb such as *et-* or *qil-*, e.g. *žarq et-* 'flash up', *ümit qil-* 'hope'.

As for the derivation of deverbal verbs, the suffix of verbal negation is *-MA*. Voice is expressed by a set of intercombinable suffixes. Reflexive verbs are marked with *-(I)n*, which often occurs after *-LA*, e.g. *šattan-* 'take pleasure in' ← *šat* 'happy'. Passive voice is expressed by *-(I)l*, e.g. *oqıl-* 'be read', after stems in *-I* by *-(I)n*, e.g. *alın-* 'be taken'. Cooperation and reciprocity is expressed by *-(I)s*, e.g. *söyles-* 'talk together'. The causative suffixes are *-Dir*, *-Glz*, *-t* and the infrequent *-Ar* and *-Ir*, e.g. *aldır-* 'let take', *otıryız-* 'let have a seat'. The variant *-t* is used after polysyllabic stems ending in vowels or sonorants, e.g. *oqıt-* 'teach'.

Compound verbs may consist of a lexical verb and a converb, e.g. *alip bar-* 'carry', literally 'take and go', contracted *apar-*. Kazakh has a limited number of auxiliary verbs used in constructions based on *-A* and *-(I)p* converbs. Thus, *-A al-* with the verb *al-* 'take' denotes ability, e.g. *Žaza alamin*, contracted *Žaz'alamin* 'I can write', *Žaza almaymin* 'I cannot write'. Most auxiliaries specify the manner of action. Thus, *-(I)p al-* with *al-* 'take' and *-(I)p ber-* with *ber-* 'give' signal orientation towards and away from the first actant, respectively, e.g. *žep al-* 'eat up', *añgime aytıp ber-* 'tell a story'. The combination *-(I)p ket-* with *ket-* 'go away' expresses movement away from something, completion and suddenness, e.g. *ölip ket-* 'die (away)', *žürip ket-* 'go away', whereas *-(I)p qal-* with *qal-* 'stay' signals thoroughness or result of the action, e.g. *Žaw kelip qaldı* 'The enemy appeared'. The combinations *-(I)p tasta-* with *tasta-* 'throw' and *-(I)p žiber-* with *žiber-* 'send' convey similar meanings of quick and thorough performance, e.g. *Žoldı kesip tastadı* '(S)he barred the way', *Külip žiberdi* '(S)he burst out laughing'. The combination *-(I)p kel-* with *kel-* 'come' denotes continuity up to some point of orientation, e.g. *Osınday bolıp keldi* 'It has become like that'. In several constructions, *-A* converbs are used in similar ways. The constructions *-A sal-* and *-(I)p sal-* with *sal-* 'lie down' denote fast and unexpected action, e.g. *atıp sal-* 'shoot suddenly'.

Finite Verb Forms

Finite verb forms consist of thematic stems expressing aspect, mood and tense plus personal markers. The Kazakh system displays similarities with the Turkish system, but also considerable differences. The personal markers are of pronominal or possessive origin (Table 19.6). The pronominal type applies to all items except the simple past and the optative-imperative. The negations are generally formed with *-MA*. The present in *-A* with the personal marker *-DI* in the third person is a general present of lower focality, used for habitual and scheduled events, e.g. *Qus uşadı* 'The bird flies'. A more focal present – with a narrower focus on the ongoing event – is formed with *-(I)p + otır, žür, žatır, tur*, with *Ø* in the third person, e.g. *Körıp turmin* 'I see', *Žazıp žatır* '(S)he is writing'. It may be negated with *-MA* followed by the *-A* converb. The auxiliaries are old presents of the verbs *otır-* 'sit', *žür-* 'go', *žat-* 'lie' and *tur-* 'stand', and thus also

Table 19.6 Personal markers

	Pronominal type	Possessive type
1p.sg.	-MIn	-m
2p.sg.	-sIn	-ŋ
3p.sg.	-θ, -dI	θ
1p.pl.	-MIz	-K
2p.pl.	-sIz(dAr), -sIn(dAr)	-ŋIz(dAr), -ŋdAr
3p.pl.	-θ, -dI	θ

convey different shades of actional meaning (continuity, momentaneity, etc.). The aorist in *-(A)r*, negated *-MAs*, with *-θ* in the third person, is mostly used with modal and prospective meanings, e.g. *Keyin köremiz* 'We will see later'.

The simple past consists of *-DI* + possessive personal markers, e.g. *Kördim* 'I saw'. The post-terminal *-GAn*, with *-θ* in the third person, mostly has a perfect meaning, e.g. *Körgeŋmin* 'I have seen'. It is negated with *-MA* or periphrastically with *zoq* or *emes* (for more emphasis). The form in *-AtIn*, after vowels *-ytIn*, with *-θ* in the third person, is a habitual past, e.g. *Dem alatınbiz* 'We used to take a rest'. The form in *-Ip*, with *-DI* in the third person, is an indirective past, e.g. *Köripti* '(S)he has apparently seen it'.

The form in *-MAK*, negated *-MAK emes*, with *-θ* in the third person, denotes intention, e.g. *Kelmekpiz* 'We want to come'. The enclitic *-šI* adds a meaning of necessity, e.g. *Satpaqšimin* 'I must sell'.

Kazakh has a mixed optative-imperative paradigm (Table 19.7). The emphatic enclitic particle *-šI* can be added in all persons.

A number of forms are constructed with the past copula particle *edi* + possessive personal markers. Thus, *-(Ip otir (~ žür ~ žatir ~ tur) edi* form focal imperfects, *-(θ)UwšI edi* a habitual or durative past, e.g. *Külüwši edim* 'I always laughed', *-GAn edi* a pluperfect, *-(A)r + edi* a habitual or a counterfactual ('would do'), e.g. *Men onı körsem, birden tanır edim* 'If I saw him, I would recognise him at once', *-AtIn edi* a habitual pluperfect or a past counterfactual, e.g. *Biz žürip ketetin edik* 'We would have left'.

The copula particle *eken*, which has indirective meaning, may combine with different items, e.g. *Keledi eken* '(S)he obviously comes'.

Table 19.7 Optative-imperative

	Singular	Plural
1p.	-(A)yIn	-(A)yIK
2p.	θ	-(I)ŋIz, -(I)ŋdAr
3p.	-sIn	-sIn

Non-finite Verb Forms

Kazakh has numerous verbal nominals such as infinitives and participles. Thus, *-(θ)Uw* (with drop of preceding stem-final high vowels) forms verbal nouns used as the lexicon form of Kazakh verbs, e.g. *oquw* 'reading, read' ← *oqi-* 'read'. The participle in *-GAn* differs from the corresponding finite item by referring to both anterior and non-anterior events, e.g. *kelgeŋ kisi* 'the person who came ~ has come ~ comes'. The infrequent item in *-A žatqan* is intraterminal, e.g. *kele žatqan kisi* 'the person who is/was coming'. The participle in *-AtIn* may refer to habitual and future events. The aorist in *-(A)r* is rarely used as a participle. Like the corresponding finite form, *-MAK(šI)* has intentional or necessitative meaning, but unlike other participles, it does not take on case, possessive or plural suffixes.

Among the converbs, the one in *-(Ip)* has simple junctive functions and is used in clause linking. The final *-I* of monosyllabic stems is sometimes lost in front of *-(Ip)*, e.g. *bop* ← *bol-* 'become'. The converb in *-A* (*-y* after vowels) is mainly used in complex verbs and has limited functions in clauses. Its negative form is used as the negation of *-(Ip)*. Among the more specific converbs, *-GAlI* has purposive and sometimes temporal meaning, e.g. *Körgeli keldi* '(S)he came to see it'.

Adverbs

Adverbs are a weakly delimited word class in Kazakh. Most adverbially used words do not show any specific morphological characteristics, e.g. *endi* 'now', *erte* 'early'. Others are fossilised converbs, e.g. *Žayawlap keldik* 'We came on foot', or nouns with case suffixes, e.g. *Birge oqıdıq* 'We studied together'. The most common adverb-forming suffix is *-šA*, e.g. *adamša söyle-* 'speak like a human being'.

Other Morphological Categories

Kazakh has numerous postpositions expressing various logical, spatial and temporal relations not signalled by the cases. Thus, *-Men(en)* expresses instrumental and comitative meanings, e.g. *oraqpen* 'with a sickle'. Postpositions may govern the nominative, such as *tuwralı* 'about', e.g. *Sen tuwralı söyledi* '(S)he talked about you', or the dative such as *deyin* 'until', e.g. *sayat segizge deyin* 'until five o'clock', or the ablative such as *soŋ* 'after', e.g. *sabaqtan soŋ* 'after the lesson'. A number of auxiliary nouns such as *üst* 'top', *ast* 'bottom' and *art* 'back' are used in possessive constructions to express spatial and other relations, e.g. *üydin artında* 'behind the house'.

Unlike more strongly Persianised Turkic languages such as Uzbek and Turkmen, Kazakh has a weakly developed system of conjunctions. The conjunction *žæne* 'and' usually only coordinates two clauses. It often occurs in constructions copied from Russian, being itself in some respects a semantic copy of Russian *i* 'and', whose use has been reinforced in the written language. *Biraq* 'but' is adversative, whereas *yæ* and *yæki* 'or', both copied from Persian, are disjunctive.

The interrogative enclitics *MA* and *DA* 'also' obey the rules of vowel and consonant assimilation. The restrictive element *Kana* and the affirmative element *Koy* have invariable vowels, but their initial consonant assimilates to the preceding word-final segment, e.g. *Ol keşe yana keldi* '(S)he came only yesterday', *Ayttim yoy* 'I did tell you'. Kazakh orthography often indicates enclitic elements by writing them separately or hyphenated.

Syntax

Nominal Phrases

Nominal phrases are structured as in other Turkic languages. There are thus nominal compounds of the common possessive type, e.g. *köz žasi* 'tear'.

The Sentence

Finite Clauses

As in other Turkic languages, the unmarked order of clause constituents is subject, object, predicate.

Comparison may be expressed by an ablative construction without any morphological marking of the adjective, e.g. *Almati Semeyden ülken* 'Almaty is bigger than Semey'.

Possession is expressed by *bar* 'existing', lack of possession by *zoq* 'non-existing'. Elements expressing the possessed entity carry possessive suffixes indicating the possessor, e.g. *Balasi bar* '(S)he has a child'.

Yes/no questions are formed with the enclitic particle *MA*, whose unmarked position is after the predicate, e.g. *Balasi bar ma?* 'Does (s)he have a child?' The variant *-MI* occurs before personal suffixes, e.g. *Aman sawmisiñ?* 'Are you fine?'

Since person is usually marked on the predicate, personal pronouns are used mainly in contrasting or focusing functions.

Non-finite Clauses

Relative clauses and complement clauses in Kazakh are left-branching, using various non-finite verb forms. Unlike highly Persianised Turkic languages such as Uzbek and Azerbaijani, Kazakh does not apply right-branching methods of clause subordination with finite verbs.

The first actant of the relative clause is not expressed overtly as the subject, if it is coreferent with the head, e.g. *kelgen ayel* 'the woman who has come'. Otherwise, the first actant may be expressed overtly, e.g. *men aytqan söz* 'the word I said'. The subject can also be expressed by a genitive with a corresponding possessive suffix on the head, e.g. *meniñ aytqan sözüm* 'the word I said'.

Complement clauses are embedded in various syntactic functions by means of non-finite forms plus case-markers. There are subject clauses such

as *Solay ekeni ras* 'It is true that this is so' or object clauses such as *Žurtıñ ne deytinine qulaq saldı* '(S)he listened to what the people said'. Participles in complement clauses may take on an additional *-LIK*, e.g. *Bügin žıynalıstı bolatındıyın estidik* 'We heard there would be a meeting today'.

While the semantic contents of simple converbs may be vague, more precise relations are expressed with complex converbs, non-finite forms with case markers or postpositions, or lexical elements with different subordinators.

Purpose clauses are expressed with *-(θ)Uw üşin*, e.g. *Almatiya oquw üşin keldim* 'I came to Almaty in order to study', or with *-sın dep*, e.g. *Balanıñ oyı össin dep dämdi sözdi izdedim* 'I searched for nice words in order for the child's mind to develop'. Causal clauses may be expressed with *-GAndIKtAn*, e.g. *elsiz bolyandıqtan* 'since it is uninhabited', *-(θ)Uw sebepti*, e.g. *kitaptıñ bolmawı sebepti* 'because there are no books'.

Clauses expressing simultaneousness may be formed with *-GAnda*, e.g. *Sonı oylayanda köp nәрse eske tüsedi* 'When (s)he thinks of it, many things come into his/her mind'. More complex forms are based on *kezde*, *waqıtta* 'at the moment, at the time', e.g. *söyleytin kezde* 'when (s)he talked'. The use of the locative suggests that these constructions are formally less grammaticalised than, for example, Turkish *-DIğI zaman*. Clauses expressing an event after which the matrix clause event takes place ('after') are formed with *-GAn soñ* or *-GAnnan keyin*, e.g. *Žoldastar tügel žıynalyan soñ oyın bastaldı* 'After all the friends had gathered, the game began', *žumisın bitirgennen keyin* 'after finishing his/her work'. Terminative clauses expressing an event up to which the matrix clause event continues ('until') are formed with *-GAnšA*, *-GAnGA deyin* or *-GAnGA šeyin*, e.g. *žıynalıstı bitkenše* 'until the meeting ends', *žumisšılar žıynalyany deyin* 'until the workers gather'.

Clauses expressing comparison with another event are formed with *-GAndAy*, e.g. *Kempir žerden altın tapqanday quwandı* 'The old woman was as happy as if she had found gold in the ground'. Clauses expressing 'without doing' are formed with *-MAstAn*, e.g. *Soyan qaramastan Batırbek ketti* 'Batırbek left without paying attention to it'. Clauses expressing 'instead of doing' are formed with *-GAnšA*, e.g. *Dalaya qomyanša molaya qon* 'Rather than settling in the steppe, settle in the cemetery'.

Concessive clauses are formed with *-GANmen*, e.g. *Özenıñ suwı tayız bolyanmen žayası tik* 'Though the water of the river is shallow, its banks are steep'.

Conditional clauses are formed with *-sA* plus possessive personal markers, e.g. *Ol kelse men ketemin* 'If (s)he comes, I will go'. As in other central Asiatic Turkic languages, *-sA* may also have temporal meaning, e.g. *šiqsam* 'when I go ~ went out'. The conditional content may, however, be specified by *eger* 'if', copied from Persian via Persianised Turkic languages, e.g. *Eger ruqsat etseniz, meniñ elge qaytım keledi* 'If you permit, I want to return home'.

Lexicon

The basic Kazakh lexicon belongs to the common Kipchak vocabulary. The lexis related to the traditional way of life, e.g. cattle breeding, feasts, food, is richly developed. There are numerous pairwords, e.g. *kiyim-kešek* 'clothing', *qarim-qatinas* 'relation'.

As in other languages of Islamic Turks, the vocabulary contains numerous lexical elements copied from Persian and Arabic. This part of the lexicon, which has entered Kazakh via Tatar and Chaghatay, covers the spheres of abstract concepts and of Oriental urban life, e.g. *aqil* 'intellect', *bazar* 'market'.

Words of Mongolic origin are also in use, e.g. *qunan* 'colt in the third year', *olža* 'booty'. Most of them date back to the eighteenth century, when Kazakh and western Mongol tribes fought for supremacy in the steppes between the Altay mountains and the Caspian Sea.

Russian words have been borrowed from the second half of the nineteenth century onwards. Their phonological adaptation was reflected in the Arabic and Roman orthography, e.g. *kiylö* 'kilogram', *telgiræp* 'telegraph', *bælsæbek* 'Bolshevik', but after the introduction of the Cyrillic script, Russian loanwords have been written in their original form. In the later Soviet period, calques and neologisms were preferred over loanwords, e.g. *xaliqaraliq* 'international', *önerkæsip* 'industry'. Many efforts have been made to develop and standardise Kazakh terminology. However, since most communication in professional domains, education and science is in Russian, the use of Kazakh terminology is more or less restricted to textbooks. There is now a weak tendency – not yet reflected in spoken Kazakh – to reduce the number of Russian loans in favour of native or Arabic–Persian lexemes.

Dialects and Variation

Kazakh differs from neighbouring languages such as Tatar, Uzbek and South Siberian Turkic by exhibiting relatively little local variation. Given the enormous extension of the Kazakh area, this may at first sight be surprising. The Kazakhs have been roaming cattle breeders with an exogamic family structure, and the permanent mobility within the population has counteracted the formation of clearly distinct dialects. Mass deportations, famine and the flight of entire clans in the 1930s alongside deportation of other ethnic groups to Kazakhstan in the 1940s has changed the population structure to the extent that methods of traditional dialectology fail.

Kazakh as spoken in the southern and the western areas still differs in some respects from the standard language, which is based on the northwestern dialects. Some sound changes characteristic of standard Kazakh are less generalised. Common Turkic *y-*, which has developed to *ž-*, is preserved in some environments. Common Turkic *č* is often preserved, while standard Kazakh prefers *š*.

Morphological particularities of southwestern dialects include an optative–imperative paradigm with 1p.sg. *-(A)y* and 1p.pl. *-(A)ll* instead of *-(A)yln* and *-(A)yIK*. Assimilations and dissimilations of suffix-initial *l*, *m* and *n* are less strong in these dialects, e.g. *tiŋlaw* 'listen(ing)' vs. standard *tiŋdaw*, whereas they are stronger in the easternmost dialects spoken in China and Mongolia, e.g. *baladar* 'children' vs. standard *balalar*. The Kazakh territories are part of a Turkic continuum of increasing assimilation processes extending from west to east and culminating in Siberian Yakut.

The literary language of the Kazakhs living in China has few lexical imitations from Chinese and is still oriented towards the standard language used in Kazakhstan.

Karakalpak

Though Karakalpak can be characterised as a slightly Uzbekised variant of Kazakh, it has preserved most phonetic shifts and morphophonemic rules characteristic of Kazakh. An exception is *ll*, which has three variants *l*, *d*, and *t* in Kazakh, but only *l* in Karakalpak, e.g. *qustar* 'birds' vs. Kazakh *qustar*. There are few morphological differences. The present in *-Atuyun* vs. Kazakh *-Atln* represents an older stage of development of *-A turyan*. The Karakalpak future in *-Ažaq* is a loan from Oghuz. The first- and second-person singular personal markers *-MAN*, *-sAŋ* vs. Kazakh *-Mln*, *-slŋ* are reminiscent of Uzbek *-mæn*, *-sæn*. Karakalpak uses different postpositions and conjunctions, some being borrowings from Uzbek and Turkmen, e.g. *šelli*, *qadar* 'until', *hæm* besides *žæne* 'and'. The lexicon differs to a certain degree from Kazakh by containing more imitations from Arabic and Persian (via Uzbek and Turkmen), e.g. *diydar* 'face' < Persian.

Karakalpak was established as a written language in 1925. After the Arabic and Latin scripts had been employed for a few years, the Cyrillic alphabet was introduced in 1940. Since the Karakalpaks live within Uzbekistan they will join the Uzbek transition to a new Latin script based on the Turkish system.

References

- Balakaev, M. B. *et al.* (eds) (1962) *Sovremennyj kazaxskij jazyk*, Alma-Ata: Izdatel'stvo Akademii Nauk Kazaxskoj SSR.
- Baskakov, N. A. (1952) *Karakalpakskij jazyk 2: Fonetika i morfologija*, Moscow: Izdatel'stvo Akademii Nauk SSSR.
- (1958) *Karakalpaksko-russkij slovar'*, Moscow: Gosudarstvennoe izdatel'stvo inostrannyx i nacional'nyx slovar'ej.
- Kirchner, M. (1992) *Phonologie des Kasachischen: Untersuchungen anhand von Sprachaufnahmen aus der kasachischen Exilgruppe in Istanbul 1–2*, Turcologica 10, Wiesbaden: Harrassowitz.
- Krueger, J. R. (1980) *Introduction to Kazakh*, Teaching Aids for the Study of Inner Asia 9, Bloomington: Indiana University.
- Menges, K. H. (1959) 'Die aralo-kaspische Gruppe', in J. Deny *et al.* (eds) *Philologiae Turcicae Fundamenta 1*, Wiesbaden: Steiner, 434–88.

- Mirzabekova, Q. (ed.) (1992) *Qazaqşa-nemisşe sözdik* (Kazakh–German dictionary), Almatı: Rawan.
- Shnitnikov, B. N. (1966) *Kazakh–English Dictionary*, Indiana University Publications, Uralic and Altaic Series 28, London, The Hague, Paris: Mouton.
- Wurm, S. (1951) 'The Karakalpak language', *Anthropos* 46: 487–610.

20 Noghay

Éva Á. Csató and Birsel Karakoç

Introduction

Noghay belongs to the South Kipchak group of Turkic languages, sharing the main characteristics of Kazakh, Karakalpak and Kirghiz. The designation goes back to the name of Emir Noghay (d. 1299). Noghay tribes played an important role in the Golden Horde, notably in the fifteenth century, their main area of influence being the steppes between the Volga and the Urals.

The earliest known Noghay text (Osmanov 1883) is written in Arabic script. The first Latin alphabet of Noghay was created in 1928, and replaced in 1938 by a new Cyrillic alphabet, which is still in use. The Noghays have a rich folk literature, of which important editions have been published. The first major grammar of Noghay was published by N. A. Baskakov in 1940.

Sound System

Phonetics and Phonology

The Noghay vowel system contains nine vowels including two front unrounded non-high vowels /e/ and /æ/ with distinctive functions, e.g. *ær* 'every' vs. *er* 'man'. The low vowel /æ/ occurs often in initial position in lexical items copied from Persian and Arabic, e.g. *æl* 'situation, state', *ævlet* 'child'.

There are 22 consonant phonemes in words of Noghay origin, the affricate /j/ occurring only in some dialects. The five sounds *f, v, c, č, šč* are only found in copied lexical items. The original Noghay labial fricative has a bilabial

Table 20.1 Vowels

	<i>Front unrounded</i>	<i>rounded</i>	<i>Back unrounded</i>	<i>rounded</i>
high	i	ü	ĩ	u
middle	e			
low	æ	ö	a	o

Table 20.2 Consonants

	Labial	Dental	Palatal	Velar	Uvular
stop	b, p	d, t		g, k	q
fricative	w, v, f	s, z	š, ž		ɣ, χ
affricate		c	j, č, šč		
nasal	m	n		ŋ	
glide			y		
liquid		l, r			

articulation, e.g. *tuwil* 'not, no'. The labiodental variant occurs in copied items, e.g. *vagon* 'railway carriage'. The consonants *k* and *g* occur only in front syllables, whereas *q* and *ɣ* occur in back environments.

An *š* in modern Noghay goes back to *s* or *č*, e.g. *šaš* 'hair' < *sač*, *aš*- 'open' < *ač*-. An original *š* has become *s*, e.g. *as* 'food' < *aš*.

The pitch accent falls on the last syllable in Noghay as usual in other Turkic languages. Exceptions include some of the numerals such as *älti* 'six', *yéti* 'seven', imperative forms such as *Óltir* 'Sit!' and some interrogative pronouns such as *qáysi* 'which'. The list of unaccentable suffixes is basically the same as in other Turkic languages.

Morphophonology

Front vs. back harmony functions as in other Turkic languages. Most suffixes have front and back variants except for some suffixes such as the participle suffixes *-Atayan*, *-(A)yatqan*, e.g. *oltiratayan* 'sit:PART.HAB.INTRATERMINAL', *keleyatqan* 'come:PART.FOCAL.INTRATERMINAL'. Some of the old compounds have also developed harmonic forms, such as *aket-* 'take away' < *alip ket-*, *akel-* 'bring' < *alip kel-* and *söytip* 'this way, so' < *solay etip*. There is no rounded vs. unrounded harmony in standard written Noghay. Suffixes with a high vowel have only two variants, a front and a back one. Thus, the genitive suffix *-D³Iŋ* has only unrounded variants, e.g. *közdiŋ* 'eye:GEN.', *ottiiŋ* 'grass:GEN.', whereas the derivational suffix *-(U)w* has only rounded variants, e.g. *satuw* 'sale' ← *sat-* 'sell' and *körüw* 'look' ← *kör-* 'see'. However, rounded variants of suffixes are frequently used in the spoken language, e.g. *Ölgön* vs. *Ölgen* '(S)he has died', *Kördü* vs. *Kördi* '(S)he saw'.

Progressive consonant assimilation includes *ml* > *mn*, e.g. *ædem* + *LAr* > *ædemner* 'man:PL.', *nl* > *nn*, e.g. *zaman* + *LAr* > *zamannar* 'time: PL.'. This assimilation is not reflected in the orthography. Other assimilations are: *md* > *mn*, e.g. *ædem* + *D³An* > *ædemnen* 'man:ABL.', *nd* > *nn*, e.g. *ton* + *D³Iŋ* > *tonniŋ* 'fur coat:GEN.'. Suffixes begin with a voiceless consonant after voiceless consonant stems, e.g. *ayaš* + *D³An* > *ayaštan* 'tree:ABL.'. The negation suffix *-MA* is pronounced as *-pA* after voiceless consonants, as *-bA* after voiced consonants and as *-mA* after vowels, e.g. *ayt-* + *-MA* > *Aytpa* 'Do

not say it!', *qaz-* + *-MA* > *Qazba* 'Do not dig!', *söyle-* + *-MA* > *Söyleme* 'Do not speak!'

The standardised forms of suffixes are to be read according to the following rules.

- A* in front words *e*, in back words *a*. As the first element of some suffixes, it may be realised as *y* after a vowel, e.g. the converb suffix *-A*, *bar* 'go' + *-A* = *bara* 'going', *barma* 'not go' + *-A* = *barmay* 'not going'.
- I* in front environments *i*, in back environments *i*.
- L* after nasals *n*, otherwise *l*.
- D²* after voiceless consonants *t*, otherwise *d*.
- D³* after voiceless consonants *t*, after nasal *n*, otherwise *d*.
- G* in back environments after voiceless consonants *q*, after voiced consonants *ɣ*, in front environments after voiceless consonants *k*, after voiced consonants *g*.
- M* after voiceless consonants *p*, after nasals *m* and otherwise *b*.
- K* in back environments *q*, in front environments *k*.

Orthography

In the currently used Cyrillic alphabet, *e* is written as <e> or <э>. The latter renders initial *e* in words of Noghay origin, e.g. *эки* = *eki* 'two'. The former is read as *ye* word-initially, e.g. *ел* = *yel* 'wind'. The letter <e> is also pronounced as *ye* after a vowel, <ь> or <ь>, e.g. *аел* = *ayel* 'family', *ие* = *iye* 'owner', *съезд* = *syezd* 'congress'. The vowel *æ* is rendered by the combination <аь> as in *аьрекет* = *areket* 'activity, work', while *ö* is rendered by <оь> and *ü* by <уь>. Both the bilabial and labiodental fricatives are written as <в>. Both *l* and *ʃ* are written as <л>. In the modern orthography, both *k* and *q* are rendered by <к>, just as *g* and *ɣ* are both written as <г>. Earlier orthographies included different letters for these sounds: <к> versus <кь>, <г> versus <гь>.

Devoicing of final consonants in copied lexical items is not reflected in the orthography. Thus, the locative forms of *doklad* 'lecture' and *klub* 'club' are written докладта, клубта.

Morphology

Nouns and Adjectives

Comparison

The comparative form of adjectives is formed with *-(I)rAK*, e.g. *üykenirek* 'bigger' < *üyken* 'big'. The superlative is an analytic form built with the particle *eŋ*, e.g. *eŋ üyken* 'the biggest'. *Bek* and *aylaq* mean 'very', e.g. *bekl aylaq üyken* 'very big'.

Table 20.3 Case suffixes

Case	Suffix	Examples			
nom.	-Ø	bala 'child'	terek 'tree'	qoyan 'hare'	atlar 'horse:PL'
gen.	-D ³ Iŋ	baladıŋ	terekiŋ	qoyanıŋ	atlarıŋ
acc.	-D ² I	baladı	tereki	qoyandı	atları
dat.	-GA	balaya	terekke	qoyanya	atlaya
loc.	-D ² A	balada	terekte	qoyanda	atlarda
abl.	-D ³ An	baladan	terekten	qoyannan	atlardan

Declension

The plural suffix is *-LAR*, e.g. *üyeler* 'houses', *sular* '(masses, kinds of) water'. The accusative suffix attached to third person possessive suffixes is *-n*, e.g. *Onıŋ üyin* [(s)he:GEN. house:3POSS.:ACC.] 'his/her house:ACC.', *Ol öziŋ dosın kördi* [(s)he own:3POSS.:GEN. friend:3POSS.:ACC. see:PST.] '(S)he saw his/her own friend'.

The dative suffix after first and second person possessive is *-A*, e.g. *menim aldım* [I:GEN. front:1POSS.:DAT.] 'towards me', *seniŋ yüreğiŋe* [you:GEN. heart:2POSS.:DAT.] 'to your heart'. It has this form also in the third person following a 'pronominal *n*', e.g. *Ol atasına usayan* [(s)he father:3POSS.:N:DAT. be similar:POSTT.PART.] '(S)he is similar to his/her father'.

The 'pronominal *n*' follows the third person possessive when the dative, locative or ablative is attached, e.g. *küzdüŋ ertengi şayında* [autumn:GEN. early time:3POSS.:N:LOC.] 'early in autumn'. This *n* triggers assimilation when the ablative follows: *-(s)I + n + D³An > -(s)InnAn*, e.g. *tawdıŋ töbesinnen* [mountain:GEN. top:3POSS.:N:ABL.] 'from the top of the mountain'.

The suffix *-D²Ay* 'like X' is also preceded by the 'pronominal *n*' when following a third-person possessive suffix, e.g. *onday* 'like it'.

The possessive suffixes are 1p.sg. *-(I)m*, 2p.sg. *-(I)ŋ*, 3p.sg. *-(s)I*, 1p.pl. *-(I)mız*, 2p.pl. *-(I)ŋız*, 3p.pl. *-LARı*, e.g. *ana* 'mother' → *anam*, *anaŋ*, *anası*, *anamız*, *anaŋız*, *analari* 'my [etc.] mother', *tawıs* 'voice' → *tawısım*, *tawısıŋ*, *tawısı*, *tawısımız*, *tawısıŋız*, *tawıslari* 'my [etc.] voice'. Note that a form such as *balalari* is ambiguous and may be translated as 'its children', 'their child', or 'their children'.

Pronouns

The plural suffix can be attached to the first- and second-person plural pronouns, *bizler* 'we', *sizler* 'you:PL.', in order to individualise the persons to which the pronouns refer.

The possessive pronouns are formed with the complex suffix *-niki* and its variants, e.g. *meniki* 'mine', *seniki* 'yours', *onıqi* 'his, hers, its', *bizdiki* 'ours', *sizdiki* 'yours', *olardıqi* 'theirs'.

The reflexive pronoun *öz* can be used as an attribute, e.g. *öz balam* 'my own child', *onıŋ öz balası* 'his/her own child', or as a noun when bearing a

Table 20.4 Personal pronouns

	men 'I'	sen 'you'	ol 'he, she, it'	biz 'we'	siz 'you:PL'	olar 'they'
nom.	men	sen	ol	biz	siz	olar
gen.	menim	seniŋ	onıŋ	bizim	siziŋ	olardıŋ
acc.	meni	seni	onı	bizdi	sizdi	olardı
dat.	mayı	saya	oya	bizge	sizge	olayrı
loc.	mende	sende	onda	bizde	sizde	olarda
abl.	mennen	sennen	onnan	bizden	sizden	olardan

possessive suffix, e.g. *Men özüm körgenmen* 'I have seen it myself'. Reciprocal pronouns are formed with *öz + öz + possessive and case suffixes*, e.g. *öz-özine* 'to each other'.

The simple demonstrative pronouns are *bu* 'this', plural *bular*, *ol* 'that', plural *olar*, *sol* 'that (same)', plural *solar*. Other demonstrative pronouns are *munaw*, *anaw*, *sosi* and combinations such as *mine-munaw* 'this here', *æne-anaw* 'that there'. The case forms of *ol* and *sol* are built according to the following pattern: *onıŋ* (GEN.), *onı* (ACC.), *oya* (DAT.), *onda* (LOC.), *onnan* (ABL.). *Bu* has the oblique stem *mu-* except in the dative: *munıŋ* (GEN.), *munı* (ACC.), *munda* (LOC.), *mundan* (ABL.), but *buya* (DAT.).

Adjectival forms of the pronouns are *munday*, *onday*, *sonday*, *munawday*, *anawday*, *sosinday*.

The interrogative pronouns include *kim* 'who', *ne* 'what', *kimdiki* 'whose', *nediki* 'of what', *neşe* 'how many/how much', e.g. *Sen neşe yasında?* 'How old are you?', *qaysı* 'which', e.g. *Qaysı awılta yaşaysıŋ?* 'In which village do you live?', *qalay* 'how', e.g. *Æliŋ qalay?* 'How is your condition?', 'How are you?' Adverbs derived from interrogative pronouns include *qayday* 'how', e.g. *Qayday kömek ettiŋ?* 'How did you help?', *qaytip* < *qalay etip* 'how', *qaşan* and *qayzaman* 'when', *qayda* and *qayerde* 'where', e.g. *Sen yay küni qayda boldıŋ?* 'Where were you in the summer?', *qaydan* and *qayerden* 'from where', *nege* 'why'.

Indefinite pronouns are *bir kim* and *birew* 'somebody', *bir zat* 'something', *eşkim* and *eşbirew* 'anybody', *eşbir* 'anything', *ærisi* 'each of them', *bærisi* 'all of them', *bæri yaslar* 'all the young people'.

Numerals

Cardinal numbers are *bir* 'one', *eki* 'two', *üš* 'three', *dört* 'four', *bes* 'five', *altı* 'six', *yeti* 'seven', *segiz* 'eight', *toyüz* 'nine', *on* 'ten', *yirma* 'twenty', *otüz* 'thirty', *qırq* 'forty', *elli* 'fifty', *alpis* 'sixty', *yetpis* 'seventy', *seksen* 'eighty', *toqsan* 'ninety', *yüz* 'hundred', *miŋ* 'thousand'.

Ordinals are formed from cardinals with *-(I)nşI*. For 'first', the expression *bas degen* or the adverbial *bas degende* is often used, e.g. *bas degen yılda* 'in the first year'. Multiplication is expressed by *kere* 'times', e.g. *eki kere* 'twice', *üš kere* 'three times'. Collective numerals are formed with *-Aw*, e.g.

altaw 'six together', from number eight onwards with *-llk*, e.g. *segizlik* 'eight together'. Distributive numerals are formed with *-(s)Ar*, *-lAp*, *-(s)ArlAp*, e.g. *altısar* 'six each', *onlap*, *onarlap* 'ten each'. Fractions are constructed according to the pattern *altıdan bir* [SIX:ABL. ONE] 'one sixth'.

Verbs

Voice

Passives are usually formed with *-(I)l*, *-(I)n*, e.g. *başlan-* 'be begun, started', *alın-* 'be taken', *tabıl-* 'be found'. But passives may also be formed with the complex suffixes *-(I)lln*, e.g. *esitilin-* 'be heard', and *-nll*, e.g. *taslanıl-* 'be thrown', *islenil-* 'be worked on'. The reflexive marker is *-(I)n*, e.g. *şeşin-* 'undress'.

Causatives are formed with *-t* and *-D²Ir*, e.g. *aşat-* 'feed', *oltırt-* 'let sit', *aştır-* 'let open', *küldir-* 'make laugh'. Some causatives are formed with *-KAr*, e.g. *yatqar-* 'lay', *yetker-* 'lead', with *-(V)r*, e.g. *uşır-* 'let fly', *işir-* 'let drink', and with *-Glz*, e.g. *kiygiz-* 'dress, put on (transitive)', *yatqız-* 'lay'. Double causative forms are used to express orders: *-Glst* < *-Glz* + *-t*, e.g. *turyüst-* 'order or force to stand up', *baryüst-* 'order to go', *-Dlrt* < *-Dlr* + *-t*, e.g. *yıydirt-* 'order to collect', *-tlr* < *-t* + *-tlr*, e.g. *oqıttır-* 'order to read', *-Art* < *-Ar* + *-t*, e.g. *qızart-* 'make red, redden'.

Cooperative-reciprocal stems are formed with *-(I)s*, e.g. *külis-* 'laugh together', *köris-* 'see each other'.

Actional Modifications

The actional meaning of lexical verbs may be modified by periphrastic verbal constructions based on the converbs in *-A* or *-(I)p* of the lexical verb and conjugated forms of auxiliary verbs. The auxiliaries (postverbs) *al-* (< 'take'), *ber-* (< 'give'), *yiber-* (< 'send, release'), *qoy-* (< 'put'), *şiq-* (< 'leave, go out') serve to build transformative actional phrases, e.g. *oltıra ber-* 'take a seat', *aytıp şiq-* 'tell to the end' (see pp. 42, 113–14). The auxiliaries *tur-* (< 'stand'), *oltır-* (< 'sit'), *yür-* (< 'walk'), *yat-* (< 'lie'), *qal-* (< 'stay') build non-transformative verbs e.g. *oqıp tur-* 'keep on reading', *uyqlap qal-* 'keep on sleeping'. *Basla-* (< begin) is used to specify the initial phase of an action, e.g. *Xattı oqıp başladı* '(S)he began to read the letter'. *Kel-* (< 'come') and *bar-* (< 'go') specify the direction towards or from the speaker, e.g. *uşıp kel-* 'fly here (hither)', *uşıp bar-* 'fly there (thither)'.

Yaz- (< 'miss') is used to express an action that has almost taken place, e.g. *Kirip kete yazdım* [ENTER:CONV. AWAY.POSTVERB:CONV. ALMOST.POSTVERB:PST.] 'I almost entered'.

Finite Verb Forms

Personal markers of pronominal origin are 1p.sg. *-MA_n*, 2p.sg. *-sl_n*, 3p.sg. \emptyset , or in the low-focal present (see below; cf. p. 43), 1p.pl. *-Mlz*, 2p.pl. *-slz*, 3p.pl. \emptyset , *-D²I* or *-D²IIAr*. Personal markers of possessive origin are 1p.sg. *-m*, 2p.sg.

-ŋ, 3p.sg. \emptyset , 1p.pl. *-K*, 2p.pl. *-(I)ŋlz*, 3p.pl. \emptyset or *-LAR*.

Intraterminal forms include the following ones. The low-focal (general) present: *-A*, negated *-MA_y*, with the ending *-D²I* in the third person, e.g. *Baraman* 'I go', *Keledi* 'It comes'. The corresponding low-focal imperfect is built with the past copula particle derived from the obsolete verb *e-*: *-A edi*, e.g. *Men onı bile edim* 'I knew that'.

The habitual intraterminal past form is *-(A)ta_yan (edi)*, negated *-MA_yta_yan (edi)*, e.g. *Ol qalada turatayan edi* '(S)he lived in the city'. The habitual form *-(A)ta_yan bol_yan* is used in a particular discourse type based on the *-GAn* forms.

The high-focal presents are *-(A)yatır*, *-(I)p turı*, *-(I)p yatır*, *-(I)p oltırı*, *-(I)p yüri*, *-A turadı*. These complex forms include auxiliary verbs such as *tur-* 'stand' and *yat-* 'lie'. The most frequently used negative forms are *-MA_y turı* and *-MA_y yatır*. The corresponding high-focal imperfects are formed with *edi*: *-(A)yatır edi*, *-A tura edi*, *-(I)p turı edi*, *-(I)p yatır edi*, *-(I)p oltırı edi*, *-(I)p yüri edi*, e.g. *Oqıp yatır edi* '(S)he was just reading'. The form *-Ayatqan bol_yan* is used in the discourse type based on the *-GAn* form.

The non-focal aorist in *-(A)r*, negative *-MA_s*, expresses modal meanings, e.g. *Aytpaspan* 'I will not say it'.

The forms expressing post-terminality include the following. The low-focal perfect is formed with *-GAn*, negated *-MAGAn*, e.g. *Barmayanman* 'I have not gone', *-GAn yoq* + personal markers, e.g. *Kelgen yoqpan* 'I have not come'. The corresponding pluperfect is *-GAn edi*, negated *-GAn yoq edi* + personal markers, e.g. *Bergen yoq edim* 'I had not given'. The low-focal in *-(I)p* also expresses indirectivity, e.g. *Yazıptı* '(S)he has apparently written'. The corresponding pluperfect is *-(I)p edi*, or *-(I)ptı edi*, e.g. *Yazıptı edi* '(S)he had written'. The high-focal (resultative) perfect is *-(I)p tu'ri*. A corresponding high-focal pluperfect is *-(I)p tu'ri edi*. Accentuation distinguishes these forms from the corresponding intraterminal ones in which the suffix *-(I)p-* is accented.

A simple past is formed with *-DI*, whereas a prospective (future) is marked with *-(A)yAK*, negated *-MA_yAK*. Conditionals are formed with *-sA*. The copula item *eken* expresses indirectivity, e.g. *Keleyatır eken* 'He is apparently coming'.

The converb in *-(I)p* of lexical verbs combined with the auxiliary verb *bol-* yields forms expressing 'be able to', e.g. *Yazıp bolaman* 'I can write'.

Non-finite Verb Forms

A habitual participle is formed with *-Atayan*, *-(U)wşı*, e.g. *Sen aytatayan zat durıs tuwil* 'It is not right what you (usually) say'. The high-focal intraterminal participle is formed with *-(A)yatqan*, e.g. *barayatqan ademler* 'people who are just going' or *-(A)yatır_yan*. A prospective (future) participle is formed with *-(A)yAK*, and a post-terminal participle in *-GAn*, e.g. *Meni sorayan kimdi?* 'Who has asked for me?'

The most frequently used converb in *-(I)p* has a vague semantic content. Its negation *-MAy* is based on the converb in *-A*. Other converb forms are based on the participle *-GAn*, e.g. *-GAnDA* 'when X', *-GAnnAn (soñ)* 'after X', *-GAnšA* 'until X', *-GAnIAy* 'as soon as X', *-GAIL* 'since X'.

The verbal noun in *-(U)w* is the form used to represent verbs in dictionaries, e.g. *söylew* 'say'. Other productive nominal suffixes include *-MA*, e.g. *kelme* 'coming', *yapma* 'making', *-(I)s*, e.g. *yüris* '(way of) walking, gait', *-GI*, e.g. *Baryim keledi* 'I would like to go'. The suffix *-MAK* is less productive.

Other Morphological Categories

Postpositions

Postpositions may govern the nominative, e.g. *sayin* 'throughout', *yil sayin* 'the whole year', the genitive, e.g. *soniñ üšin* 'because of it', the dative, e.g. *deyim* 'until', or the ablative, e.g. *berbetin* 'since'. Some postpositions which normally govern the genitive of pronouns also occur with the nominative, e.g. *sen üšin* 'for you'. Frequently used postpositions of converbial origin are *başlap* 'from', *alıp* 'from', e.g. *sol künnen başlap* [that day:ABL. since] 'from that day', *qarap* 'towards', e.g. *üyine qarap* 'towards his house'. Postpositions taking possessive suffixes agree with the genitive attribute, e.g. *menim keleyek aqimda* [I:GEN. come:PART. concerning:IPOSS.:LOC.] 'concerning my coming'.

Syntax

Nominal Phrases

Nominal phrases are built in the normal Turkic way, e.g. *uyalşan bir qız* [adjective attribute indefinite article noun] 'a shy girl'. Preposed relative clauses are based on participles, *-Atayan*, *-(U)wšı*, *-Ayatıryan*, *-(A)yatqan*, *-(A)yAK*. The first actant of the relative clause may be coreferential with the head of the construction as in the following examples: *eski ædetti tutatayan noyaylar* 'the Noghays who used to keep the old traditions', *şawıp keleyatıryan qart* [run:CONV. come:PART. old man] 'the old man who comes running'. The first actant may also be expressed in the relative clause, e.g. *Mašina avtobus toqtayyan yerge keldi* [car bus stop:PART. place:DAT. come:PST.] 'The car came to that place where the bus stops', *siz oqıyan francuzlar man bolyan soyıs* [you read:PART. French:PL. with.POSTP. be:PART. war] 'the war with the French about which you read'. If the first actant is in the genitive, the head noun carries a possessive suffix agreeing with it, e.g. *oniñ oqıtqan oquwşılari* [he:GEN. teach:PART. student:PL.:3POSS.] 'the students whom he teaches'. The *-GAn* participle may be used in a diathetically neutral way, e.g. *tuwyan til* [be born:PART. language] 'native language', *Tüye Töbeli degen awıl* [Tüye Töbeli say:PART. village] 'the village which is called Tüye Töbeli'.

Nominal compounds may be built without a possessive suffix, e.g. *yemis terek* 'fruit-tree', *toy keşe* 'wedding night'.

The possessor in genitive constructions may be in the nominative even if it has a specific reference, e.g. *Ayşat üyi* [Ayşat house:3POSS.] 'Ayşat's house'.

The Sentence

The normal Turkic subject-object-predicate order characterises the Noghay syntax. Topicalised constituents may be marked by *deseñ* [say:COND.:2SG.] e.g. *Predsedatel' deseñ, qarttiñ sözün maqul köretayaniñ bildirdi* 'As for the chairman, he announced that he found the words of the old man appropriate'.

The pronominal copula forms are attached to nominal predicates, e.g. *Men Noyayman* 'I am a Noghay'. The subject representative pronominal copula may be missing, e.g. *Sen tükenşi* 'You are a shop assistant'.

Non-finite Clauses

Most of the complement clauses are based on non-finite verb forms, such as participles, infinitives and verbal nouns, e.g. *Menim kelüwim iyttiñ qarawılşılıq sesin şiyardı* [I:GEN. come:INF.:IPOSS. dog:GEN. watchman:DER. sound:3POSS.:ACC. bring out:PST.] 'My coming caused the dog to produce a warning sound'. Temporal clauses are usually based on converbs expressing different temporal and aspectual meanings.

Temporal clauses may be formed with the suffix *-GAnDA*, e.g. *bular qırıya şıqqanda* 'when they leave'. There are also high-focal forms such as *-(A)yatqanda*, e.g. *ol qayıp keleyatqanda* 'when (s)he was returning'. Corresponding constructions with *zaman* 'time' are *-GAn zamanda* and *-(A)yatqan zamanda*, e.g. *ol balalardı oqıtıma baryan zamanda* 'when (s)he taught the children', *ol üyge keleyatqan zamanda* 'as (s)he was coming home'.

The converbs can also take possessive suffixes agreeing with the subject: *-(A)yatqanımda*, e.g. *ol oramnan ozayatqanımda* 'when I was passing along the street'. The form in *-sA* can also be used temporally, e.g. *Qarasa, ædemner yol yasap turi ediler* 'When (s)he looked, the men were building the way'. The meaning 'before' may be expressed by the periphrastic form *-MAy turıp*, e.g. *Ol azbarinnan şıqpay turıp, xattiñ şıyarıp qaradı* [the courtyard:3POSS.:ABL. leave:NEG.CONV. postverb:CONV. letter:3POSS.:ACC. take out:CONV. look:PST.] 'Before (s)he left his/her courtyard, (s)he took out his letter and looked at it'.

Clauses expressing causality include those based on *-GAnnAn* (*-GAnInnAn*), e.g. *Ol yüregi süyingennen, bas degende ne aytayayın da bilmedi* 'Because (s)he was so happy, (s)he did not know what to say at the beginning', *-GAnnA köre*, e.g. *Aziq ækelmege qiyin bolyanna köre, köbisinşe aš bolatayan ediler* 'Because it was difficult to bring food, most of them were hungry'. The adverbial expressions *nege dese* and *nege deseñ* 'because' can introduce a finite causal clause, e.g. *1924 yılda Peterburg V. I. Leninniñ örmetine Leningrad dep atalyan, nege dese ol bu qalada yaşayan em islegen*.

'In 1924 St Petersburg was given the name Leningrad in honour of V. I. Lenin, because he lived and worked in that town'. Juxtaposed causal clauses may be introduced by adjunctors such as *sonnan sebep* or *soya köre* 'because of that, therefore', e.g. *Ušip bolmaydı, soya köre yasirinyan* '(S)he could not fly, and therefore he hid'. Comparison is expressed with constructions of the type *Ol kelgennen kelmege ni köp iygi* [(s)he come:PART.:ABL. come:NEG.:PART.:3POSS. much good] 'His/her not coming is better than his/her coming'.

Examples of non-finite forms used in final clauses include those based on the aorist plus postposition, *-(A)r üšin*, e.g. *Ol isti qutarar üšin men de baraman* 'I will also go to settle this matter', the infinitive plus postposition, *-(U)w üšin*, e.g. *Yas balasına nizam beriw üšin şalısatayan edi* '(S)he made efforts in order to discipline the young boy'. Concessive clauses are formed with *-sa* plus the particle *DA*, e.g. *özi körmese de* 'even if (s)he does (did) not see it herself/himself'.

Optative-imperative forms may also be used in embedded clauses. In the following example the embedded clause based on the optative is followed by a postposition: *osal belgi almasın üšin, yoyarı klassqa köşsin üšin* 'so that they would not get a bad character, and so that they would pass to the higher class'.

Coordination

Nouns are coordinated with the postposition *MAn* 'with' attached to the first constituent, e.g. *Ayşat pan Murat* 'Ayşat and Murat'. *Ne . . . ne* means 'neither . . . nor', and *ya . . . yade* means 'either . . . or'.

Lexicon

The Noghay lexicon includes many lexical items copied from Arabic and Persian. The use of such elements are a particular characteristic of the cultural vocabulary before the Soviet era. Today copies of Russian items dominate. The Noghay lexicon also includes many lexical items copied from Mongolic.

Dialects

There are three main dialects of Noghay: Aq-Noghay, Qara-Noghay and Central Noghay. Standard Noghay is not based on one particular dialect, but rather consists of a mixture of features. The most important isoglosses include the following: an initial *ž* in Aq-Noghay corresponds to initial *y* in Qara-Noghay and initial *j* in Central Noghay. The Aq-Noghay participle suffix *-(I)yAK* corresponds to Qara-Noghay *-(A)yAK* and Central Noghay *-(A)jAK*.

References and Further Reading

Baskakov, N. A. (1940) *Nogajskij jazyk i ego dialekty. Grammatika, teksty i slovar'*, Moscow-Leningrad: Izdatel'stvo Akademii Nauk SSSR.

- (1963) 'Očerok grammatiki nogajskogo jazyka', in N. A. Baskakov (ed.) (1963), 497–562.
- (ed.) (1963) *Nogajsko-russkij slovar'*, Moscow: Gosudarstvennoe izdatel'stvo inostrannyx i nacional'nyx slovar'ej.
- (ed.) (1973) *Grammatika nogajskogo jazyka, I. Fonetika i morfologija*, Čerkessk: Karačaevo-čerkeskoe otdelenie Stavropol'skogo knižnogo izdatel'stva.
- Falev, P. A. (1918) 'Nogajskaja skazka ob Aq-Köböke', *Sbornik Muzeja Arxeologii i Etnografii Rossijskoj Akademii Nauk*, St Petersburg, 189–96.
- Menges, K. H. (1959) 'Die aralo-kaspische Gruppe', in J. Deny et al. (eds) *Philologiae Turcicae Fundamenta 1*, Wiesbaden: Steiner, 434–88.
- Milyx, M. K. (1947) 'Očerki po nogajskomu sintaksisu', in *Učenie zapiski Rostovskogo na Donu Gosudarstvennogo Universiteta 5, 2*, Taganrog, 5–35.
- (1949) 'Nogajskie teksty', in *Jazyki severnogo Kavkaza i Dagestana 2*: 248–95.
- Osmanov, M. (1883) 'Nogajskie i kumyckie teksty', in *Sbornik po izučeniju plemen Kavkaza 17: 3*, St Petersburg, 3–59.

21 Kirghiz

Mark Kirchner

Introduction

Kirghiz belongs to the Kipchak group of Turkic. Modern Kirghiz is closely related to Kazakh. In the nineteenth century, the Kazakhs were even called Kirghiz, and the Kirghiz Kara-Kirghiz. The language shares some features with South Siberian Turkic, but it is still unclear whether today's Kirghiz are direct successors of Old Turkic Kirghiz. In the eighteenth and nineteenth centuries, under the Khanate of Qoqan (p. 28), the Kirghiz and their language were exposed to strong influence from the Persianised Uzbek area. After the Kirghiz territory was conquered by the Russian Empire in the second half of the nineteenth century, Russian influence became predominant in all spheres of life including the linguistic one.

After the October Revolution in 1917, Kirghiz was established as a literary language, based on the northern dialects. Several Kirghiz texts had been written down by Russian scholars before that. Thus the *Manas* Epos was edited by Wilhelm Radloff in 1885. The Arabic script was introduced in 1924 but was given up after four years in favour of a unified Latin-based alphabet. A modified form of the Cyrillic alphabet has been in use since 1941. In the Soviet period, a standard language was developed that could be used in all spheres of modern life including science and politics. On the other hand, Russian became the dominant language of daily urban life, higher education, industry and administration. This situation is still valid, though Kirghiz became an official language in 1989, and Kirghizstan attained independence in 1991. Whether the exodus of Russians and other minorities will lead to a dominance of the Kirghiz language is still an open question.

Sound System

Phonetics and Phonology

Contemporary Kirghiz has fourteen vowel phonemes, eight short and six long ones. The whole range of distinctions is realised in the first syllable, whereas the number of phonemes in the following syllables is reduced because of neutralisations. The short vowels form a symmetric system with three distinct features:

Table 21.1 Short vowels

	<i>a</i>	<i>i</i>	<i>o</i>	<i>u</i>	<i>e</i>	<i>i</i>	<i>ö</i>	<i>ü</i>
back	+	+	+	+	-	-	-	-
high	-	+	-	+	-	+	-	+
round	-	-	+	+	-	-	+	+

The phoneme /e/ is realised somewhat higher in first syllables than in others. Initial /ö/ and /o/ tend to be pronounced with a prothetic bilabial glide. Unstressed high vowels may be reduced to various degrees.

The six long vowel phonemes, represented by digraphs in Kirghiz orthography, are not relics of early Turkic length, as in Khalaj, Yakut and Turkmen, but result from consonant loss in native words and from diphthongs in loanwords: *ā*, e.g. *jā-* 'rain' < *yay-*, *ē*, e.g. *bē* 'mare' (cf. Kazakh *biye*), *ō*, e.g. *tō* < *tay* 'mountain', *ō̄*, e.g. *dō̄löt* 'wealth' (cf. Arabic *daula*), *ū*, e.g. *ulū* < *uluy* 'great' and *ū̄*, e.g. *elū̄* < *ellig* 'fifty'. Long vowels have a more limited occurrence than short ones. The lack of *i* and *ī* leaves an asymmetry in the system. Though word-final *iy* and *īy* are occasionally pronounced as long vowels, they are followed by post-consonantic suffix variants, e.g. *biy-i* 'his judge'. But word-final sounds as in *kū̄* 'melody' are long vowels, as shown by the postvocalic suffix variants, e.g. *kūsū̄* 'his melody'. The corresponding Kazakh word has a final glide: *küyü* 'his melody'.

The consonant system comprises, depending on the method of analysis, at least 16 phonemes. Table 21.2 shows the main variants:

Table 21.2 Consonants

	<i>Labial</i>	<i>Alveodental</i>	<i>Alveopalatal</i>	<i>Velar</i>
stop	b, p	d, t		g, k, q
affricate			ʃ, ʧ	
fricative	v, f	z, s	ʒ, ʂ	ɣ, χ
nasal	m	n		ŋ
glide			y	
liquid		r, l, ʎ		

The phonemes /t/, /z/, /s/, /ʃ/, /m/, /n/, /ŋ/ and /t/ have clear-cut realisations. The phonemes /g/, /k/ and /l/ exhibit clearly discernible front and back variants, *g* vs. *ɣ*, *k* vs. *q* and *l* vs. *ʎ*. These variants are not distinguished in the orthography. Most stops tend towards slightly fricative realisations between vowels. This is especially valid for /b/ → *b*, *v*, but also for /d/ → *d*, *δ* (reminiscent of the interdental realisation in Bashkir). As for /p/ → *p*, *f*, the latter variant only occurs in loanwords but is represented by a special grapheme. The same is true of *χ*, which may be considered an allophone of

/k/. Initial /y/ is realised as *j* or *ž* in urban pronunciation, e.g. *jay* 'summer'. Initial /j/ sometimes becomes *y* in sandhi across word boundaries. Though Common Turkic *č* is orthographically marked as such in all environments, there is a tendency towards syllable-final neutralisation of the opposition /č/ vs. /š/.

Accent is, as a rule, on the last syllable. Exceptions include, as in many other Turkic languages, onomatopoeic and expressive elements, intensives such as *jápjašil* 'grass-green' and interrogatives such as *qáysı* 'which'. Certain suffixes and particles do not carry accent, e.g. the particles *Go*, *da* 'too', copula suffixes, the interrogative *-BI*, the emphatic suffix *-čI*. In some cases, accent may disambiguate otherwise homophonous word forms, e.g. *qóyču* 'let it be' vs. *qoyčú* 'shepherd'.

Phonotactics and Morphophonology

Besides the general rule that suffix variants with voiced initial sounds are added to voiced stem-final sounds and *vice versa*, there are important morphophonemic alternations in suffixes with initial *l* and *n*. The suffix-initial consonant remains unassimilated after final vowels, but is *t* after voiceless consonants, and *d* after voiced consonants, *l* being sometimes preferred after *r* and *y*. Suffixes with initial *N*- include the genitive suffix *-NIn*, e.g. *balanin* 'of the child', *qardin* 'of the snow', *attin* 'of the horse', and the accusative suffix *-NI*, e.g. *balani* 'child', *qardi* 'snow', *atti* 'horse'. Suffixes with initial *L*- include the plural suffix *-LAr*, e.g. *terezeler* 'windows', *kündör* 'days', *attar* 'horses', and the denominal adjective suffix *-LÜ*, e.g. *balalü* 'having children', *jeldü* 'windy', *attü* 'having a horse'.

Strong sound harmony is one of the features that bind Kirghiz closer to some Siberian Turkic languages. The distinctions front vs. back and rounded vs. unrounded are mostly neutralised in non-first syllables. As for the vowels, there is then only a low archiphoneme /A/, represented by /a/, /e/, /o/, /ö/, and a high /I/, represented by /i/, /ı/, /u/, /ü/. The choice of the representatives is determined by features of the preceding syllable, e.g. *joldoštor* 'friends' ← *jol* 'way' + *-LAš* + *-LAr*, *Berbeyşiŋbi?* 'Do you not give it?' ← *ber-* 'give' + *-BA* + *-y* + *-sIŋ* + *-BI*, *Tüšünböğön* '(S)he has not understood' ← *tüšIn-* 'understand' + *-BA* + *-GAn*.

The vowel harmony system has considerable asymmetries. Thus, *u* in the preceding syllable is not followed by *o*, e.g. *učunda* 'on its top', *ürda-* 'steal', and *ö* is followed by *a* in several derivational suffixes, e.g. *özan-* 'talk' (cf. *özdön* 'from the mouth'). Certain long suffix vowels only have rounded variants, e.g. the infinitive in *-Ü* → *-ü*, *-ü* and the collective numeral suffix *-Ö* → *-ö*, *-ö*. The rounded vs. unrounded harmony rules are not applied consistently in spoken Kirghiz. The written form *үйлөндүм* 'I got married' ← *üy* 'house' + *-LAN* + *-DI* + *-m*, may, like the corresponding Kazakh form *үйлендім*, be pronounced *üylöndüm* or *üylendim*. The different orthographic vowel representations thus conceal the similarities between Kirghiz and Kazakh.

Loanwords often deviate from the harmony schemes: a word such as *Özbekstanda* 'in Uzbekistan' contains rounded and unrounded as well as front and back vowels.

Morphology

Nouns and Adjectives

Word Formation

Suffixes forming denominal nominals include the highly productive adjective suffix *-LÜ*, e.g. *aqıldü* 'intelligent' ← *aqıl* 'intellect', *küçtü* 'strong' ← *küç* 'strength' and *-LIK*, which primarily form abstract nouns, e.g. *balalıq* 'childhood' ← *bala* 'child', but also adjectives (often neologisms) meaning 'belonging to', e.g. *industriyalıq* 'industrial' ← *industriya* 'industry'. A frequent suffix in neologisms is *-čIl*, denoting propensity to something, e.g. *ulut* 'nation' ← *ulutčul* 'nationalist'. The augmentative suffix *-(I)rÄK* is added to adjectives to express comparison, often within ablative constructions, e.g. *Bal qanttan tattüräq* 'Honey is sweeter than sugar'.

A number of suffixes derive nominals from verbs: the productive suffix *-mA*, forming nouns, e.g. *körgözmö* 'exhibition' ← *körgöz-* 'show', and *-GIč*, used in many neologisms, denoting entities capable of performing the action expressed by the verb, e.g. *učquč* 'pilot' ← *uč-* 'fly', etc.

Declension

The plural suffix is *-LAr*, e.g. *jigitter* 'young men', *tölor* '(several) camels'. Nouns used in a collective sense are not marked for plural, e.g. *tö* 'camel, camels'.

Possessive suffixes hold the second position in the nominal morpheme chain:

Table 21.3 Possessive suffixes

	Singular	Plural
1p.	-(I)m	-(I)lız
2p.	-(I)ŋ	-(I)ŋAr
3p.	-(s)I	-LArI

Case suffixes hold the third position. The basic paradigm, applied to nouns without possessive suffixes, is shown in Table 21.4.

Nouns with possessive suffixes show minor deviations from this scheme. *-A* is the allomorph of the dative suffix after first- and second-person singular possessive suffixes. Allomorphs with a 'pronominal *n*' after third-person

Table 21.4 Case suffixes

nom.	∅
gen.	-NIn
acc.	-NI
dat.	-GA
loc.	-DA
abl.	-DAn

possessive suffixes are dative *-nA*, accusative *-n*, locative *-ndA* and ablative *-nAn*.

Pronouns

The inflection of the singular personal pronouns differs from the basic case paradigm given above.

Table 21.5 Personal pronouns

	<i>1p.</i>	<i>2p.</i>	<i>3p.</i>
nom.	men	sen	al
gen.	menin	senin	anin
acc.	meni	seni	ani
dat.	maya	saya	aya
loc.	mende	sende	anda
abl.	menden	senden	andan

The plural forms are *biz* 'we', *siz* (polite to one addressee), *siler* (plural), *sizder* (polite to more than one addressee) 'you', *alar* 'they'.

The inflection of demonstrative pronouns is similar. Most of them have an optional *-l* in the nominative. The pronoun *bul* 'this' has initial *m* in all oblique cases except the dative. *Uşul* is used for close objects known to the speaker, *ošol* for remote and known objects, and *tigil* for objects outside the conversational setting.

Pronouns also include the reflexive *öz*, the interrogatives *emne* 'what', *kim* 'who', *qaysi* 'which', etc.

Numerals

The cardinal numerals for digits are *bir* 'one', *eki* 'two', *üç* 'three', *tört* 'four', *beş* 'five', *altı* 'six', *jeti* 'seven', *segiz* 'eight', *toyuz* 'nine'. The decades are expressed with *on* 'ten', *jüyirma* 'twenty', *otuz* 'thirty', *qırq* 'forty', *elü* 'fifty', *altümiş* 'sixty', *jetimiş* 'seventy', *seksen* 'eighty', *toqson* 'ninety'. Further opaque numbers are *jüz* 'hundred', *miñ* 'thousand'. As in other Turkic languages, complex numbers are formed by compounding, the highest decimal numeral being the leftmost one, e.g. *jüyirma jeti miñ beş jüz toqson bir* '27591'.

Ordinal numbers are formed with the suffix *-(I)nčI*. A special collective suffix *-(θ)ō* (with drop of stem-final vowels) is used for the numbers 1 to 7, e.g. *ekō* 'two together'. Distributives are formed with the ablative, e.g. *Ar bir gruppada otuzdan student bar* 'There are thirty students in each group'. Approximative and multiplicative numbers can be expressed with *-DAy*, e.g. *on miñdey bala* 'about ten thousand children', *-čA*, e.g. *otuzča* 'about thirty', *-LAGAn*, e.g. *jüzdögön* 'hundreds'.

Copulas

The following copula suffixes are used to mark person and number in nominal sentences:

Table 21.6 Copula suffixes

	<i>Singular</i>	<i>Plural</i>
<i>1p.</i>	-mIn	-BIz
<i>2p.</i>	-sIñ	-sIz
<i>3p.</i>	∅	∅

The second-person plural suffix *-sIz* is used for polite address. The simple plural form is *-sIñAr* (*<-sIñ + LAr*), whereas *-sIzdAr* is a polite form for more than one addressee.

Verbs

Denominal Verb Derivation

Most suffixes deriving from verbal stems have equivalents in other Turkic languages, but may differ from them in phonological and morphophonological respects. Thus, the widespread denominal suffix *-LA* has twelve variants in Kirghiz, e.g. *oroqto-* 'cut with a sickle' ← *oroq* 'sickle'.

Among suffixes deriving from deverbal verbs, the voice suffixes fill the second slot in the verbal morpheme chain: passive *-(I)l*, reflexive *-(I)n*, cooperative-reciprocal *-(I)š*, causative *-DIr*, *-Ar*, *-(I)t* etc. One major difference is that *-(I)š* is also used to indicate the third-person plural of verbs (see p. 42). The next morpheme in the chain is the verbal negation suffix *-BA*.

Like many other Turkic languages, Kirghiz displays actional constructions in which converbs of verb lexemes combine with a limited set of auxiliary verbs (see pp. 113–14). Thus, *-A al-* 'take', often contracted to *-al-*, denotes ability, e.g. *Oquy alam* 'I can read'. The constructions *-(I)p al-* and *-(I)p ber-* signal orientation towards and away from the first actant, respectively, e.g. *Qolunđu jüp alči* 'Please wash your hands', *Bu qatti oqup ber* 'Read this letter (to someone)'. Continuity up to a given time may be indicated by *-(I)p kel-*, e.g. *Bir neče jıldan beri joldoš bolup keldik* 'We have been friends for a

number of years'. Transformation, with nuances of completion and suddenness may be expressed with *-(I)p ket-* and *-(I)p sal-*, e.g. *Qizarip ketti* '(S)he (suddenly) turned red', *Unutup saldim* 'It slipped from my mind'. Non-transformation with durative and similar actional meanings is expressed with *-(I)p* plus *jat-* 'lie', *otur-* or *oltur-* 'sit', *jür-* 'go' and *tur-* 'stand'.

Thus, *-(I)p tur-* may express repetition, e.g. *Anda sanda alardan qat kelip turat* 'Sometimes letters from them arrive', and *-(I)p jür-* long-lasting continuity, e.g. *Al Biškekte oqip jüröt* '(S)he goes to the university of Bishkek'.

Finite Verb Forms

Kirghiz has two sets of endings to indicate person and number of finite verbs, one of pronominal and one of possessive origin. The latter type of personal marker is only used with the simple past and the conditional.

The third person is indicated by *-(I)š*, formally identical with the reciprocal suffix. It has maintained its place in the verbal morpheme chain, e.g. *Kelišti* 'They came'.

The richly developed system of aspects, tenses and moods shares some basic properties with the Kazakh system, though it often conveys different shades of meaning.

A general intraterminal present is formed with *-A* + pronominal personal markers. In the first-person singular, the short form *-m* is preferred, e.g. *Šarda turam* 'I live in the city'.

A more focal present with a narrower focus on the ongoing event is formed with *-(I)p* + present forms of *jat-* 'lie', *otur-* 'sit' etc. + pronominal personal markers, e.g. *Ištep oturmun* 'I am working (just now)', *Čurqap jatišat* 'They are running', *Oylop jatam* 'I am thinking'. With the lexical verbs *bar-* 'go' and *kel-* 'come', *-A* converbs are preferred, e.g. *Bolot bara jatat* 'Bolot is coming'. There are similar constructions with *tur-*, *jür-* and *otir-*, with different semantic nuances. A similar form, *-ŪdA* + pronominal personal markers, is predominantly used in formal written style, e.g. *Ündör uyulūda* 'Sounds are (being) heard'.

A simple past is formed with *-DI* + possessive personal markers, e.g. *Jürgünčülör bulaqtan sū içišti* 'The travellers drank water from the well'. A

Table 21.7 Personal markers

	Pronominal type	Possessive type
1p.sg.	-mIn	-m
2p.sg.	-sIn	-ŋ
3p.sg.	Ø, -t	Ø
1p.pl.	-blz	-k
2p.pl.	-slz(dAr), sInAr	-ŋlz(dAr), -ŋAr
3p.pl.	-Iš; -IšAt	-Iš

post-terminal with perfect functions (see p.44) is formed with *-GAn* + pronominal personal markers, e.g. *Bolottun atı ašūda qalyan* 'Bolot's horse has stayed at the pass'. The negation is formed with *-BA* or with *joq* or *emes*. The short form *-GAm* is sometimes used for the first-person singular. The *-GAn* form may also suggest remoteness and inference. The suffix *-DIr* adds a presumptive meaning, e.g. *Ūlum oyyonyondur* 'I guess my son is awake'. An indirective past (see p.45) is formed by *-(I)ptIr* + pronominal personal markers, e.g. *Elge tünčtiq berbeptir* '(S)he apparently did not give peace to the people'. The negation is formed with *-BA*. A frequently used habitual or durative past is formed with *-čI* + pronominal personal markers, negated with *emes*, e.g. *Ušul jerge kelčübüz* 'We used to visit that place'.

A modal and prospective item is formed with *-(A)r* + pronominal personal markers, negated with *-BAs*, e.g. *Jolun tabarbız* 'We will find the way', *Ketpesmin* 'I will not go'. Intention is signalled by *-MAKčI* + pronominal personal markers, negated with *emes*, e.g. *Jazmaqčimin* 'I want to write'.

Several simple tenses combine with the past copula particle *ele* and the indirective copula particle *eken*. For example, *-(A)r ele* forms an intraterminal past (imperfect), e.g. *Tigi döbödö bir jaman tam bolor ele* 'There was a shabby hut on that hill', and *-čU eken* expresses an indirective habitual past, e.g. *Tinččiliq kezinde bu jerde miñ adam iştečü eken* 'A thousand men are said to have worked at this place in peacetime'. Combinations with *ele* can also express counterfactual content, e.g. *Kečē ele kelet ele* '(S)he would have come yesterday'.

There is a special optative-imperative paradigm. All forms are negated with *-BA*. Second-person forms in *-GIn* and *-GIIA* add the connotation of a wish or a request.

Non-finite Verb Forms

One of the productive verbal noun suffixes is *-(Ø)Ū ~ -(Ø)Ö*, which forms action nouns used as the lexicon form of Kirghiz verbs, e.g. *barū ← bar-* 'go', *biylō ← biyle-* 'dance'. (Stem-final vowels fuse with the suffix vowel, determining its height.) Verbal nouns denoting the manner of performing an action are formed with *-(I)š* forms, e.g. *araba aydaši* 'the way (s)he drives the cart'.

The *-GAn* participle has a wide range of functions and may refer to anterior as well as non-anterior events. Constructions such as *-(I)p + jat-* + *-GAn* are intraterminal, referring to ongoing events, e.g. *kele žatqan kiši* 'the person just

Table 21.8 Optative-imperative

	Singular	Plural
1p.	-(a)yIn	-(A)II(K)
2p.	Ø, -GIn	-(I)ŋlz(dAr), -GIIA
3p.	-sIn	-(Iš)sIn

coming', e.g. *nari jaqta iştep jatqan kelinder* 'the young women working over there'. A durative or habitual participle is formed with $-(\bar{U})\check{C}U \sim -(\bar{O})\check{C}U$, e.g. *bizdin mektep turçu döbö* 'the hill where our school used to stand'. The aorist participle in $-(A)r$, negated $-BAs$, has modal or prospective reference, e.g. *Kim bolorun kim bilet?* 'Who knows what (s)he will be?' The item $-A elek$ denotes an event that has not yet taken place. Though its last element is written separately, this marker is pronounced as a harmonic suffix, e.g. *bütölök* 'unfinished'.

The most frequent converb is the one in $-(I)p$. Stem-final labial consonants may be dropped, causing contraction in the length of the vowel, e.g. *tāp* ← *tap-* 'find'. The $-A$ converb is formed with $-A$ after consonants and with $-y$ after vowels, negative $-Bay$. Both converbs have various functions in complex verbs and complex clauses. There are also terminative, abtemporal, purposive and other converbs, e.g. $-GAnčA$, $-GInčA$ 'until', 'instead of', $-GAnI$, $-GAlI$ 'since', 'in order to'.

Adverbs

Adverbs are a weakly delimited and formally mixed class. The most common adverb-forming suffix is $-čA$. A number of lexemes are primarily adverbs, e.g. *bat* 'fast' and *emi* 'now'. There are also fossilised forms with case suffixes and converb suffixes. Some adverbs are based on pronouns, e.g. *qanday* 'how', *qačan* 'when'. Superlative of adjectives is formed with the preposed element *eŋ*, e.g. *eŋ jaqši* 'best'.

Other Morphological Categories

Among the simple postpositions, *menen* 'with' governs the nominative, and the accusative of certain pronouns, e.g. *atası menen* 'with his father', *meni menen* 'with me'. The postpositions *arqilū* 'through' and *siyaqtū* 'like' govern the nominative, e.g. *Joldoşum arqilū bildim* 'I learned it from my friend', *biz siyaqtū jigitter* 'young men like us'. Others, such as *deyre* and *čeyin* 'up to, until' govern the dative, e.g. *sāt jetige deyre* 'until seven o'clock', *tōyo čeyin* 'up to the mountains'. Some, such as *kiyin* 'after' govern the ablative, e.g. *elū jıldan kiyin* 'after fifty years'. A set of complex postpositions based on possessive nominal forms indicate temporal and spatial relations (see pp. 48, 112–13, 221–2, 370–1), e.g. *üydün aldına* 'to the front of the house'.

Since Iranian influence on Kirghiz has not been very intensive, the system of conjunctions is only weakly developed. Though *jana* may often be translated as 'and', its use is rather restricted. Elements such as *al* and *biroq* 'but', traditionally described as adversative conjunctions, tend to function as conjunctive adverbs.

Enclitic particles fulfil important functions in the Kirghiz sentence. A very common one is the interrogative $-BI$. The particle *da* 'also' is, unlike its counterparts in other Turkic languages, non-harmonic. The restrictive particle *ele* – not to be confused with the past copula particle *ele* – is also frequent, e.g. *Üč ele kün qaldı* '(S)he only stayed for three days'.

Syntax

Nominal Phrases

Nominal compounds of the type *jekşembi künü* 'Sunday', sometimes lack the possessive suffix on the second member, e.g. *jıyač usta* 'carpenter'.

In genitive constructions, the attribute carries the genitive suffix, whereas the head is marked with a third-person possessive suffix, e.g. *şardıñ borboru* 'the centre of the town'.

The Sentence

Finite Clauses

The unmarked constituent order is subject, object, predicate. Since constituents are mostly case-marked, the basic order can be changed for topicalisation or contrast, e.g. *Sayıñdım žerimdi* 'I longed for my native place'.

Plural agreement of subjects and predicates is observed more often than in other Turkic languages, and may be due to the formal difference between nominal and verbal plural suffixes, e.g. *Baldar kelišti* 'The children came'.

Since person is normally marked on the predicate, personal pronouns are used mainly in contrasting or focusing functions.

In nominal sentences, existential statements can be expressed without copulas, whereas the negation copula is *emes* 'is not'. Possession may be expressed with *bar* 'existing', negated *joq* 'non-existing'. Elements denoting the possessed entity carry possessive suffixes referring to the possessor, e.g. *Atı bar* '(S)he has a horse'.

The infrequent conjunction *jana* mostly coordinates clause constituents. Clauses may be coordinated with simple juxtaposition.

Non-finite Clauses

Relative clauses and complement clauses are left-branching and based on various verbal nominals.

Relative clauses are embedded in front of their heads, e.g. *bizge eškiñi açqan kişi* 'the person that opened the door for us'.

Complement clauses are embedded in various syntactic functions by means of participles. In *Jō jürgöñdön tō mingen jaqši* 'It is better to mount a camel than to walk', the subject clause is based on the nominative participle *mingen*, and the adverbial clause on the ablative-marked *jürgöñdön*. The object clause *Menin mında ekenimdi bilbeyt* '(S)he does not know that I am here' is based on an accusative-marked participle. The construction $-GI +$ possessive suffix + *kel-* 'come' in a third-person form expresses the wish to carry out an action, e.g. *Baryñbiz kelet* 'We want to go'.

Non-modifying $-(I)p$ converbs may be used for clause chaining, e.g. *Men erteñ menen turup zaryadka žasap kiyinip žünup mektepke baražatam* 'I get

up in the morning, do gymnastics, dress, wash myself and go to school' (Imart 1981: 600).

The semantics of simple converbs are often vague. More specific semantic relations are expressed by adverbial clauses embedded by means of complex converbs, non-finite forms, often based on lexical elements, with case markers and postpositions. Clauses expressing simultaneousness may be formed with *-GAndA*, e.g. *Asan közün açqanda töşöktö jatqanday sezdi* 'When Hasan opened his eyes he felt as if he was lying in the bed'. More complex forms are based on a noun denoting 'time', *kez* or *ubaq* plus participles, e.g. *Belge çiya bergen kezde qar jādī* 'When they reached the pass, it snowed'. The same construction with *zamat* 'moment' expresses immediate succession, e.g. *Arabadan tüşer zamat Asan qaptardī taşiy baštadi* 'As soon as Hasan got down from the carriage, he started to carry the sacks'. This can also be expressed by *-Ar(I) menen*, e.g. *Oqū büütör menen oqūčular dem alişqa ketiştī* 'As soon as school ended, the pupils went on holiday'.

Concessive clauses are formed with *-GAn menen*, e.g. *ot jayılyan menen* 'though the fire was lit'. Purposive clauses are expressed with *-GANI*, *-(I)ş üčün*, *-Ü üčün* etc., e.g. *Jaştar bilim aliş üčün bardıq şarttar tüzülgön* 'All conditions are met in order for the youth to acquire knowledge', *Oquyani keldi* '(S)he came to study'. Causal clauses may be expressed with *-GAndAn*, e.g. *Uyalyanımdan men žer qarap qaldım* 'I looked at the ground because I was ashamed', *-GAN sebeptü*, e.g. *Bolot uyalyan sebeptü qizarıp ketti* 'Bolot blushed because he was ashamed', etc. Conditional clauses are formed with *-sA* plus possessive personal markers. The conditional relation may be emphasised with *eger* or *egerde* 'if', a particle copied from Persian, e.g. *Eger maqul bolsoñ aytam* 'I will tell you, if you agree'. Clauses expressing 'instead of doing' may be formed with *-GAndAn körö*, e.g. *beker жүргөндөн körö* 'rather than running around uselessly'.

Lexicon

Common Kipchak lexemes constitute the basis of the Kirghiz lexicon. It contains several archaic words, e.g. *uy* 'cow' (Old Turkic *ūd* 'ox') instead of *siyir* etc. in Kazakh and other modern Turkic languages. Some widespread words are of obscure origin, e.g. *borbor* denoting the central point of a circle and used in the literary language for 'centre, capital'. Besides many older common Turkic-Mongolic words, numerous Mongolian loans from the Middle Ages can be identified, e.g. *sonun* 'remarkable', *dülöy* 'deaf'.

Arabic and Persian lexical elements, copied via Chaghatay and Uzbek, form a considerable part of the lexicon. They cover all fields of traditional Islamic society and have partly superseded the Turkic vocabulary. They have generally adapted more strongly to the native phonological system than their counterparts in neighbouring Turkic languages, e.g. *jöp* 'answer' << Arabic *jawāb*, *insap* 'fairness, justness' << Arabic *insāf*. The northern dialects, on

which the literary language is based, are less influenced by Islamic vocabulary.

Russian provides the most recent layer of loanwords. While in the literary language, the use of Russian elements is restricted to the domains of European civilisation, imitations from Russian are more dominant in informal spoken Standard Kirghiz. In the speech of non-bilinguals, the phonetic shape of Russian loanwords is mostly adapted to the Kirghiz system. Thus, the former name of the Kirghiz capital is written *Frunze*, but often pronounced *Purunza*. Many lexemes are copies of Russian ones, e.g. *birdiktülük* modelled on *edinstvo* 'unity' and *baldar baqčasi* modelled on *detskij sad* 'kindergarten' (with *baqča* 'garden' < Persian).

Efforts have been made in the post-Soviet period to reduce the immense amount of Russian technical terms by creating neologisms on the basis of Turkic and even Arabic-Persian lexical material.

Dialects and Variation

The Kirghiz dialects can be divided into a southern and a northern group, the latter forming the base of the literary language.

The dialects of the south, mainly those spoken in the Ferghana basin, show different degrees of Uzbekisation. Some characteristics that separate Kirghiz from other Turkic languages are absent there. Thus *-LAr* is used in the third-person plural verb forms, e.g. *Aytılar* 'They said' instead of *Aytıştı*. Long vowels of the standard language are often represented by diphthongoids as traces of the lost γ , e.g. *tow* 'mountain' instead of *tō*. The past copula particle *ede*, which is more similar to *edi* in Uzbek and other Turkic languages, is used instead of *ele*. Uzbek influence has led to partial loss of the 'pronominal *n*' and other changes in the nominal inflection, e.g. *balasini* 'his child (accusative)' instead of *balasın*. There is an æ in the neighbourhood of front consonants and in Arabic and Persian loanwords, e.g. *æger* 'if' instead of *eger*. In the southern dialects, lexical elements copied from Arabic and Persian via Uzbek preserve their original phonetic shape to a high degree.

The most interesting feature of the northern dialects is the loss of a clear contrast between /s/ and /z/, e.g. *qısil* 'red' instead of *qızıl* and *balazi* 'his/her child' instead of *balası*. This is one of the properties that draw Kirghiz closer to the South Siberian Turkic languages.

The written language of the Kirghiz living in China is largely oriented towards the norm used in Kirghizstan. Nevertheless, the different script and numerous Chinese loanwords have led to the emergence of a specific variety of Kirghiz in China. (For the Fu-yü language, see pp. 10, 83.)

References and Further Reading

Hebert, R. J. and Poppe, N. (1963) *Kirghiz Manual*, Indiana University Publications, Uralic and Altaic Series 33, The Hague: Mouton.

- Imart, G. (1981) *Le kirghiz* 1–2. Aix-en-Provence: Publications de l'Université de Provence.
- Judaxin, K. K. (1965) *Kirgizsko-russkij slovar'*, Moscow: Sovetskaja Ėnciklopedija.
- Wurm, S. (1949) 'The (Kara-)Kirghiz language', *Bulletin of the School of Oriental Studies* 13: 97–120.
- Žapar, A. (1992) *Sintaksičeskij stroj kyrgyzskogo jazyka* 1–2, Biškeek: Mektep.
- Zaxarova, O. V. (ed.) (1987) *Grammatika kirgizskogo literaturnogo jazyka 1: Fonetika i morfologija*, Frunze: Ilim.

22 Uzbek

Hendrik Boeschoten

Introduction

Uzbek, both in terms of cultural importance and number of speakers, is the second Turkic language after Turkish. As the latter is the successor of Ottoman Turkish, Uzbek as a literary idiom is the continuation of Chaghatay in what was once the focal area of use for that written language (see Chapter 9).

Contact with other languages has been, and still is, important for a number of features of Uzbek. Among the other languages of Uzbekistan, Tajik is mainly spoken in the oases of Bukhara and Samarkand, and in the Ferghana Valley, while Russian is mainly spoken in the capital Tashkent.

The language situation is extraordinarily intricate due to the complex ethnic make-up of the Uzbek nation. Uzbek has inherited the mixed character of its literary predecessor Chaghatay. The Kipchak Uzbeks were the latest group to arrive in Central Asia, but have been dominant politically ever since. The original Turkic population has sustained closer ties with the Tajiks, and their brand of Turkic has made the largest contribution to the structure of modern Uzbek. A third linguistic grouping are the Oghuz Uzbeks in Khwarezm and adjacent areas in Karakalpakistan and Turkmenistan. All these varieties function in relationship to one literary language, modern standard Uzbek.

Uzbek has several features, notably phonological properties, in common with modern Uyghur. It is also rather close to Kazakh with respect to a number of synchronic features.

Uzbek has inherited from the Chaghatay period several focal areas, e.g. the oases of Samarkand and Bukhara, Tashkent and the Ferghana valley. Even in a description of standard Uzbek, it is thus difficult to present, for example, the pronunciation in a uniform way. In the present survey, the language material will be presented essentially in transliteration of the written standard, with some extra indications of phonetic characteristics. Dialects spoken outside Uzbekistan will not be dealt with here.

Apart from the existing variational patterns in spoken Uzbek, political developments in the 1990s have led the language into a new transitional period. The new Latin alphabet enacted by law in 1993 suggests that the new

attitudes are, on the whole, conservative, as it essentially represents a transliteration of the Cyrillic spelling. After earlier proposals, those responsible finally seem to have settled on a Latin alphabet with digraphs, only using ASCII symbols (see Appendix, Table 1, p. 454).

Sound System

Phonetics and Phonology

Many problems of Uzbek phonology still await thorough treatment in a modern framework.

The vowel system is particularly difficult to describe. Modern spelling applies a system of six vowel signs, identical to the system used for Tajik. It thus does not reflect the fact that the distinctions between back and front vowels have been largely preserved, notably in *u* vs. *ü* and *o* vs. *ö*.

The oppositions between back and front syllables are traditionally accounted for in terms of pairs of back and front consonants: *q* vs. *k* and *ɣ* vs. *g*. A corresponding pair *t* vs. *l* is mostly not assumed, since the dominant spoken varieties have not retained this opposition. It has been postulated that the phonemes /a/, /h/, /o/, /u/ would be realised as back when occurring with *q* and *ɣ*, and as front when occurring with *k* and *g*. This may, in some varieties, work out for *i* vs. *ï*, or even for *a* vs. *ä*. But it certainly does not for *u* and *o*, since there are minimal pairs such as *bol* - 'become, be' vs. *böl* - 'divide' and *uč* 'extremity' vs. *üč* 'three'. Such pairs are homographic in modern spelling.

The phonetic realisations of *u* vs. *ü*, *o* vs. *ö*, *a* vs. *ä* and *i* vs. *ï* tend to vary greatly on a basically overlapping continuum. The qualities of *a* vs. *ä* and *i* vs. *ï* are particularly variable. Thus, *ä* is typically realised as [ɛ] and [æ]. The vowel *a* ranges from a back [ɑ] to centralised variants. There is also a labialised back [ɒ] that will be transcribed with *â* in this chapter. High unrounded vowels are realised as [i] if they are adjacent to velar or uvular consonants. They are lowered considerably in open syllables, and reduced or lost in closed syllables before the continuants *š*, *s*, *r*, *l*, *z*, *n*, e.g. *bir* [b^hr] 'one', *kiši* [kfɪ] 'man'. In the present survey, *o* vs. *ö*, *u* vs. *ü*, and (tentatively) *a* vs. *ä* will be distinguished, while *i* vs. *ï* will not.

Many spoken varieties of Uzbek exhibit a mixture of the common Turkic and the Tajik system. Dialects and local standard-like varieties vary considerably in this respect. The Turkic system assigns the features front vs. back to primary stems, and may, by application of intersyllabic harmony, apply them to words. The Tajik system is characterised by a distinction between the 'stable' vowels *é*, *â*, *o* and the 'unstable' vowels *i*, *u*, *a*, a distinction reflecting an old opposition between long and short vowels. Due to the overall variability of the vowel system, it is not possible to treat the manifestations of vowel harmony in the present survey.

Traditional accounts of the consonant system may be presented as follows:

Table 22.1 Consonants

	Labial	Prepalatal	(Post)palatal	Velar, uvular, glottal
stop	p, b	t, d	k, g	q
fricative	f, w	s, z	š, ž	ɣ, χ, h
nasal	m	n		ŋ
affricate			č, j	
glide			y	
liquid		l, r		

Voiceless stops are mostly aspirated. Besides the opposition voiced vs. unvoiced, the distinction between weak and strong consonants – lenes (weak) and fortes (strong) – is also important. The velar *ŋ* is, depending on the background of the speaker, often produced with a following *g*-sound, e.g. *köŋ^gil* 'heart', *yâŋ^gâq* 'walnut'. Many speakers pronounce *k* and *q* as fricatives.

The spelling of loanwords is mostly etymological. Arabic–Persian elements reflect the original forms as written with Arabic script. Some Cyrillic letters, e.g. *ц* and *ш*, only occur in Russian loanwords. No separate symbols exist for the sounds *c* and *šč* in the new Latin alphabet.

Morphophonology

The realisations of suffixes are highly regular. The standard spelling is basically morphological and thus normally does not indicate vowel harmony or consonant assimilations. Only the causative derivation gives rise to a system of allomorphs.

Like other Turkic languages, consonants are devoiced after voiceless consonants. This assimilation is mostly represented in the spelling of word-formation suffixes added to native stems, e.g. *türtki* 'push' vs. *sêwgi* 'love' and *yâtqiz* - 'cause to lie down' vs. *kirgiz* - 'bring in'. Inflectional suffixes are, however, mostly spelled invariably, e.g. *мактабда* 'at school', *ишга* 'to work', *ичди* 'drank', even if the suffix-initial consonants are devoiced. Exceptions include suffixes with initial -*G*, e.g. the dative marker and the participle suffix -*Gän* (-*gän*, -*kän*, -*qän*).

After the consonants *k*, *q*, *g*, *ɣ*, stem-final *g* and *ɣ* are devoiced regressively, e.g. *qulâqqa* 'into the ear' ← *qulâq* 'ear', *tâqqa* 'the mountain' ← *tây* 'mountain', *bâqqan* 'looking' ← *bâq* - 'look', *sâqqan* 'milking' ← *sây* - 'milk', *tékkän* 'reaching' ← *têg* - 'reach'.

The assimilations are obscured in the written forms of Arabic–Persian elements, e.g. *нисбат* 'relation', *тадқиқ* 'investigation', *шубҳа* 'doubt' for *nispât*, *tâtiq*, *šüphâ*.

The opposition voiced vs. unvoiced is not generally neutralised in word-final position. Intervocalic lenes are voiced, and also often further weakened

by fricativisation. The morphophonological rules for the distribution of fortes and of voiced and voiceless lenes is partly sensitive to a distinction between mono- and polysyllabic stems, e.g. *yük* 'burden' → *yüki* 'its burden', *eshik* 'door' → *eshigi* 'its door' (see pp. 32, 98).

Sandhi phenomena are common, e.g. *on bir* 'eleven' > *ombir*, *mümkün mi?* 'is it possible?' > *mümkümmi*, *sätib äldi* 'bought' > *sätiwäldi*, *kéräk emäs* 'is not necessary' > *kérägémäs*.

Orthography

The Cyrillic alphabet has been in use since the 1940s. The new Latin script was enacted by law in 1993, but is only slowly being put into use. It preserves the principles of the Cyrillic one and thus is by no means a return to the Latin alphabet of the 1930s. One other proposal under discussion is the replacement of non-ASCII symbols with digraphs, e.g. <sh> for š.

Morphology

There are two basic morphological classes in Uzbek, verbs and non-verbs. At the functional level, nouns, pronouns, adjectives, adverbs and postpositions can be distinguished.

Word formation will not be dealt with here. Derivational suffixes can be found in morphological appendices to dictionaries (e.g. Waterson 1980).

Nouns and Adjectives

Declension

The plural ending is *-lar*. The possessive suffixes are 1p.sg.-(*i*)m, e.g. (*mëniñ*) *äpam* 'my aunt', (*mëniñ*) *üyim* 'my house', 2p.sg.-(*i*)ñ, (*sëniñ*) *äpañ*, (*sëniñ*) *üyiñ*, 1p.pl. -(*i*)miz, e.g. (*bizniñ*) *äpamiz*, (*bizniñ*) *üyimiz*, 2p.pl.-(*i*)ñiz, e.g. (*sizniñ*) *äpañiz*, (*sizniñ*) *üyiñiz*, 3p.sg. -(*s*)i, e.g. *äpasi* 'his/her aunt', *üyi* 'his/her house', 3p.pl. *äpalari* 'their aunt', *üylari* 'their house'. Possessor complements take genitive markers, e.g. *Kämälniñ äpasi* 'Kamal's aunt', *qoñsilärniñ üyi* 'the neighbours' house'.

Case markers are suffixed on the head of nominal phrases marking arguments of verbs or serving to form complement phrases, locative phrases, temporal phrases, etc. The genitive suffix marks complements of nominal phrases. Uzbek has no 'pronominal *n*' in the declension of nouns with a third-person possessive suffix, e.g. *üy-i-dä* 'in his house' (cf. Turkish *ev-i-n-de*).

The unmarked nominative is the default case, serving various functions from indicating the subject to forming adverbial complements. The accusative suffix *-ni* marks specific direct objects. The genitive suffix *-niñ* is mostly pronounced as *-ni*, thus coinciding with the accusative. This is reflected in the spelling of free possessives such as *bizniki* < *bizniñki* 'belonging to us'. The dative suffix *-Gä* marks indirect objects, as in *Kitäbni sizgä äldim* 'I bought the book for you', and directional complements as in *Akäm Moskvagä këtädi*

'My brother has gone to Moscow'. Temporal complements as in *Bëš küngä mühlät sorädi* '(S)he asked for five days' respite' are less frequent.

The locative suffix *-Dä* serves to form locative and temporal complements. *-gi* is used to transform such phrases into adjective phrases, e.g. *mäktäb sahnidägi öriklär* 'the apricot (trees) in the school yard' ← *mäktäb sahnidä* 'in the school yard'. The ablative suffix *-Dän* expresses local or temporal points of departure and sources, e.g. *Èrgäš mäktäbdän keldi* 'Ergash came from school', *Miltiqtäñ ätdi* 'He shot from a rifle', *Mën uni qadimdän bilämän* 'I know him from old times'.

There are also good arguments for considering the equative suffixes *-däy* and *-čä* case markers. The suffix *-däy* – with the variants *-däy* and *-däk* – is used for qualitative comparison, e.g. *Šundäy bolsin!* 'Let it be thus!' A counter-argument might be provided by its use with pronouns carrying a genitive suffix, e.g. *sunig-däk köp sözlär bilän* 'with that many words'. The equative in *-čä* is used for quantitative comparison and to indicate manner, e.g. *ät räfličä* 'extensively', *dostičä* 'friendly'. Its status as a case marker is more problematic. It might be better viewed as a suffix for the derivation of adverbs, e.g. *Özbekčä gäpirmäymän* 'I do not speak Uzbek'. There are extensions such as *-čäsiga*, consisting of *-čä* + third-person possessive suffix + dative, e.g. *qahramänčäsiga* 'heroically'. On the other hand, *-čä* is part of the complex postposition *-Gäčä* 'as far as, until', e.g. *üygäčä* 'up to the house', *yaqingäčä* 'until recently'. This type of shading of derivational elements and postpositions into the category of case is worth further investigation.

Adjectives

Though adjectives in principle are formally indistinguishable from nouns, some derivational suffixes, such as *-li* < *-lig*, e.g. *zärarli* 'harmful', the privative suffix *-siz*, e.g. *kücsiz* 'powerless', and the diminutive suffix *-Ginä*, e.g. *yumšäqqinä* 'rather soft', are specific to the category of adjectives. Intensification is often achieved by reduplication, e.g. *issiq-issiq* 'very hot'. The suffix *-räq* is mostly used in comparatives, e.g. *Näq älmädän širinräq* 'The pear is sweeter than the apple'. However, the ablative case is the central element in this construction, as evidenced by examples like *bäldän širin* 'sweeter than honey', in which *-räq* is not used. A superlative is expressed with the particle *ëñ*, e.g. *ëñ mäydä* 'smallest'.

Some adjectives such as *bäšqa* 'other, different', *özgä* 'other', *yaqin* 'near, close' and *äšiq* 'more' govern the dative or the ablative, e.g. *bunlärädän bäšqa* 'other than/different from these', *šähargä yaqin* 'close to the city', *bëš yildän äšiq waqt* 'a period longer than five years'.

Pronouns

The personal pronouns are *mën* 'I', *sën* 'you', *biz* 'we', *siz* 'you' (collective plural and polite form for one addressee), *sënlär* 'you' (informal plural), *bizlär* 'we' and *sizlär* 'you' (both individualising plurals). The case marking

generally conforms to the system employed with nouns. Some differences can be seen in Table 22.2.

The demonstrative pronouns form a rather complex four-place system consisting of *bu* 'this', *šu*, *oşa* and *u* 'that'. The oblique forms are *bun-*, *şun-*, *oşan-* and *un-*, to which possessive suffixes may also be added, e.g. *şunimiz* 'that one of us'. The *n* is not doubled in the accusative and genitive. The plurals are *bulâr*, *şulâr* and *ulâr*. The pronouns *šu* and *oşa* are used for objects and persons in view, the latter being more remote, whereas *u* signals reference outside the conversational setting. Both *bu* and *šu* are used for anaphoric textual deixis, the latter being more 'expressive', e.g. *Yânâ bittâ aytiñ şu gâpni!* 'Just repeat what you said there!' For cataphoric reference there is a preference for the use of *šu*, but *bu* occurs as well, e.g. *Şuni yaxşı biliñ: öçmâs nâmiñiz* 'Know this well: your name will not be extinguished', *Buni âniñ bilgil: êrklik tâlâb qilar* 'Know this for sure: freedom is exacting'. *Oşa* is also often used anaphorically.

The overall demonstrative system is more elaborate than this. The primary pronouns combine with the deictic particles *mânâ* 'here' and *ânâ* 'there', e.g. *mânâ bu kitâb* 'this book just here', *ânâbu âdâm* 'this man over here', *Sizni sorâgân ânâ şu kişi êdi* 'The one who asked for you was that very person'. Reflexive pronominal forms are derived from *öz* 'self'. *Öz* is also used attributively, e.g. *öz kitâbim* 'my own book', or substantivised with possessive suffixes, e.g. *özim* 'I (myself)'. These pronouns can also be used anaphorically between sentences. The pragmatic characteristics associated with *öz* are rather complex, e.g. *öz kitâbim* 'my own book', *özimniñ kitâbim* 'my own book', *âtâmniñ özi* 'my father himself'. The focusing properties of *öz* are seen in examples such as *U kabinetdâ özi otiribdi* '(S)he was sitting in the room', i.e. (s)he alone, *Kârimâ özi kirib keldi* 'Karima came herself', i.e. of her own free will.

Reciprocal expressions are formed with *bir bir* + possessive suffix + case.

Table 22.2 Personal and demonstrative pronouns

	Personal pronouns				Demonstrative pronouns		
	1p.sg.	2p.sg.	1p.pl.	2p.pl.	'this'	'that'	'that'
nom.	mên	sên	biz	siz	bu	šu	u
gen.	mêniñ	sêniñ	bizniñ	sizniñ	buniñ	şuniñ	uniñ
acc.	mêni	sêni	bizni	sizni	buni	şuni	uni
dat.	mêñâ	sêñâ	bizgâ	sizgâ	bunâ	şunâ	unâ
loc.	mêndâ	sêndâ	bizdâ	sizdâ	bundâ	şundâ	undâ
abl.	mêndân	sêndân	bizdân	sizdân	bundân	şundân	undân
equ. I	mêndây	sêndây	bizdây	sizdây	bundây	şundây ~ şundây	undây
equ. II	mênimçâ	sêniñçâ	bizniñçâ	sizniñçâ	bunçâ	şundây	unçâ
pl.	–	sênlâr	bizlâr	sizlâr	bulâr	şulâr	ulâr

Collective pronouns are derived from *hâmmâ*, *barçâ*, *bari* 'all', e.g. *Barçâ ornidân turdi* 'They all got up from their seats', *Hâmmâmiz keldik* 'We all came', *Bari yiyildi* 'They all assembled'. Generalising expressions are mostly formed with *hâr* 'every', often together with interrogative pronouns, e.g. *hâr kim* 'everybody', *Hâr bir qişlâq şiringinâ şâhârçâ bolişi kêrak* 'Every single village must become a nice little town'. The word *heç* is found in negative statements, e.g. *Heç kim gâp bâşlâmâdi* 'No one started to speak'. Indefinite pronouns may be formed with *allâ-*, e.g. *allâkim* 'some person', *allânimâ* 'whatever', or with *-dir*, e.g. *kimdir* 'somebody'. Interrogative pronouns are discussed on p. 373.

Numerals

The cardinal numbers one to nine are *bir*, *ikki*, *üç*, *tört*, *beş*, *âlti*, *yetti*, *sâkkiz*, *toqqiz*, *on* 'ten', *yigirmâ*, *ottiz*, *qirq*, *êllik*, *âltmiş*, *yêtmiş*, *sâksân*, *toqsân*, *yüz* 'hundred', *miñ* 'thousand', *million* 'million'. Optionally *-tâ* may be added, in particular to lower numerals, e.g. *bittâ*, *ikkitâ*, *üçtâ* ~ *üştâ*. Ordinal numerals are formed with *-(i)ñci*, e.g. *birinçi*, *ikkinçi*, *nêçânçi* 'the how maniest' ← *nêçâ* 'how many'. Two suffixes form collective numerals, *-âw* for one to seven and *-âlâ* for two to seven, e.g. *birâw* 'an individual', *ikkâw* 'the two', *ikkâlâsi* 'both of them', *törtâwimiz* 'the four of us'. Apart from *yarim* 'one/a half', *çâràk* 'one/a quarter' and *nimçâràk* 'one eighth', fractions are expressed by means of the ablative, e.g. *ikkidân üç* 'two thirds'.

There are several frequently used quantifier nouns such as *bây* 'bundle, sheaf', *bâş* 'head' for animals, e.g. *bir bây bêdâ* 'a sheaf of clover', *üç bâş qoy* 'three heads of sheep', *bir bâş üzüm* 'a bunch of grapes'.

Copulas

Nominal sentences have copula suffixes, particles or verbs indicating person and number, e.g. *Işçimân* and *Işçi bolâmân* can both mean 'I am a worker'. The latter form is semantically bleached to an extent that the dynamic meaning of *bol-* 'become' is lost. *Inžener bolâmân* may mean both 'I am an engineer' and 'I am going to be an engineer'. Past tense is expressed with *êdi* 'was'. Other copula particles are the conditional *êsâ* and the indirective *êkân* and *êmiş*, which form paradigms in combination with copula suffixes. Negation in nominal sentences is expressed with the negation particle *emâs*. In speech, *êdi*, *êkân* and *êsâ* are frequently suffixed as *-di*, *-kân* and *-sâ*, which is sometimes also rendered in written Uzbek.

Verbs

Deverbal Verb Derivation: Voice, Negation, Actionality

Verbs with cooperative-reciprocal meaning are formed with *-(i)ş*, e.g. *öp-* 'kiss', *öpiş-* 'kiss one another', *ur-* 'beat', *uriş-* 'fight', *sâlâmlâ-* 'greet', *sâlâmlâş-* 'greet one another', *toxtâ-* 'stop', *toxtâş-* 'stop all at once', *çaqir-* 'call', *çaqiriş-* 'call one another, call all at once'. A middle voice in *-(i)n* is

Table 22.3 Copula forms

Person	Personal suffixes pro-nominal type	Personal suffixes pos-sessive type	Past	Con- ditional	Indirective	Negation /present	Negation /past
1p.sg.	-mân	-m	édim	ésâm	ékân mân	émâsmân	emâs édim
2p.sg.	-sân	-ŋ	édiŋ	ésaŋ	ékânsân	émâssân	emâs édiŋ
1p.pl.	-miz	-k/q	édik	ésâk	ékân miz	émâsmiz	emâs édik
2p.pl.	-siz	-ŋiz	édiŋiz	ésaŋiz	ékânsiz	émâssiz	emâs édiŋiz
3p.sg.	-∅	-∅	édi	ésâ	ékân	émâs	emâs édi
[pl.]	[-lâr]						

derived from a limited number of verbs, e.g. *qaši-* 'scratch', *qašin-* 'scratch oneself' (also *özini qaši-*). Passives are derived with *-(i)l* and, from stems containing *ll*, with *-(i)n*, e.g. *oqi-* 'read', *oqil-* 'be read', *sür-* 'push', *süril-* 'be pushed', but *aldâ-* 'deceive', *aldân-* 'be deceived'. The passives can be impersonal and are fully productive also for intransitive verbs, e.g. *Bu yerdâ otirilmâsin* 'On this place one must not sit', *Poçtâgâ qandây bârilâdi* 'How can one go to the post office'. An agent may be introduced by a postpositional phrase with *tâmânidân*, e.g. *oyli tâmânidân yâzilgân xat* 'a letter written by his/her son'.

There are several causative suffixes, *-Dir*, *-(i)t*, *-Giz*, *-Gâz*, *-(i)z*, e.g. *tâptir-* 'let find', *süpürtir-* 'make sweep', *qaynât-* 'boil', *yâtkiz-* 'cause to lie down', *körgâz-* 'show', *âqiz-* 'make flow'. By causativisation intransitive verbs are made transitive and the original subject is made the direct object, e.g. *Kâmâl öldi* 'Kamal died' > *Kâmâl ni öltirdilâr* 'They killed Kamal'. With transitive verbs, the original subject is put in the dative case, e.g. *Bâlâ şe'rni oqidi* 'The child read a/the poem' → *Bâlâgâ şe'rni oqidi* '(S)he made the child read a poem'.

Main verbs with the converb suffixes *-â* and *-(i)b* can combine with auxiliary verbs to express semantic modifications. Thus, innate ability is expressed with *-â bil-*, e.g. *Yâza bilmây mân* 'I do not know how to write'. Circumstantial ability, on the other hand, is expressed with *-âl-* < *-â âl-*, e.g. *Xatni bugün yâzâ âlmâdim ~ yâzâlmâdim* 'I could not write the letter today'.

Verbs are normally negated with *-mâ*. The infinitive in *-mâq*, the verbal nouns in *-(i)š* and *-(U)w*, the present in *-mâqtâ* and the intentional in *-mâqçi* are not formed from negative *-mâ* stems. For other methods of negation, see p. 373.

Many constructions with auxiliary verbs express modifications of the intrinsic actional value (*aktionsart*) of the main verbs (see p. 42). Among the constructions with *-â* converbs, *-â bâr-* has durative meaning, e.g. *Bugün biznikidâ qâlâ bâriŋ* 'Stay with us today', *-â bâšlâ-* ingressive meaning, e.g. *yâzâ bâšlâ-* 'start writing', *yâzmây bâšlâ-* 'stop writing', *-â sâl-* coincidental meaning, e.g. *U kelâ sâlib mâjlisni âçdi* 'As soon as (s)he arrived (s)he opened the meeting', *-â bêr-* continuative meaning, e.g. *Çâpa bêrâdi* '(S)he keeps running', *-â yâz-* propinquitive meaning, e.g. *Ölâ yâzdi* '(S)he almost died', and *-(i)b kör-* desiderative or conative meaning: 'try, wish to', e.g. *Mên miltiqân atib kördim* 'I wanted to shoot the rifle'.

Among the constructions with *-(i)b* converbs, *-(i)b âl-* expresses orientation towards the first actant, e.g. *Fânni bilib âliqlâr!* 'See that you learn science!', *Yep âldim* 'I ate it all'. It is sometimes ingressive, e.g. *Sözlâšip âldik* 'We started a conversation'. The combination *-(i)b bâr-* expresses a gradually developing action, e.g. *yaxšilârib bâr-* 'get better and better'. Transformativity or completion is expressed with *-(i)b bol-*, *-(i)b çiq-*, *-(i)b ôt-*, *-(i)b qoy-*, *-(i)b sâl-*, *-(i)b tâšlâ-*, e.g. *Mên yuwinib boldim* 'I have already washed myself', *Oqib çiqmaqçi* '(S)he intends to read it all', *Qisqaçâ sözlâb ötmâqçimân* 'I will explain it shortly', *yâzib qoy-* 'write down', *Bâla piyâlâni sindirib qoydi* 'The child broke the glass', *urub tâšlâ-* 'give a good hiding', *Xânniŋ zulmi mehnâtkâš xalqni yêb sâldi* 'The cruelty of the ruler totally destroyed the people'. Besides, *-(i)b kêt-* expresses continuation up to a certain point of time, e.g. *Ikki yil išlâp keldim* 'I have been working for two years', *-(i)b kêt-* movement away from something and suddenness, e.g. *uçub kêt-* 'fly away', *isib kêt-* 'suddenly become hot', *Qayiq aydarilib kèdi* 'The boat suddenly turned over'. Constructions such as *-(i)b otur-*, *-(i)b tur-*, *-(i)b yür-* express non-transformativity with durative, continuative and habitual nuances, e.g. *Šu gâpni aytib otirišibdi-ya!* 'That is what they were saying all the time', *Gâpirib tur!* 'Keep talking', *Oqib yüribdi* '(S)he read on'. Finally, *-(i)b qâl-* signals a change of state, e.g. *sinib qâl-* 'get broken', and *-(i)b yubâr-* suddenness and completion, e.g. *Hâmmâ bâlâlâr birdân qiç-qirib yubârdilâr* 'All children shouted at once'. Occasionally, a finite form is repeated instead of the construction with the converb, e.g. *Oxladı-qâldi* '(S)he fell asleep'.

Thematic Suffixes

A variety of thematic suffixes may be added to verb stems. According to their morphosyntactic functions, finite, nominal and adverbial forms can be distinguished. Some forms may combine two or all of these functions. In the case of *-â/-y*, the finite and adverbial forms are etymologically not quite identical. The marker *-(U)w* is realised as *ū* and *ū* after consonants, and as *-w* after vowels. A stem-final *-â* is assimilated to *â*, e.g. *išlâw* 'working, work'.

Finite Verb Forms

The finite forms normally combine with copulative elements that express person and number. The copula may be a suffixed marker of the pronominal

Table 22.4 Thematic suffixes

Form	Finite	Personal suffix	Nominal	Adverbial (converbs)
-di	simple past	possessive type	–	[-di mi] (with Q-particle)
-á (after vowel: present/future -y)		pronominal type	–	in compounds, doubled: manner
-mâqtâ	progressive	pronominal type	–	–
-r	presumptive	pronominal type	participle	... -(a)r ... -mas
(negated -mâs)	future			
-mâslik	–		verbal noun	
-yâtir	non-habitual present	pronominal type		
-yap	cursive (focal) present	pronominal type		
-wât	present			
-GAn	post-terminal past	pronominal type	participle	verbal noun
-gAnlik	–		verbal noun	
-mâqçi	intentional future	pronominal type		
-(U)wçi	–		active participle	
-(â)yâtkân	–		past habitual participle	
-âdigân	future	pronominal type	future participle	
-(i)b	narrative past	pronominal type	–	conjunctive
(negated -mây)				
-Gâç				post-terminal purposive 'until ...'
-Gâli/-Gâni				conditional
-GUnçâ				
-sâ	suppositive modality	possessive type	–	
-mâq			infinitive	–
-mâqlik				–
-(i)š			verbal noun	–
-(i)šlik			verbal noun	–
-(u)w			verbal noun	–
-(y)âjâk	definite future [literary style]	pronominal type	participle	–

or possessive type, e.g. *išlâgânman* 'I have worked'. It can also be morphologically complex, yielding different aspecto-temporal values, e.g. *išlâgân êdim* 'I had worked'.

Finite verb forms are mostly negated with *-mâ*, e.g. *Kêlâmân* 'I will come', *Kêlmâyman* 'I do not come'. The form in *-(A)r* is negated with *-mâs*, e.g. *kêlârman*, *kêlmâsman*. The present tense in *-mâqtâ* is negated with *êmâs*, e.g. *išlâmâqtâmân* 'I am at work', *išlâmâqtâ êmâsmân* 'I am not at work'. In some cases negation is possible with *-mâ*, *êmâs* and *yoq*,

e.g. *Kêçâ qâr yâymâgân* 'It has not snowed tonight' with *Kêçâ qâr yâqqan êmâs* 'No snow has fallen tonight' and *Qâr yâqqan(i) yoq* 'No snow has fallen at all'.

Tense, Aspect, Mood

Uzbek is a language in which 'viewpoint' aspect, referred to in the following as *aspect*, is grammaticalised. Temporal distinctions are also well represented in the basic units. An important distinction is that between post-terminal ('perfect') and non-post-terminal forms, e.g. the question whether or not past events are presented from the point of view of their results. The proximity of the event to the time of speech also seems to be a feature encoded in past time forms. There are a number of present and past tense forms, both simple and compound. The present tense forms are the following:

A general intraterminal present in *-â* (after consonants)/*-y* (after vowels) has a wide range of uses including habitual and also future readings. The personal markers are of the pronominal type, the third person being marked with *-di*, e.g. *išlâyman* 'I work', *išlâysân* 'You work', *išlâydi* '(S)he works', *išlâyimiz* 'We work', *išlâysiz* 'You work' (plural/polite), *išlâydilâr* 'They work', *išlâysizlâr* 'You work' (infrequent).

A focal intraterminal (progressive) present in *-(â)yâtir*, with personal markers of the pronominal type, denotes that the event unfolds during a certain time span around the speech event, e.g. *Yâzâyâtirmân* 'I am writing', *Oqiyâtirlâr* 'They are reading'.

Another focal present in *-mâqtâ*, with pronominal personal markers, is typical of a factual literary or newspaper style, e.g. *Bârmâqtâimiz* 'We are going'. The negation is formed according to the nominal paradigm, e.g. *Bârmâqtâ êmâsmiz* 'We are not going'.

A high-focal, momentary present is the one in *-(â)yap < -â yâtib*, with pronominal personal markers and with *-ti* in the third person, e.g. *Yâzâyapti* ~ *Yâzyapti* '(S)he is writing (just now)', *Kûlmâyapmiz* 'We are not laughing'. This form indicates that the time of the event more or less coincides with the time of speech. It is not frequently used in the written language. A variant that may occur is *-wât*, a regional form from the Tashkent area.

The forms in *-(â)yap* and *-(â)yâtir* have developed from composite forms with *yât-* 'lie' as a descriptive verb. Other present forms with various nuances can be formed with similar verbs, e.g. *Yâzib turibmân*, *Yâzib oturubmân* 'I am writing', *Yâzib yürubmiz*, *Yâzib yâtibmiz* 'We are writing'.

The past tenses include the following forms:

The simple past in *-di*, with personal suffixes of the possessive type, is a general past tense with an inherent 'perfective' aspectual value, i.e. not viewing a developing event intraterminally, e.g. *Kêldim* 'I came', *Kêldiñ* 'You came', *Kêldi* '(S)he came', *Kêldik* 'We came', *Kêldiñiz* 'You came'.

A perfect in *-Gân* with personal suffixes of the pronominal type, presenting the event in a post-terminal perspective, e.g. *Kêlgânman* 'I have come'. The

form is often employed as the base form in narratives. Together with the past copula, it expresses a pluperfect or a distant past.

An indirective past is built with *-(i)b*, e.g. *Kitâbni bugün âlip kelişni unutibmân* 'I seem to have forgotten to bring the book today'. The personal suffixes are of the pronominal type; the third person is mostly marked with *-di*, e.g. *Kelibmân, Kelipsân, Kelipti ~ Kelibdi, Kelibmiz*. Third-person forms of verbs in *-l* are often shortened, e.g. *bopti < bolipti, Âpti < âlipti*.

Past tenses expressing intraterminal (imperfective) aspectual meaning are formed with the past copula particle *êdi*. Thus, *-(A)r êdim* may have habitual meaning, e.g. *U biznikigâ hâr hâftâ kélâr êdi* '(S)he came to our house every week'. The forms in *-mâqtâ* and *-yâtir* have past equivalents in *-mâqtâ êdi* and *-(â)yâtkân êdi*, the latter being more common than *-(â)yâtir êdi*. Examples of focal intraterminals: *Kârim kirgânidâ, mên xat yâzayâtkân êdim* 'When Karim entered I was writing a letter', *Ulâr yoldâ ketâr êkân, turli quşlârnig âwâzi êşitilib turmâqtâ êdi* 'While they were going their way, the voices of all sorts of birds were heard'.

In these types of complex tenses, the indirective copula particle *êkân* can take the place of *êdi*, e.g. *Bârmâgân êkânsân* 'Apparently, you have not gone'.

Morphological means are used to convey modal meanings. A primary class of modal forms is formed by the so-called optative-imperatives. The first-person suffixes *-(â)y* and *-(â)yin*, plural *-(â)ylik* form voluntatives, e.g. *Bu kitâbni âlayin* 'I will buy this book', *Futbol oyinigâ bârâylîk mi?* 'Shall we go to the football match?' A deontic-permissive modality is expressed by the third-person suffix *-sin*, e.g. *Bilsinlâr* 'They should know, let them know'. A strong imperative is formed with the conditional plus *-çi*, e.g. *Yâzsâgiz-çi* 'Come on, write!' The second-person singular imperative consists of the bare stem, to which *-Gin* may be added, e.g. *Oqigin* 'Read!', *Çiq, Çiqqin* 'Get out!' The polite form ends in *-(i)ñ*, e.g. *Oqiñ, Çiqiñ*. There are several plural forms: *-igiz, -igizlâr* and *-iñlâr*, e.g. *Oqigiz(lâr), Oqiñlâr, Çiqiniz(lâr), Çiqiñlâr*.

The so-called aorist in *-(A)r* with pronominal personal markers and its negation in *-mâs*, indicates modal meanings, e.g. *Bârârmân* 'I (suppose I) will go', 'I might go', *Bârârsiz* 'You will presumably go', *Bârâr* '(S)he should, might go'. It is used in generic statements, such as in proverbs, e.g. *Bilgân bilgânini işlâr, bilmâgân bârmâyini tişlâr* 'The knowledgeable man processes his knowledge, the nitwit bites his finger'. Apart from this, it is restricted to modal future time reference, e.g. *Mên bugün siznikigâ kélârmân* 'I (probably, possibly) will come to your house today', *Bu kişi bâzârdân quruq qaytmâs* 'This person would not return empty-handed from the market'.

The future in *-(y)âjâk*, with personal suffixes of the pronominal type, is restricted to the literary language and has its background in Oghuz Uzbek dialects. It expresses a stronger modality than *-(A)r*, e.g. *Bâtirlâr xalqniñ xâtirâsidâ hâmişâ yâşâyâjâklâr* 'The heroes will stay in the people's memory forever'.

The finite use of the participle in *-âdigân/-ydigân* seems to gain ground as a means to express a strong shade of future epistemic meaning, e.g. *Âftidân, u êrtâgâ dalâgâ ketâdigân* 'As it seems, (s)he will go to the field tomorrow'.

Intentionality can be expressed with *-mâqçi*, e.g. *Sengâ bir yaxşilik qilmâqçimân* 'I am going to do you a favour', *Mên bugün üydâ dâm âlib yâtmâqçi êdim* 'I was going to take a rest at home today and lie down', *Âfândi êşâgini sâtmâqçi boldi* 'The Efendi decided to sell his donkey'. Occasionally only *-mâq* is used, e.g. *Ikkâwi âwgâ çiqmâq boldi* 'The two of them decided to go hunting'. Intention is also expressed by *-âdigân/-ydigân*, e.g. *Kêçgâ yaqin uçraşadigân bolib xayrlâştik* 'We said goodbye deciding to meet again in the late afternoon', *Ulâr şu toyridâ gâpirmâydigân boldilâr* 'They decided not to speak about this'.

Non-finite Verb Forms

Nominal forms are integrated into the sentence through application of nominal morphology, case, possessive suffixes, forming the basis for complement, relative and adverbial clauses. There are simple converb suffixes such as *-â/-y* *-(i)b* (negated *-mây*), *-Gâç*, *-Gâni ~ -Gâli*, *-GUnçâ ~ -GUçâ*. The conditional is a special sort of converb, *-sâ* being marked for person with suffixes of the possessive type. The more complex converb suffixes include *-(A)r + êkân > -rkân* etc. Apart from these, there is a wide range of morphologically complex converb markers, mostly involving a verbal noun, a possessive personal marker and a case suffix and/or a postposition.

Adverbs

There are few morphologically marked adverbs. Some suffixes such as the equative *-çâ* typically have the function of deriving adverbial phrases, e.g. *mârdçâ* 'manly', *sâatlârçâ* 'for hours'. The suffix *-çâsigâ*, consisting of *-çâ* + possessive suffix + dative, yields adverbs of manner, e.g. *qahramânçâsigâ* 'heroically'. Another adverb suffix is *-lâb*, e.g. *Oquwçilârnî hârflâb êmâs, boyinlâb oqişkâ örgâtiş kërâk* 'One must teach the students not to read in letters, but in syllables', *Küzdâ xâsil tonnâlâb âlingân* 'In autumn the harvest has been reaped in tons'.

Some temporal adverbs are formed with the third-person possessive suffix, e.g. *kêçâsi* 'at night', *bir kûni* 'one day'. Spatial adverbs include *taşqari* 'outside', *içkâri* 'inside', *yuqâri* 'up', *quyi* 'down', *qarşi* 'opposite', *toyri* 'directly', *ilgâri* 'forward', *qaytâ* 'back'. There are also adverbs expressing degree such as *jüdâ* or *ğâyât* 'very'.

Other Morphological Categories

Postpositions

Certain postpositions govern the nominative, e. g. *bilân* (pronounced *blân*) 'with', *üçün* 'for', *kâbi, siğâri* 'like' (for qualitative comparison), *qadâr* 'as

much as, as great as' (for quantitative comparison), *ârqali* 'through', *sayin* 'every', *sari* 'towards', *boylâb* 'the whole distance or period of', *boyiçâ* 'in accordance with', *üzrâ* 'on', *tüfâyli* 'because of', e.g. *ârawâ bilân* 'with a car', *Kârim bilân birgâ* 'together with Karim', *Kârim üçün* 'for Karim', *hâmmâ ânâlar kâbi* 'like all mothers', *tâş parçâlâri siğâri* 'like fragments of a stone', *bir boyrâ qadâr çây* 'a brook as big as a reed mat', *nê qadâr* 'how much', *poçta ârqali* 'by mail', *yil sayin* 'every year', *çâr-âtrâf sari* 'in all four directions', *bütün dünüyâ boylâb* 'in the entire world', *bu yil boylâb* 'during this whole year', *qânun boyiçâ* 'in accordance with the law', *âq qâyâz üzrâ* 'on white paper', *mêñâ qilgân yaxşiligiñiz tüfâyli* 'on account of the goodness you have shown towards me'. Focused pronouns can stand in the genitive, e.g. *sêniñ üçün* ~ *sên üçün* 'for you'.

Certain postpositions govern the dative, such as *qadâr* 'until', *tâmân* 'towards', *körâ* 'according to, on account of', *qarab* 'towards', *qarşi* 'against', *qarâgândâ* 'compared with, considering', *qarâmây* 'notwithstanding', e.g. *értâlâbgâ qadâr* 'until the early morning', *üygâ tâmân* 'towards the house', *kâsâlligigâ körâ* 'because of his illness', *qişlâqqâ qarab* 'towards the village', *küçingâ qarab işlâ* 'work according to your strength', *uniñ kütgânigâ qarşi* 'contrary to his expectation', *Sizgâ qarâgândâ u yâs körinâdi* '(S)he looks younger than you do', *taşqi körinişigâ qarâgândâ* 'to judge from his outward looks', *Dekabr' bolışigâ qarâmây, kün issiq* 'Although it is already December, the days are warm'. *Tâmân* can also be used without the dative, or take on dative or ablative suffixes itself, e.g. *mâktâb tâmângâ* 'in the direction of the school', *mâktâb tâmândân* 'from the direction of the school'.

Other postpositions govern the ablative: *awwal*, *âldin*, *burun*, *ilgâri* 'before', *kéyin*, *soñ* 'after', *bêri*, *bâşlâb* 'since', *buyân* 'ever since', *tâşqari* 'besides, in addition', e.g. *bundân âldin* 'before this', *bayramdân ilgâri* 'before the holiday', *bundân üç kün awwal* 'three days ago', *hâmmâdân burun* 'before anything else', *bir haftâdân kéyin* 'after one week', *ikki yildân bêri* 'two years ago', *kêçâdân bâşlâb* 'since yesterday', *üç kündân buyân* 'for the last three days'.

For spatial and temporal reference, nominal kernels of the type *iç* 'inside' plus possessive suffixes are used (see pp. 48, 112–13, 221–2). The preceding nominal may be in the genitive or the nominative. The choice is sometimes optional. The most important nominal kernels are *iç* 'inside', *üst* 'upper part, surface', *têpâ* 'top', *âst* 'bottom', *tâg* 'lower part, bottom', *âld* 'front', *ârqa* 'back', *qâş*, *bâş* 'presence', *yân* 'side, presence', *ortâ* 'middle', *âra* 'space in between', e.g. *üy içigâ* 'into the house', *bir haftâ içidâ* 'in one week', *çalq içidâ* 'among the people', *köpriç üstidâ* 'on the bridge', *pul üstidâ* 'about money', *stol tepâsidâ* 'above the table', *suw âstidân* 'out of the water', *mêniñ âldimdâ* 'in front of me', *ârqamizdâ* 'behind us', *çirmân bâşidâ* 'at the threshing floor', *yânimgâ* 'to me', *qişniñ ortâsidâ* 'in the middle of winter', *ikki dërâzâ ârasidâ* 'between the two windows', *çalq ârasidâ* ~ *ortâsidâ*

'among the people', *tabib bilân qassâb ârasidâ* 'between the doctor and the butcher'. Spatial adverbs may also serve as kernels of postpositions, e.g. *şâhâr taşqarisidâ* 'outside of the city', *paçtâ toyrisidâ* 'about cotton', *qarşimdâ* 'opposite to me'. As is obvious from some of the examples given, spatial expressions could have extended, more or less metaphorical, meaning. The connection with spatial meaning is less direct in cases such as *âxiridâ* 'at the end of' ← *âxir* 'end', *haqidâ* 'concerning, about' ← *haq* 'right', *ornigâ* 'instead of' ← *orin* 'place', *tâmânidân* 'by' (in agentive phrases) ← *tâmân* 'side'.

Syntax

Nominal Phrases

The basic structure of the nominal phrase is determiner + quantifier (phrase) + adjective (phrase) + noun. All attributes, of course, are optional. The determiner is a pronoun. The adjective slot can also be filled with a relative clause. Under certain conditions, both quantifier and determiner may move to the right of the adjective. The conditions under which this happens are not easy to describe; compare *bir yaxşi ât* 'a/one good horse' (from the class of good horses) vs. *yaxsi bir ât* 'a good horse' (some indefinite horse that happens to be a good horse). Determiners and quantifiers other than *bir* ~ *bittâ* do not normally occur after an adjective, but do so when the adjective slot is filled with a relative clause, e.g. *kulâyâtkân şu bâlâ* ~ *şu kulâyâtkân bâlâ* 'that child that is laughing'.

The third-person possessive suffix can be used partitively, e.g. *dâstânlârdân biri* 'one (a certain one) of the epics', *Mêñâ qaysisi yaraşadi, qizilimi, âqimi?* 'Which one suits me, the red one, or the white one?'

There is agreement between the genitive attribute and the possessive suffix on the head in possessive constructions, e.g. *bâlâ-niñ kitâb-i* 'the book of the child'. There is no obligatory number agreement: *bâlâ-lâr-niñ kitâb-i* 'the book of the children'. Exceptions are rare and are confined to complements with first- and second-person pronouns, e.g. *bizniñ kitâb* 'our book'. The same type of agreement is required whenever the subject of a relative clause is marked with the genitive.

Plurality is expressed in a several ways. As usual, the nominal plural marker *-lar* often has an individualising quality. Nouns with quantifiers normally remain unmarked for plural, e.g. *üçtâ bâlâ* 'three children'. However, the plural suffix may be used in these cases to indicate the individuality of animate nouns, e.g. *Âkâmniñ üçtâ bâlâlâri bâr* 'My brother has three children', *köp âdamlâr* 'many persons' vs. *köp âdâm* 'many people'. Indefinite plurality is often indicated by reduplication, e.g. *qâp-qâp un* 'many sacks of flour', *Kêçâ kim-kim* ~ *kimlâr keldi* 'What persons have arrived tonight?'

There is no grammatical category of gender. Expression of female sex in some nouns borrowed from Arabic and Russian is a marginal phenomenon, e.g. *šâirâ* 'poetess' ← *šâir* 'poet', *studentka* 'female student' ← *student* 'student'. Natural gender, i.e. femininity, is indicated by using certain nouns as attributes, e.g. *âyâl*, *çâtin* 'woman', *qiz* 'girl', *uryâçi* 'female (animal)'. There is a neutral or a masculine reading for the unmarked case. The attributes *ér* 'man', *érkâk* 'male' are often used to express masculinity. Examples: *oyil bâlâ* 'male child', *qiz bâlâ* 'female child', *oraqçi (érkâk)* 'mower', *oraqçiâyâl ~ çâtin* 'female mower', *uryâçi arslân* 'lioness'.

Though animacy is not formally grammaticalised, the feature 'animacy' plays a certain role in syntax. Thus, the optional plural marking of verb forms in sentences with plural subjects is largely restricted to cases in which these subjects are animate, e.g. *Bâlâlâr küylâdi ~ küylâdilâr* 'The children sang', but *Daraytlâr kökârdil*kökârdilâr* 'The trees sprouted leaves'. A collective plural for verb forms with animate subjects is formed with the voice suffix *-(i)š*, e.g. *Kelâdi* '(S)he comes', *Kelâdilâr* 'They come' (individualising), *Kelišâdi* 'They come' (collective).

The Sentence

Finite Clauses

There is no subject-verb agreement as far as number is concerned. In the first and second person, the person is marked with a suffix on the verb. Personal pronouns may be used for focusing or topicalisation, e.g. *Kitâb yâzdim* 'I wrote a book' ~ *Kitâbni mên yâzdim* 'It was I who wrote the book', *Mên institutadâ edim* 'I (= topic) was at the institute' vs. *Institutadâ edim* 'I was at the institute'.

Existential sentences are constructed with *bâr* 'existent' or *yoq* 'non-existent' in predicate position, e.g. *Köp kiši bâr* 'There are many people', *Bugün nân yoq êkân* 'Today there appeared to be no bread'.

Possessive sentences are expressed with the possessed + possessive suffix + *bâr* and *yoq*, e.g. *Sâatiñ bâr mi?* 'Do you have a watch?' Alienable possession can also be rendered with the locative, e.g. *U kitâb mênâ yoq* 'I do not have that book'. A third way to express possession employs the noun *êgâ* 'possessor' governing the dative, e.g. *Kattâ tâjribâgâ êgâ boladi* '(S)he has much experience'.

Adjectives and nominal phrases can be used adverbially, e.g. *Kêç yâtdim* 'I have gone to bed late', *Bu yil qâr yâymâdi* 'This year no snow has fallen', *Yèrni qariš-qariš êgallâdi* 'They occupied the territory piece by piece'. Adverbial expressions are also formed with lexicalised converbs in *-(i)b*, e.g. *Tiküwçi bolib işläymân* 'I work as a tailor', *Uni prezident qilib saylâdilâr* 'They elected him president', *Mên sizni deb keldim* 'I came for your sake/ because of you' (with the accusative *siz-ni*).

Constituent Order

As for the constituent order in simple sentences, the first slot is the topic position. The closer a constituent is to the verb, the more it tends to be new information. The post-verbal position indicates topic continuance. An example of focusing: *Mên yiylâtkân êmâsmân, öziñ yiylâgânsân* 'It is not I who made you cry, you cried yourself'. An example of topic continuance: *Qizlâr zvenosi bu* 'This is a girls' squad', vs. *Bu, qizlâr zvenosi*. A focus site is created for the subject in *Sizni axtârgân tâbib mênîñ özim bolâmân* 'I am the doctor that cured you'. There is also a rather free ordering of converb clauses and their main clauses, e.g. *Lomonozov nâmlî universitetkâ Özbeklâr bârâdi âlim bolgâni* 'To the Lomonosov University the Uzbeks go in order to become scholars'.

Negation

Verbal negation with *-mâ* has been dealt with above. Nominal negation is formed with the particle *êmâs*, e.g. *Šu kitâb yaxši êmâs* 'That book is not good'. With verbal predicates, *êmâs* can be used for contrastive negation, e.g. *Paytani qol bilân êmâs, mašina bilân têramiz* 'We harvest the cotton not by hand, but with a machine'. Existential negation is expressed with *yoq*, the affirmative counterpart being *bâr*. A strong negation may be expressed by the type *Bârgisi yoq* '(S)he has no intention at all of going'. The Persian type of negation with *nâ...nâ* 'neither...nor' is also used, e.g. *Râsmni nâ körsâtdi, nâ bêrdi* '(S)he neither showed nor gave me the picture'.

Interrogatives

Yes/no questions are mostly expressed by *-mi*, which normally follows the predicate. It is a clitic, accompanied by a rising pitch in the preceding syllable. It precedes the copula suffix of nominal predicates, but follows the copula of verbal predicates, e.g. *Sâymisiz?* 'Are you well?', *Oqituwçimi êdiñiz?* 'Were you a teacher?', but *Kelâsizmi?* 'Will you come?', *Yâzmâgân êdiñmi?* 'Have you not written?' In the last example the question particle can also stand in front of the copula, i.e. *yâzmâgânmi êdiñ ~ (colloquial) yâzmâgânmidin*.

The most common question words are *kim* 'who', *nimâ* 'what', *qançâ* 'how much, how many', *nêçâ, nêç(â)tâ* 'how many', *qani* 'where', *qayêrdâ, qayâqtâ* 'where', *qayêrgâ, qayâqqa* 'where to', *qayêrdân, qayâqdân* 'from where', *qaysi* 'which', *qalây, qandây, qanaqa* 'how', *nâgâ, nimâgâ, nâçük, nimâ üçün* 'why', *qaçân* 'when'. The word group containing the question word typically stands in the focus position of the sentence, in front of the predicate, or forms the predicate itself, e.g. *Zaldâ qançâ âdâm bâr?* 'How many people are in the hall?'

Expression of Modal Nuances

Modality can be expressed, for instance, with the verbal nouns *-(i)š ~ -(U)w* or the conditional plus *mümkin* or *bolâdi* 'possible', e.g. *Bu magazindâ qâyâz*

tâpiş mümkin mi? 'Can paper be found in this store?', *Su kitâbni âlişim ~ âlsam mümkin mi?* 'Can I buy this book?', *Bu işni bir kündâ bâjarsam bolâdi* 'I can finish this job in one day', *Yérni haydab ~ haydasâ bolâdi* 'The soil can be ploughed'.

Presumption can be expressed by adding *-dir* to the predicate, e.g. *Qârniñ âčkândir* 'You must be hungry', *Aytkândirmân, êsimdâ yoq* 'I must have said so, I do not remember'. Other shades of modality are also expressed with *-dir*, e.g. *Râisi bolmasâ sekretari bârdir* 'Even if they have no chairman, at least they have a secretary', *Kêtişkâ mâjburdirsiz* 'You really have to go'. *-dir* also can be attached to question phrases, e.g. *Kimdândir êşitdi* '(S)he heard it from whoever it was', *qandâydir* 'in whatever way'. Another presumptive form is *kêrak* combined with the conditional, e.g. *Sên mêninğ gâplârimdân gümânsir-âğân bolsâñ kêrak* 'You must be inclined to question what I am saying'. Modal meanings of doubt or suspicion are conveyed by clitics such as *-çi* and *-(y)â*, e.g. *Sên âpaniñ oylisân-â?* 'So you are aunt's son?', *Kêlsâ-çi?* 'But if (s)he really comes?' Stronger doubt is expressed by *-mi êkân*, *-mikân*, *-mikin*, e.g. *U kitâbni keltirgânmikin?* 'Have you really brought the book?' Supposition can be expressed with the conditional suffix *-sâ*, e.g. *Qani, bugün biznikigâ kêlsâñ?* 'How about it, would you not come to us today?', *Êrtâgâ yolgâ çiqsâm?* 'Could I not leave tomorrow?'

Indirective meaning can be carried by the past tense on *-(i)b*. In most cases, however, it is expressed by the copula particle *êkân*, e.g. *Kâsâl êkânsiz* 'You are ill (as I have heard)', *Xatni yâzmâs êkânsân* 'You will not write the letter (so it seems)', *Kârim qislâqqa kêtâmâğân êkân* 'Karim had obviously not gone to the village'. The interrogative *-mi êkân > -mikân*, *mikin* expresses doubt, e.g. *Kêlgânmikân?* 'Has (s)he really come?'

The verb *kêl-* may combine with the otherwise obsolete verbal noun suffix *-Gi* or with *-Gân* to express desire to do something, e.g. *Çêkkilâri kelmâyapti* 'They do not feel like smoking right now', *Kitâbni oqigâni keldi* '(S)he felt like reading the (whole) book'.

A frequent construction with deontic meaning is the combination of verbal nouns with *kêrak ~ lâzim* 'necessary', e.g. *U xat yâzişi kêrak êdi* '(S)he had to write a letter' (with nominative subject), *Qoŋgirâqni çaliş kêrak!* 'One must ring the bell' (impersonal). A strong shade of modality is obtained by negating the permissive in *-sâ + boladi*, e.g. *Bârmâsâk bolmaydi* 'We absolutely must go'.

Coordination

There exist a number of coordinating conjunctions, parallel ones such as *wâ*, *hâm*, *-dâ* 'and', adversative ones such as *ammâ*, *lêkin*, *birâq* 'but, however, although', and separative ones such as *gâh ... gâh*, *ba'zân ... ba'zân*, *dâm ... dâm* 'sometimes ... at other times', *yâ ... yâ* 'either ... or', *nâ ... nâ* 'neither ... nor'.

Non-finite Clauses

Relative Clauses

To form relative clauses, normally one of the participles in *-Gân*, *-(â)yâtkân* and *-âdigân* is used. In the following sentence, *-Gân*, indicates anteriority, *-(â)yâtkân* simultaneity, and *-âdigân* posteriority relative to the speech event, e.g. *Sizniñ üstinğizdâ bolgân, bolayâtkân wa bundân keyin bolâdigân hâmmâ gâplârdân xabârdârmân* 'I am informed about all rumours that have been told about you, that are being told and that will be told from now on'.

If the nominal head is also the first actant (subject) of the relative clause, it is not expressed, e.g. *xatni yâzgân kişi* 'the man who wrote the letter'. In other cases there are two strategies. The subject of the relative clause may be in the genitive as in *kişiniñ yâzgân xati* 'the letter the man wrote' or *(mêninğ) bârayâtkân yol-im* 'the road I was travelling along'. Here the head carries a possessive suffix as an agreement marker. The subject can also be in the nominative, as in *kişi yâzgân xat* or *mên bârayâtkân yol*, in which case no agreement is marked.

An example of relativising the possessor within a possessive construction is *boyinlârigâ qoŋirâqlâr âsilgân tüyâlâr* 'camels, around whose necks bells are hung'. There are also verbless constructions that are syntactically akin to these clauses, e.g. *kâllâsi kâttâ kişi* 'a man whose head is big', where *kâttâ* 'big' does not modify the head noun.

Other participles are *-(A)r*, mainly occurring in fixed expressions, e.g. *mêninğ turur yâyim* 'the place where I stay', and *-(U)wçi*, a habitual participle, e.g. *âdâb bilüwçi âdâm* 'a man who knows what is proper'.

Complement Clauses

To form complement clauses, a number of verbal nouns may be used. The subject of the subordinate clause may be marked with the genitive, e.g. *Yigitâ(niñ) nimâ demâqçi bolgâni aŋlaşildi* 'It became clear what the boy wanted to say'. Other examples with *-Gân* and *-Gânlîk* are *Sên bilân gâplâşmâğânim yaxşi* 'It is good that I have not spoken with you', *Kêlgâniğâ xursândmân* 'I am glad you have come', *Nimâ qilgânlîgini yâzgân êdi* 'She had written what she had done'. Nominal clauses can be subordinated with *êkân* or *êkânlîk*, e.g. *Ziyâdâxânniñ sözigâ küliş kêrak êkâni keyin êsigâ keldi* '(S)he realised afterwards that what Ziyadkhan said should be laughed at'.

Other verbal nouns used are *-mâq* and *-mâqlik*, *-(i)ş* (negation *-mâslik*) and *-(u)w*, e.g. *Aybni yaşirmâq, nâdânlâr işi* 'To cover up a fault is the work of stupid people', *Êmdi sêni körmâqlik mâlâl* 'It is not convenient now to see you', *Nêgâ mên u bilân uçraşuwdân hayiqay?* 'Why should I be afraid of meeting him?'

The clauses may also occur as complements of nominal phrases, e.g. *Oyliniñ ölgânlîgi xabâri keldi* 'The news arrived that his son had died', *Jawâb bêriş waqti keldi* 'The time has come to give an answer'.

Complement clauses can also be marked with a preposed *ki*, e.g. *Mèn èšitdimki sèn kèčà mukâfât âlibsin* 'I heard that you have got a reward' vs. *Mèn sèniñ kèčà mukâfât âlgâniñni èsitdim*.

Converb Clauses

Converb clauses with *-â* indicate the manner in which the main event evolves, e.g. *gâpigâ pâr wâ qilmây* 'not paying attention to what (s)he said'. The converb is normally doubled, either on the same verb, e.g. *Hikâyâni külä külä aytâ bâšlâdim* 'I began to tell the story laughingly', or on different verbs, e.g. *Aytmây indâmây çiqib kètdi* 'Without saying anything (s)he went out'.

Clauses with *-(i)b* indicate succession or simultaneity of events, e.g. *Ulâr yânâ yarim saât išlâb pařtaxârdân qaytišdi* 'They worked for an extra half hour and (then) went home from the cotton field', *külib gâpirdi* '(s)he laughed and spoke/(s)he spoke laughing'. The negated counterpart is *-mâstân*, e.g. *U šâsilmâstân qulay pâytni kütdi* 'Not taken aback, (s)he waited for the right moment'.

Clauses with *-Gâč* indicate a post-terminal perspective on the event, e.g. *Tün kètgâč tân âqârgâč qušlâr sâyraý bâšlâdi* 'After the night had gone and dawn had broken, the birds began to sing'. Clauses with *-Gâni ~ -Gâli* express purposive meaning, e.g. *Sèni körgâli kèldim* 'I have come to see you'. Clauses with the limitative-terminative converb in *-GU(n)čâ* set a temporal frame for the main event in the sense of 'as long as' or 'until', e.g. *Körmâgünčâ išanmây mân* 'I will not believe it until I see it (as long as I do not see it)', *U kèlgünčâ kütiñ* 'Wait until (s)he comes'. The meaning can also be preferential ('rather than ...'), e.g. *Qâyâz gâ yâzgunčâ, dâftârgâ yâz* 'Write it in the notebook rather than on a piece of paper'. Possessive suffixes occasionally occur between *-GUN* and *-čâ*, e.g. *Siz kèlgünigizčâ mèn bu yèrdâ bolâmân* 'I will stay here until you come'. Nominal clauses with *ékân* and verbal clauses with *-(A)r + êkân > -rkân* express temporal overlapping ('while'), e.g. *u sözlârkân* 'while (s)he is speaking'.

The conditional suffix *-sâ* is used in conditional clauses, with or without conjunctions such as *âgâr* 'if' and *bârdi-yu* 'and if', e.g. *U qorslik qilsâ, mèn nimâ dèymân?* 'If (s)he acts rudely, what am I to say?', *Âgâr mèn kèçikib qâlsâm, mâjlisni sèn âçasân* 'If I am late, you must open the meeting'. It is used in the protasis of counterfactual conditional clauses, e.g. *Èh, yâš waqtim bolsâ, undâ bâšqa nârsâ dèr èdim* 'Well, if I were young, I would tell him something different'. The conditional may be followed by the past copula particle *èdi*, e.g. *Sèn kèlsân èdi, muzeygâ bârgân bolâr èdik* 'If you had come, we could have gone to the museum'. Clauses of contradiction are formed with *-sâ + hâm* 'and', e.g. *Quyâš bâtib, yèrgâ qârângilik tüšgân bolsâ hâm, biz dâlâdân çiqmâgân èdik* 'Although the sun had set and darkness had descended on the earth, we had not left the field'. The suffix is also used in combination with question words in generalising expressions

such as *Kim dèwârniñ nârigi tâmânidâ bolsa, u âmân qâldi* 'Whoever was on the other side of the wall was safe', *Nimâ èksân, šuni orâsân* 'Whatever you sow, that you will reap'. The conditional suffix is also used as a finite verb in (elliptical) proposals and exclamations, e.g. *Bugün bir muzeygâ bârsâk* 'What about us going to a museum today?', *Kâški bilsâm èdi!* 'If only I had known!' (counterfactual). It is also common in temporal usage, e.g. *Âynadân qarâsâm, u kèlâyâtkân êkân* 'When I looked out of the window, (s)he was coming'.

Adverbial clauses are also based on more complex forms. The subject of such clauses is not normally put in the genitive. Most junctors are postposed, whereas others are preposed, e.g. the ones based on *ki*.

Temporal clauses are formed in many ways. The junctors include *-Gân + possessive suffix + case*, e.g. *Mèn kèlgânimdâ, siz yoq êkânsiz* 'When I came, you were not there', *Kèlgünigâ kütdik* 'We waited until you came', *-(i)s + possessive suffix + -Dâ*, e.g. *Mâktâbtân qaytišimdâ uyâ aytâ bèrdim* 'When I returned from school I told him', *-(i)š + possessive + bilân*, e.g. *Yâtiši bilân uyqugâ kètdi* 'The moment (s)he laid down, (s)he went to sleep', *-Gân + zamân*, e.g. *uygâ kèlgân zamân* 'as soon as (s)he came home', *-dimi*, formally identical with an interrogative simple past, e.g. *Pařta âçildimi, tèrim bâšlânâdi* 'As soon as the cotton blossoms, the harvest begins'. Coincidence is expressed with *-Gân ~ -(i)š + waqtâ ~ çâyda ~ pâyttâ*, e.g. *Piyâdâ âskârlâr hüjum bâšlâgân waqtâ, kütilmâgân hâdisâ roy bèrdi* 'When the infantry proceeded to attack, something unexpected happened'. Some junctors consist of *-Gân (+ possessive suffix) + -Dân + postposition*, e.g. *sözlâšgândân keyin ~ soñ* 'after having spoken', *sèn kèlgânigândân bèri* 'since you have come', *sekretar bolgânidân buyân* 'ever since (s)he became secretary'. *-Gân + sayin ~ sari* expresses gradually evolving events, e.g. *Quyâš tikkâygan sayin hawâ küçlirâq qiziydi* 'As the sun climbed ever higher, the air got hotter and hotter'. *-GUN + -Gâ + qadâr*, e.g. *mâktâbni bitirib çiqqungâ qadâr* 'until finishing school'.

Causal clauses may be formed with *-Gân + possessive suffix + -Dân*, e.g. *Köčâ tutâš balçiq bolgânidân, yürišgâ imkân yoq èdi* 'Because the street was mud all over, it was impossible to walk there', or with *-Gân + possessive suffix + postposition*, e.g. *u kâsâl bolgâni sâbabli ~ sâbabdân ~ tüfâyli* 'because (s)he was ill'. The construction finite tense + *šekilli* indicates the reason for a conclusion expressed in the main clause, e.g. *Kârimjân kèldi šekilli, âwâzi kèlyapti* 'Karimjan must have come, as his voice is being heard'. Other possibilities are offered by *çünki, nègâki, sâbabki* 'because'.

Purpose clauses are formed with *-sin üçün, -sin dèb* and *-mâq üçün*, e.g. *èšitsin üçün* 'in order that (s)he listens', *körišmâq üçün* 'in order to meet'. Comparative clauses expressing 'like, just as' are formed with *-Gân + possessive suffix + -dây ~ -dèk ~ siğari ~ kâbi*. Clauses expressing preference are formed with *-Gân ~ -(i)š + -Dân + körâ*, e.g. *Mèndân külgânigândân körâ, râhmiñ kèlsin* 'Instead of laughing at me, you should have pity'. The

construction *-(i)s* + possessive suffix + *čâ* expresses 'as far as', e.g. *mëniġ ešitišimčâ* 'as far as I have heard'.

These are only examples of morphologically composite conjunctive devices. This system is essentially an open one, and many more combinations exist.

A loose connective device is offered by the clitic *-ki*. Other clitics are *-ku* and *-(y)u*, e.g. *Këčâsi qalin qâr yâġdi-yu*, *hawâ unčâlik sâwuġ êmâs edi* 'Much snow has fallen tonight, although it was not very cold outside'.

Lexicon

The long-standing intensive contact with Iranian languages has resulted in numerous loanwords. A special feature of nominal morphology is the rather widespread use of Persian affixes. Characteristically, their application is not confined to loanwords, e.g. *tilšunâs* 'linguist', *tilšunâslik* 'linguistics', *otinfuruš* 'seller of firewood'. Prefixed forms include *nâtoġri* 'untrue', *bëorin* 'out of place'.

Dialects

The best overviews of Uzbek dialects are Rešetov and Sâabdurahmanov (1978) and Sâabdurahmanov (1984). From a comparative point of view much would be gained from a discussion of the dialects, with the inclusion of the dialects spoken in northern Afghanistan, a field of research almost totally neglected in Uzbek dialectology.

References and Further Reading

- Borovkov, A. K. (ed.) (1959) *Uzbeksko-russkij slovar'*, Moscow: Gosudarstvennoe izdatel'stvo inostrannyx i nacional'nyx slovarej. (Contains a concise grammar.)
- Gabain, A. von (1945) *Özbekische Grammatik*, Leipzig: Harrassowitz.
- Kamâl, F. (ed.) (1957) *Hâzirgi zamân Özbek tili*, Tashkent: Özbekistan SSR Fanlar Akademiyasi Našriyâti.
- Kononov, A. N. (1960) *Grammatika sovremennogo uzbekskogo jazyka*, Moscow, Leningrad: Izdatel'stvo Akademii Nauk SSSR.
- Nigmatov, K. G. (1992) 'Fonologičeskaja sistema sovremennogo uzbekskogo jazyka i ee stanovlenie', *Tjurkologija* 1992, 3: 3–10.
- Raun, A. (1969) *Basic Course in Uzbek*, Indiana University Publications 59, Uralic and Altaic Series, Bloomington: Indiana University; The Hague: Mouton.
- Rešetov, V. and Sâabdurahmanov, Š. (1978) *Özbek dialektologiyasi*, Tashkent: Fan.
- Sâabdurahmanov, Š. (1984) *Özbek xalq ševalari morfologiyasi*, Tashkent: Fan.
- Sjoberg, A. F. (1963) *Uzbek Structural Grammar*, Bloomington: Indiana University; The Hague: Mouton.
- Waterson, N. (1980) *Uzbek-English Dictionary*, Oxford: Oxford University Press.
- Wurm, S. (1945) *Der özbekische Dialekt von Andidschan. Phonetische und morphologische Studien, Texte 1, Sitzungsberichte 224: 3*, Vienna: Akademie der Wissenschaften in Wien.
- (1959) 'Das Özbekische', in J. Deny *et al.* (eds) *Philologiae Turcicae Fundamenta 1*, Wiesbaden: Steiner, 489–524.

23 Uyghur

Reinhard F. Hahn

Introduction

The language varieties that are nowadays collectively referred to as Modern Uyghur, *hâzirqi zaman uyyur tili* 'present time Uyghur language', used to be known as Eastern Turki in the west. The collective ethnic label *uygur* was introduced in 1921. Its choice was based on the assumption that the present-day Uyghurs are the direct descendants of the Old Uyghurs. The present-day Uyghurs are evidently of multi-ethnic, though assumedly of predominantly Turkic, descent. The language varieties they use are typologically separate from Old Uyghur.

Modern Uyghur, hereafter referred to as Uyghur, is most closely related to those language varieties that are collectively known as Uzbek. There is a particularly high degree of mutual intelligibility between most varieties of Uzbek and Uyghur.

Uyghur is the native language of almost all Chinese citizens that are classified as ethnic Uyghurs. Furthermore, it has also become the first language of a number of Chinese citizens with different official ethnic labels, especially those whose ancestors speak or spoke other Turkic languages that are less widely used in Xinjiang, for example Tatar, Uzbek and Ili Turki. Uyghur is the official language of Xinjiang, along with Standard Mandarin which has official currency throughout China. Thus, Uyghur is also widely used as a second language, namely as an inter-ethnic regional lingua franca, by a considerable number of non-Uyghur residents of Xinjiang, for example by Uzbeks, Kazakhs, Kirghiz, Tatars, Oirat-Mongols, Dungans (Hui), Sarkolis, Wakhis and Russians, and to a lesser extent by Han Chinese.

Uyghur has been a literary language for several centuries. The preferred pre-twentieth-century literary language was Chaghatay, mostly forms of Chaghatay that had been influenced by regional dialects. Two closely related standardised written Uyghur languages were introduced into Xinjiang and the Soviet Union respectively no earlier than the mid-twentieth century. Most speakers of Uyghur are literate in their own language. Most of those in Xinjiang also speak Mandarin as a second language, and a large and growing number is literate in Mandarin as well. Most of those in the central Asian

republics do not know Mandarin but use Russian as well as the respective majority languages along with Uyghur, their native language. Proficiency in Russian is also widespread among Uyghur speakers in Xinjiang, especially among those who received their formal education before the late 1960s.

The Sound System

Phonetics and Phonology

Uyghur is characterised by the following features listed in traditional classifications of Turkic languages: development of **ð* to *y*, e.g. *adaq* > *ayaq* 'foot', retention of initial **b*- in *bol*- 'become', vs. Turkish *ol*-, retention of final **G*, e.g. Uyghur *tay* vs. Kazakh *taw*, Turkish *dağ* [da:] 'mountain', Uyghur *səriq* 'yellow' vs. Kazakh *sari*, Turkish *sarı*, and retention of the consonantal onset in suffixes such as the participial marker *-GAN*, e.g. Uyghur *qalyan* vs. Turkish *kalan* 'remaining'. Uyghur belongs to the group of Turkic languages in which there is essentially no voiced vs. unvoiced opposition among non-labial initial consonants, e.g. *taş* 'rock', *tö(r)t* 'four' vs. Turkish *taş* and *dört*.

Vowels

Disregarding vowel length distinction, the vowel inventory consists of the following phonemes, which may be subcategorised on the basis of height, backness and roundness. It has been argued, within a lexical phonology framework, that /è/ has a back counterpart /ɤ/.

The phonetic default realisations of the short vowels are as follows: /i/ → [ɪ], /ü/ → [y], /i/ → [ɪ], /u/ → [ʊ], /è/ → [e], /ö/ → [ø], /o/ → [o], /e/ → [ɛ], /a/ → [ʌ].

The high vowels undergo some tensing when they occur adjacent to alveolars, palatals and alveodental continuants, e.g. *čiraq* [tʃʰiraq] 'lamp', *yüz* [jyz] 'face', *suda* [su:da] 'in/at (the) water', *jenubiy* [dʒɛnu'bi:] 'southern'.

Both /i/ and /ü/ undergo apicalisation after alveodental continuants in unstressed syllables, e.g. *siler* [sɪ'læ:] 'you (plural)', *ziyan* [zɪ'ja:n] 'harm'. They are medialised after /ɣ/ or before /l/, e.g. *til* [tʰil] 'tongue', *xiżmet*

Table 23.1 Vowel phonemes

	Unrounded		Rounded	
	front	back	front	back
high	i	ĩ	ü	u
mid	è			
low	e	a		

[xiż'met] 'service'. After velars, uvulars and /f/ they are realised as [e], e.g. *giram* [ge'ram] 'gramme', *xelqi* [xel'qʰe] ~ [xæ:'qʰe] 'his [etc.] nation', *finn* [fen] 'Finn'. Between two syllables that contain a rounded back vowel each, they are realised as back, e.g. *qolimu* [qʰo'luumu] 'also his [etc.] arm'. Any vowel undergoes laxing and backing when it occurs in uvular and laryngeal environments, e.g. *qiz* [qʰyz] 'girl', *hisap* [hɪs'ʌp] 'number', *eyiz* [ʔe'æz] 'mouth', *halqa* [hal'qʰa] ~ [hɑ:'qʰa], 'ring', *qum* [qʰum] ~ [qʰom] 'sand', *qolay* [qʰo'lɑɪ] 'convenient', *qeyez* [qʰæβ'æz] 'paper'.

Lowering tends to apply to the non-high vowels when a syllable-final liquid assimilates to them, e.g. *kör* [kʰœ:] 'look!', *boldi* [bɔ:'dɪ] 'he [etc.] became', *ders* [dæ:s] 'lesson', *tar* [tʰɑ:] 'narrow'.

There are also arguments for a vowel length distinction, for underlyingly long vowels which resist devoicing, umlauting and raising, attract secondary stress in non-final syllables, and require that *y* or *r* intervene between them and a following possessive suffix. Long vowels may be pronounced slightly longer in certain environments, but detailed phonetic data are lacking. Vowel length is not distinguished in the official orthographies.

Consonants

The inventory of consonant phonemes is shown in Table 23.2. Commas separate voiceless and voiced consonants that otherwise share place and manner of articulation. Of these phonemes, only /ŋ/ and /N/ cannot occur word-initially. It has not been determined whether they are two distinct phonemes, but it is assumed to be the case in this description.

The glottal stop [ʔ] never closes a syllable. As for instance in German, it occurs word-initially wherever a word is perceived as beginning with a vowel. So far it has not been determined whether or not a word-initial glottal stop is phonemic or inserted by rule. For the sake of simplicity in this introductory description we shall assume that they are inserted. They seem to be phonemic in non-initial position, which is always in loanwords, e.g. *tebi'i* 'natural' < Arabic, *faŋ'en* 'scheme' < Mandarin, *jem'i* 'altogether' < Arabic.

The phonemes /f/, /fj/, /ʃ/, /ɣ/ and /h/ occur only in loanwords. Thus /fj/ and /ʃ/ do not participate in voicedness alternation processes with native /ç/ and /ʃ/ respectively. Other than in loanwords, the /h/ occurs only in

Table 23.2 Consonant phonemes

	Labial	Alveodental	Alveopalatal	Velar	Uvular	Laryngeal
stop	p, b	t, d		k, g	q, ɢ	ʔ
affricate			č, j			
fricative	f (v)	s, z	š, ž		χ	h
nasal	m	n		ŋ	N	
liquid		l, r				
glide	w		y			

a few native interjections. It is assumed here that /χ/ is loanword-specific. The grapheme <x> in native words represents /q/ in its fricativised realisation. Many speakers realise /h/ as voiced [ɦ], while others realise it as voiceless [h].

The voiced uvular stop /g/ alternates with the voiceless uvular stop /q/ in the same ways as /g/ alternates with /k/ in the velar group. It is realised as a fricative everywhere except after non-labial nasals, e.g. *γazγa* [ʁɑ:z'βɑ] 'to a/the goose', *tay* [tʰɑ:ʁ] 'mountain', *bayče* [bɑʁtʃɛ] 'park', *nanyɑ* [nɑn'gɑ] 'to the bread'.

Whether what is represented here as /w/ is a glide or a fricative consonant depends on the dialect spoken. In some dialects it is realised as a labiovelar or labiodental glide [w], [v], while in other dialects it is realised as a bilabial or labiodental fricative [β], [v]. Some speakers distinguish two phonemes, /w/ and /v/, in which case /v/ occurs only in loanwords, e.g. *vitamin* 'vitamin'. In writing, this distinction is only observed in the Latin alphabet-based system.

Uyghur aspiration is very similar to aspiration in most English and German dialects. In ordinary speech a voiceless stop or affricate preceding a vowel is aspirated unless it is directly preceded by a voiceless consonant, and it is weakly or not at all aspirated elsewhere, e.g. *pütün* [pʰʰʰtʰʰʰɯn] 'entire', *qalpaq* [qʰʰʰɑl'pʰʰʰɑqʰʰʰ] 'hat', *kök* [kʰʰʰøʰʰʰ] 'sky', *terep̄te* [tʰʰʰɛrɛpʰʰʰtɛ] 'on a/the side'.

Prosody

As a general rule, primary accent falls on the last syllable of a word. If it falls on a non-final syllable of what is written and perceived as a word, the following unaccented syllable(s) ought to be assumed to be either enclitics or separate words. Long vowels in non-final position, especially in penultimate position, tend to attract secondary accent, e.g. *tetq̄q̄āt + dA/* → *tetqiqatta* [tʰʰʰɛtq̄w(:)qʰʰʰɑ(:)tʰʰʰɑ] 'in (the) research'.

Morphophonology

Vowel Devoicing

Short high vowels are devoiced when they occur between two voiceless consonants (including glottal stops), in a word with more than one syllable usually only in unstressed position, e.g. *tik* [tʰʰʰk] 'steep', *isim* [ʔʰʰʰsɪm] 'name', *küt* [kʰʰʰʏt] 'wait!', *qušlar* [qʰʰʰʊʃ'lɑ:] 'birds'. Between a glottal stop and a voiceless stop or affricate in the standard language (also elsewhere in some other dialects) this tends to lead to total devocalisation: /i/ and /i/ become [ʃ], /ü/ becomes [ϕʰ], a voiceless [ɣ]-coloured bilabial fricative, and /u/ becomes [ϕ], a voiceless bilabial fricative, e.g. *it* [ʔʰʰʰt] 'dog', *ikki* [ʔʰʰʰke] 'two', *üčün* [ʔʰʰʰʏtʃɯn] 'for', *uka* [ʔʰʰʰkɑ] 'younger brother'.

Umlauting

A typical feature of Uyghur not including the Lopnur varieties is an umlauting phenomenon: a low-level short vowel in an open syllable turns into a mid-level vowel if the following syllable contains a high unrounded vowel, e.g. /et+i+/ → *éti* 'his [etc.] flesh', /at+i+/ → *éti* 'his [etc.] horse'. Long vowels are exempt from this rule, e.g. /pēr+i+/ → *peri*, not **pèri*, 'his [etc.] feather', /gāz+i+/ → *γazi*, not **γézi*, 'his [etc.] goose'.

Front vs. Back Harmony

Suffixes can contain high or low vowels, not mid-level vowels, as shown in Table 23.3. It is only these types of vowels that participate in vowel harmony. Suffix vowels are underlyingly specified as high or low and as unrounded or rounded. Whether they are realised as front or as back vowels depends on the stem to which they come to be attached. A suffix vowel must be front if the stem is front, and it must be back if the stem is back. This front vs. back harmony affects not only vowels but also consonants in suffixes. Velar and uvular stops are realised as *g* and *k* in front suffixes, and as *γ* and *q* in back suffixes. Harmonic assimilation will be illustrated with the nominal suffix *-IAr*, the desiderative suffix *-GU* and the verbalisation suffix *-IA*. The segments represented by capital letters are pre-harmonic archiphonemes that become fully specified only when they have been integrated into a word. Thus *-IAr* has front realisations such as in *kišiler* 'people', *künler* 'days', *kešler* 'slippers', *sözler* 'words', and back realisations such as in *qušlar* 'birds', *dostlar* 'friend', *qaplar* sacks, *tillar* ← /tʰʰʰl+IAr/ 'tongues'. Front suffix variants are found in *kiygü* 'desire to wear', *köygü* 'desire to sing', *kelgü* 'desire to come', *türligü* ← /tʰʰʰr-IA-GU/ 'desire to sort', and back variants in *qoyyu* 'desire to put', *qalyu* 'desire to stay', *qiyyu* ← /qʰʰʰy-GU/ 'desire to cut' and *bašliyu* ← /baš-IA-GU/ 'desire to start'.

Loanwords may contain both front and back elements, but any suffix attached to such a loanword stem takes its harmonic information only from the last syllable of the stem. This may be illustrated by means of the suffixes *-IAr* and *-dA* attached to *adem* 'human being' < Arabic *Sādam*, *polek* 'Pole' < Russian *poljak*, *šenduŋ* 'Shandong' < Mandarin *Shāndōng* and *Xunen* 'Hunan' < Mandarin *Húnán*: *ademler* 'human beings', *polekler* 'Poles', *Šenduŋda* 'in Shandong', *Xunende* 'in Hunan'.

An apparent exception to the front vs. back harmony rule, namely a case

Table 23.3 Vowel phonemes in suffixes

	Unrounded		Rounded	
	front	back	front	back
high	i	ɨ	ü	u
low	e	a		

of true vocalic neutralisation, involves the diminutive–approximative marker *-če* in central dialects. Being an enclitic, it has an underlyingly specified vowel, the front vowel /e/. While it thus ought to take on front suffixes, it does so only as long as it does not undergo raising, e.g., *neyčemde* ‘in my little flute’, *kitapčemde* ‘in my booklet’. As soon as it is raised, its vowel becomes transparent. Rather than determining harmony in suffixes that follow it, it yields to the harmonic category of the stem it modifies, e.g. *neyçide* ‘in a/the little flute’, *kitapçida*, not **kitapçide*, ‘in a/the booklet’, *uyğurçida*, not **uyğurçide* ‘in Uyghur’.

Suffix-initial *-G* must harmonise with respect to front vs. back. It must be velar after front-vocalic stems, and uvular after back-vocalic stems, as shown in the assimilation of the dative suffix *-GA*.

/töpe+GA/ →	töpige	to a/the peak
/taGa+GA/ →	tayıya	to an/the uncle
/köz+GA/ →	közge	to an/the eye
/qız+GA/ →	qızıya	to a/the girl
/sey +GA/ →	seyge	to a/the vegetable
/say+GA/ →	sayya	to a/the river gorge
/xūnen+GA/ →	xunenge	to Hunan
/şendun+GA/ →	şendunğa	to Shandong
/terep+GA/ →	terepke	to a/the side
/tetqıqāt+GA/ →	tetqıqatqa	to (the) research

In a few instances this leads to the formation of internally disharmonic suffixes, e.g. *-ka* and *-qe*, namely where the preceding syllable ends with a back consonant of the opposite category to that of the preceding vowel, e.g.

/taktak+GA/ →	taktakka	to a/the clapper
/xelq+GA/ →	xelqqa	to a/the nation

Rounded vs. Unrounded Harmony

In many Turkic languages, vowels also undergo assimilation to rounding if the vowel in an adjacent syllable is rounded. There is a great deal of variation with regard to which kinds of vowels and which kinds of environments are involved in labial harmony. In what is considered to be the Uyghur language, this applies to high vowels in some dialects, but not in the standard dialect. Extensively applying labial harmony is found only in the Lopnur varieties, which display rounded high and low vowels in many suffixes, e.g. *körsö* ‘if he sees’, *körmöylü* ‘let us not see!’, *qoyçu* ‘shepherd’, *quştu* ‘bird’ (accusative).

In Standard Uyghur this type of roundness assimilation applies only to certain high suffix vowels such as in the first-person possessive suffixes *-(I)m*, *-(I)mlz*, the adjective-forming suffix *-IK* and the ordinal numeral suffix

-(I)nčl. Such vowels also undergo front–back assimilation. Labialisation is inconsistently represented orthographically, e.g. *kölüm* ‘my lake’, *yolum* ‘my road’, *kölimiz* [k^høly‘mız] ‘our lake’, *yolimiz* [jolo‘mız] ‘our road’, *sütlük* ‘having milk’, *tuzluq* ‘having salt’, *oninči* [ʔo‘nontʃı] ‘tenth’.

Vowels in the third-person possessive suffix *-(s)I*, the ablative suffix *-dIn*, etc. do not undergo labial assimilation in the standard language, e.g. *köli* ‘his [etc.] lake’, *yoli* ‘his [etc.] road’, *köldin* ‘from a/the lake’, *yoldin* ‘from a/the road’, not **kölii*, **yolu*, **köldün*, **yoldun*.

Vowel Raising

Vowel raising is another striking feature of Uyghur phonology. In some form or other it applies in all central and Khotan dialects. In the standard language, a low short vowel in a non-initial unstressed open syllable turns into the corresponding high vowel, e.g. /somka+dA/ → *somkida* ‘in a/the bag’, /teyxə+dA/ → *teyxide* ‘on a/the plate’, /ata+lar+i/ → *atilir* ‘their fathers’. In ordinary speech, vowel raising also applies in word-final open syllables, unless there is a pause, but in such instances it is not represented in the official orthographies, e.g. *ata-ana* [ʔat^hıʔa‘nA] (‘father-mother’) ‘parents’.

Consonant Assimilation

Voiced back stops become fricative between vowels, in an initial syllable and also in coda position, e.g. *tögige* [t^høγı‘γɛ] ‘to his [etc.] camel’, *ayiyiya* [ʔajəwə‘wə] ‘to his [etc.] foot’, *jigdige* [dʒıγdi‘γɛ] ‘to (the) oleaster’, *tay* [t^hɑʃ] ‘mountain’.

Alveolars assimilate to a following /s/, e.g. *ketsem* [k^hessem] ‘if I walk’, *sözsiz* [sössız] ‘wordless’, *kücsiz* [k^hyssız] ‘powerless’, *sözleşsek* [söz‘leşsek] ‘if we converse’. An /n/ assimilates to a directly following /l/, e.g. *qalyanlar* [q^halɤal‘la:] ‘remaining ones’. A labial labialises a directly preceding /n/, e.g. *on bir* [ʔom‘bı:] ‘eleven’. An /n/ becomes velar before a velar, and it becomes uvular before a uvular, e.g., *künge* [k^hɤŋ‘gɛ] ‘to a/the day’, *nanya* [nan‘gə] ‘to (the) bread’.

In ordinary speech, an /t/ in coda position assimilates to the preceding vowel. The result is a long vowel and, in the case of /e/ and /a/, lowering, e.g. *bir* [bı:] ‘one’, *bar!* [ba:] ‘go!’, *ber!* [bæ:] ‘give!’, *berdim* [bæ:‘dım] ‘I gave’. Many speakers apply the same rule to /l/ and /y/ as well, especially in non-final position, e.g. *keldim* [k^hæ:‘dım] ‘I came’, *hoyla* [fio:‘la] ‘courtyard’.

A suffix- or enclitic-initial voiced stop becomes voiceless if the preceding syllable ends with a voiceless or devoiced consonant, e.g. *terepke* ‘at a/the side’, *terepke* ‘to a/the side’, *tetqıqatta* ‘in (the) research’, *tetqıqatqa* ‘to (the) research’, *attek* ‘horse-like’, *çüstek* ‘dream-like’. Two voiced back stops of the same category are devoiced if they come to be adjacent to each other, e.g. /beg+GA/ → *bekke* ‘to a/the bey [Turkic governor]’, /tag+GA/ → *taqqa* ‘(on)to a/the mountain’.

Stops closing non-initial syllables undergo devoicing. This applies cyclically, so that a following morpheme undergoes initial devoicing if applicable.

A /b/, /d/, /g/ or /G/ reveals its true identity as a lenis (weak; see pp. 32, 98), if it comes to be followed by a vowel. In this case, /b/ turns into [w], in some dialects only after a short vowel. Devoicing of /mekteb/ 'school', /zawud/ 'factory', /jümeġ/ 'tap' and /ayag/ 'foot' results in: *mektep*, *zawut*, *jümek*, *ayaq*, with the locative suffix -dA: *mektepte*, *zawutta*, *jümekte*, *ayaqta*. Compare the forms occurring before the vowel of the third-person possessive suffix -(s)I 'his [etc.]': *mektiwi*, *zawudi*, *jümiġi*, *ayiyi*.

In most works published in China and in the Soviet Union, this has been described as intervocalic voicing rather than as final devoicing. It is assumed that a voiceless stop undergoes voicing if it comes to precede a vowel. This analysis has influenced the latest orthographic reform. That it is incorrect is proven by examples in which a stop is not realised as voiced; cases which are explained as 'exceptions', e.g. /šerep/ 'honour' /apārāt/ 'device', /rāk/ 'cancer', /ittipāq/ 'alliance', are those which exhibit the third-person possessive forms *šeripi*, *aparati*, *raki*, *ittipaqi*, not **šeriwi*, **aparadi*, **ragi*, **ittipayi*.

When it closes a syllable, an affricate turns into the corresponding fricative, e.g. *ačquč* [ʔaʃ'quʃ] 'key'.

Consonant Dissimilation

A stop becomes fricative if it precedes a consonant, both word-internally and, in the absence of a pause, across word boundaries, e.g. *mektep* [meç'tep] 'school', *teġdim* [tʰeɣ'dim] 'I reached', *yataqta* [jatʰaɣ'ta] 'at a/the lodging', *tamaqni* [tʰamaɣ'ni] 'food' (accusative), *bek tar* [beç'ta:] 'very narrow', *neq pul* [neɣ'pul] 'cash, money'.

Syllable Structure

Uyghur syllables may be structured as CV(V)C, where C represents a consonant and V represents a vowel. VV may be either a long vowel, e.g. /pēr/ *per* 'feather', or a vowel followed by a glide, e.g. /qayt/ *qayt* 'return'.

Other descriptions list the possibilities V, e.g. *u* 'he/she/it', VC, e.g. *at* 'horse', VCC, e.g. *ast* 'area below', and CVCC, e.g. *qorq* 'fear'. However, V, VC and VCC cannot occur since a word that is perceived as beginning with a vowel really begins with a glottal stop and a vowel, e.g. [ʔu]. Although written as such, a syllable cannot end with two consonants. In such cases one of the consonants is either dropped, e.g. *dost* [dos] 'friend', *kent* [kʰen] 'town', or assimilated, e.g. *qorq* [qʰo:q] 'fear', and many speakers assimilate a pre-consonantal glide as well, e.g. *qayt*- [qʰa:t] 'return'. In less natural speech, vowel epenthesis may be used and thus an extra syllable may be created to break up a consonant cluster, e.g. *qorq* [qʰoroq] 'fear'.

Orthography

Few languages have experienced as many orthographic changes as Uyghur. Until the early part of the twentieth century, most Uyghurs used to write varieties of literary Chaghatay, or they used to employ Chaghatay-based

orthographies to render their own language. These orthographies utilised the Arabic script with the inclusion of Iranian-derived devices from Farsi and Tajik and special characters. In 1937, in an attempt to render the system more suitable for the Uyghur language, nearly consistent vowel representation was introduced, and Arabic-specific letters were dropped from the alphabet. This system underwent several modifications. In the meantime, Latin- and Cyrillic-based systems had been introduced to write Soviet Standard Uyghur. In the 1950s, a Cyrillic-based orthography was also introduced in Xinjiang but was replaced by a Latin-based orthography in 1960. This 'New Writing System', *yèŋi yèziq*, was the system officially used throughout the Cultural Revolution in China. However, private use of the Arabic-based orthography continued, and this 'Old Writing System', *kona yèziq*, was officially reinstated in 1983. At the same time, the introduction of two diacritic symbols to distinguish *ü* from *u*, and *ö* from *o* improved the vowel representation system. The latest reform involves the representation of final consonants in the Old Writing System and seems to have decreased consistency within the orthographic system.

The modern Arabic-, Latin- and Cyrillic-based orthographies operate by and large upon the same set of principles. Specific to the Latin-based system is that it utilises three character combinations representing retroflex consonants typical of Mandarin: *ch* [tʃʰ] *zh* [tʃ] and *sh* [ʃ], even though in Uyghur these are realised as alveopalatal *č* [tʃ], [dʒ] and [ʃ] respectively. Specific to the Cyrillic-based system is that Russian loanwords are spelled exactly as in Russian and that thus characters are employed that are alien to Uyghur and are rendered by means of native phonics or by zero.

All three systems now represent and distinguish all vowel phonemes with the exception that /i/ and /i:/ are represented by the same grapheme and that vowel length is not indicated. All three systems go beyond mere phonemic representation. They represent the output of harmonic rules, umlauting, vowel raising and consonant devoicing.

As a result of reform in Xinjiang in 1985, consonant devoicing is now represented inconsistently. It was decided that voiced consonant phonemes should be represented as voiced in all environments, but only in certain words. In all other cases morpheme-final stops are to be written as voiceless in any environment, even between vowels. Thus /jümeġ/ 'tap' is written *jümek*, and the form with the third-person possessive suffix [dʒymɪ'ye] 'his [etc.] tap' is written *jümiki*, previously *jümiġi*.

Morphology

Uyghur morphological principles are by and large the same as those in all Turkic languages. Morphological derivation and marking is accomplished by means of postpositive devices, be they suffixes, enclitics or separate words. In the following, derivational morphemes that obey all rules of harmony and

do not block word-final stress will be considered suffixes, e.g. *baliliriya* [bəlilirə'kə] 'to his [etc.] children'. Those that harmonise but block word-final accent will be called harmonic enclitics or pre-stressed suffixes, e.g. *barsam* ['bɑ:sɑm] 'if I go', *kelsem* ['k'hæ:səm] 'if I come'. Those that do not harmonise but cannot act as independent words, and still undergo consonant devoicing if their first consonant is voiceless, will be referred to as disharmonic enclitics, e.g. *taydek* [t'hɑ'dək] 'mountain-like', *tayaqtek* [t'hɑ'kɑχtek] 'comb-like'.

Morphological modification involves only two morphological classes, nominal and verbal. A nominal morpheme can take on only denominal suffixes or enclitics, while a verbal morpheme can take on only deverbal suffixes or enclitics. A suffix or an enclitic may act as a switch to the opposite category, such as the denominal verbalising suffix *-IA*, e.g. *sözle-* 'converse' ← *söz* 'utterance'.

Nouns

Nouns are not modified by means of definite or indefinite articles. Whether a noun is definite or indefinite can only be deduced from the context in which it is used. Furthermore, there is no morphological gender distinction. Also, there appears to be a reluctance to import feminine-marked loanwords, and the few that have been imported tend to be neglected in favour of more general, unmarked nouns, e.g. general *mu'ellim* 'teacher' < Arabic *mu'falim* being preferred over gender-specific *mu'ellime* 'female teacher' < Arabic *mu'falima*.

Declension

Plural Marking

Plural marking precedes possession marking and case marking. The plural marker is *-IAr*, e.g. *balilar* 'children', *közler* 'eyes'. However, plurality is morphologically less consistently expressed than for example in Indo-European languages. An unmarked noun may occasionally connote plurality, e.g. *bala bar* (literally 'child exists') 'there is a child', 'there are children'. The use of *-IAr* is redundant if plurality is expressed otherwise, e.g. *üç bala* 'three children', *köp bala* 'many children'. Using *-IAr* in such cases expresses variety, e.g. *köp balilar* 'many children of various types'. The same morpheme may also be used to express representation, e.g. *Mirsultanlar* 'Mirsultan and others', 'Mirsultan and his folks', 'Mirsultan and friends'.

Possession Marking

Preceding case marking, possession marking is accomplished by means of person-specific suffixes. The third-person marker is *-(s)i*. The formerly second-person plural possessive suffix *-(I)ñiz* has come to serve as the polite possessive suffix used when addressing one person, and *-IAr* has been added for the equivalent used when addressing more than one person. The other suffixes take on a fully harmonising high vowel when attached to a stem

ending with a consonant. If a word ends with a long vowel, which happens only in loanwords, then *y*, in some dialects *r*, intervenes between the noun stem and any possession marker except the one for the third person.

Case Marking

There are six morphologically marked cases, nominative, genitive, accusative, dative, locative, ablative. Nominative forms are marked by \emptyset , e.g. *Bala oyniwatidu* 'A/the child is playing', *Çay içiñ* 'Drink (some) tea!' vs. accusative *Çayni içiñ* 'Drink the tea!'

Table 23.4 Possessive markers

	<i>1p.sg.</i>	<i>2p.sg.</i>	<i>2p.sg. polite</i>	<i>3p.sg.</i>
<i>bala</i> 'child'	-m balam	-ñ balañ	-ñIz baliñiz	-(s)I balisi
<i>tügme</i> 'button'	tügmem	tügmeñ	tügmiñiz	tügmisi
<i>tay</i> 'mountain'	téyim	téyiñ	téyiñiz	téyi
<i>bel</i> 'waist'	bélim	béliñ	béliñiz	béli
<i>toy</i> 'wedding'	toyum	toyuñ	toyiñiz	toyi
<i>köl</i> 'lake'	kölüm	kölüñ	köliñiz	köli
<i>balilar</i> 'children'	balirim	baliriñ	baliriñiz	baliliri
<i>bala</i> 'calamity'	balayim	balayiñ	balayiñiz	balasi
<i>radiyo</i> 'radio'	radiyoyum	radiyoyuñ	radiyoyiñiz	radiyosi
	<i>1p.pl.</i>	<i>2p.pl.</i>	<i>2p.pl. polite</i>	<i>3p.pl.</i>
<i>bala</i> 'child'	-mlz balimiz	-ñ+IAr balañlar	-ñIz+IAr baliñizlar	-(s)I balisi
<i>tügme</i> 'button'	tügmimiz	tügmeñlar	tügmiñizlar	tügmisi
<i>tay</i> 'mountain'	téyimiz	téyiñlar	téyiñizlar	téyi
<i>bel</i> 'waist'	bélimiz	béliñlar	béliñizlar	béli
<i>toy</i> 'wedding'	toyimiz	toyuñlar	toyiñizlar	toyi
<i>köl</i> 'lake'	kölimiz	kölüñlar	köliñizlar	köli
<i>balilar</i> 'children'	balirimiz	baliriñlar	baliriñizlar	baliliri
<i>bala</i> 'calamity'	balayimiz	balayiñlar	balayiñizlar	balasi
<i>radiyo</i> 'radio'	radiyoyimiz	radiyoyuñlar	radiyoyiñizlar	radiyosi

Table 23.5 Case marking

	'child'	'lake'	'bird'	'his home'	'my buttons'
nom.	bala	köl	quš	yurti	tügmilirim
gen.	baliniñ	kölñiñ	qušniñ	yurtiniñ	tügmilirimniñ
acc.	balini	kölñi	qušni	yurtini	tügmilirimni
dat.	baliya	kölge	qušqa	yurtiya	tügmilirimge
loc.	balida	kölde	qušta	yurtida	tügmilirimde
abl.	balidin	köldin	quštın	yurtidin	tügmilirimdin

Pronouns

The pronominal inventory is fairly symmetrical. Notable among the personal pronouns (Table 23.6) are a few changes in the second-person series. The dative form of *men* 'I' and *sen* 'thou' are not predictable **menge* and **senge* but backvocalic *maņa* and *saņa*, and *sen* has become *si*... in pluralised forms, e.g. *siler* instead of predictable **senler*. Also noteworthy among both personal and demonstrative pronouns is that some of them require an intervening *-niņ* in certain cases. The former second-person plural form *siz* has become the polite form used when addressing one person, and is pluralised by *-ler*. Added to this scheme has been a second-person deferential series: singular *özle(r)*, literally 'selves', and plural *herqaysiliri*, literally 'every which one of them'. The singular demonstrative pronouns are shown in Table 23.7. Plural forms are derived with *-lar* and undergo perfectly predictable case suffixation. Notable among the demonstrative pronouns is the use of compounds to denote considerable distance: *a bu* > *awu* 'that (over there)', *ene šu* > *ešu* ~ *ašu* 'that (way over there)'. The demonstrative pronoun *a* 'that (over there)' is semi-defunct; it is being replaced by *awu* and tends to be avoided if suffixing is required.

The interrogative pronouns *kim* 'who', *nime* 'what', *qaysi* 'which' and *qandaq* 'what kind' are inflected in the predictable manner.

Verbs

Verb Derivation

Actional Modifications with Auxiliary Verb Constructions

The native and most natural manner of expressing actional modifications in Uyghur is by means of auxiliary verbs that follow the main verb marked by the converb in *-(I)p*. In such constructions, *al-* 'take' expresses own benefit, e.g. *Ismiņizni yēziwaldim* 'I have written down your name [for my benefit]', *et-* 'do, make' expresses intensive, sudden action, e.g. *külüwetti* '(S)he burst out laughing', *baq-* 'look, watch' expresses tentative action, e.g. *Yep beqiņ!* 'Please try to eat! Please have a taste!', *bar-* 'go and arrive' expresses progress, e.g. *Buday ösiņ baridu* 'The wheat grows', *ber-* 'give' expresses benefit to others, e.g. *Hikeye eytip berdim* 'I told a story', or action without restraint, e.g. *Külüwe(r)dim* ~ *Külüwedim* 'I laughed and laughed', *bol-* 'become, be' expresses completion, e.g. *Xetni yēzip boldum* 'I finished writing the letter', *çiq-* 'emerge' expresses thoroughness, e.g. *Xetni oqup çiqtim* 'I read the letter carefully from beginning to end', *kel-* 'arrive' expresses consistent action, e.g. *Poloni mundaq qilip keldim* 'This is how I have been making pilaf all along', *ket-* 'depart' expresses sudden change of state, e.g. *Bēsim ayrip ketti* 'I suddenly got a headache', *kör-* 'see' expresses tentativeness, e.g. *Tētip köreyli!* 'Lets taste (it)!', *oltur-* 'settle' expresses brief and/or frequent action, e.g.

Table 23.6 Personal pronouns

	Nominative	Genitive	Accusative	Dative	Locative	Ablative
1p.sg.	men	mēniņ	mēni	maņa	mende	mēdin
2p.sg.	sen	sēniņ	sēni	saņa	sende	sendin
2p.sg. polite	siz	siziņ	sizni	sizge	sizde	sizdin
2p.sg. definite	özle(r)	özle(r)niņ	özle(r)ni	özle(r)ge	özle(r)de	özlerdin
3p.sg.	u	uniņ	uni	uniņa	uniņda	uniņdin
1p.pl.	biz	biziņ	bizni	bizge	bizde	bizdin
2p.pl. polite	siler	silerniņ	silerni	silerge	silerde	silerdin
2p.pl. definite	sizler	sizlerniņ	sizlerni	sizlerge	sizlerde	sizlerdin
2p.pl.	herqaysiliri	herqaysilirniņ	herqaysilirni	herqaysilirige	herqaysiliride	herqaysiliridin
3p.pl.	ular	ularniņ	ularni	ularğa	ularda	ulardin

Table 23.7 Demonstrative pronouns

	Nominative	Genitive	Accusative	Dative	Locative	Ablative
this	bu	buni	buniņ	buniņa	buniņda	buniņdin
such (close)	bundaq	bundaqni	bundaqniņ	bundaqqa	bundaqta	bundaqdin
that	mundaq	mundaqni	mundaqniņ	mundaqqa	mundaqta	mundaqdin
that (far)	šu	šuni	šuniņ	šuniņa	šuniņda	šuniņdin
such (far)	šundaq	šundaqni	šundaqniņ	šundaqqa	šundaqta	šundaqdin
that (there)	u	uni	uniņ	uniņa	uniņda	uniņdin
such (very far)	undaq	undaqni	undaqniņ	undaqqa	undaqta	undaqdin
that (close)	awu	awuni	awuniņ	awuniņa	awuniņda	awuniņdin
that (far)	ene šu	ene šuni	ene šuniņ	ene šuniņa	ene šuniņda	ene šuniņdin
	ešu	ešuni	ešuniņ	ešuniņa	ešuniņda	ešuniņdin
	ašu	ašuni	ašuniņ	ašuniņa	ašuniņda	ašuniņdin

Nime qilip olturisz? 'What are you doing?', *öt-* 'pass' expresses singular occurrence, e.g. *Bu işlarni éytip öttüm* 'I mentioned this matter once', *qal-* 'remain' expresses finite state, e.g. *Bowam ölüp qaldi* 'My grandfather has died', *qoy-* 'put, place' expresses preparatory or customary action, e.g. *Dersni oqup qoydum* 'I studied the lesson', *Salamlışip qoyduq* 'We greeted one another', *sal-* 'place' expresses inadvertent action, e.g. *Yiylap saldim* 'I started crying', *taşla-* 'throw' expresses sudden, uncontrolled action, e.g. *Waqirap taşlidim* 'I gave a sudden scream', *tur-* 'stand' expresses brief or repetitive action, e.g. *Olturup turuq* 'Please take a seat for a moment!', *Bunigya köp kélip turamsiz?* 'Do you come here often?', *yat-* 'be settled' expresses progression, e.g. *Uxlawatidu* 'He is sleeping', *yür-* 'proceed' expresses continuous or regular action, e.g. *Biznikige kélip kétip yürisz* 'You have been coming to our place on a regular basis'.

Finite Verb Forms

Uyghur has developed a rather complex system of tense marking, mostly using compounding rather than simple suffixation. All verbs are inflected in a totally consistent manner, and there are no phonological alternations other than the generally applicable ones.

What appears to be simple suffixation applies only in two cases. To express a simple past, the suffix *-d* is attached to the verb stem and followed by the respective personal possessive marker (with 1p.pl. *-uq* instead of *-mlz*), e.g. *Kirdim* 'I entered', *Kirdi* '(S)he entered', *Kirduq* 'We entered', *Tuttum* 'I seized', *Tutti* '(S)he seized', *Tuttigizlar* 'You seized'. The negation marker is *-mA*, e.g. *Kirmidim* 'I did not enter'.

A so-called indefinite past is expressed by means of the participial suffix *-GAn*. Personal marking may be optionally accomplished by means of a pronominal enclitic, in the case of the third person by means of the enclitic *-dur*. Examples: *Kirgenmen* 'I entered (at some time)', *Tutqandur* '(S)he seized (at some time)', *Kirmigensiz* 'You did not enter (at some time)'. The same suffix is used with the past copula *idi*, the simple past of the defunct verb *i-* be, to mark the pluperfect, e.g. *Kirgen idim* 'I had entered', *Tutqan idinizler* 'You had seized', *Kirmigen idim* 'I had not entered'.

The converb suffix *-b* in conjunction with the past copula *idi* etc. expresses a perfect-like past. The copula may be cliticised and the *b* in the preceding suffix undergoes fricativisation, e.g. *Kirip idim ~ Kiriwidim* 'I have entered', *Tutup iduq ~ Tutuwiduq* 'We have seized', *Kirmep idim* 'I have not entered'.

An indirective past is formed with *-b* plus an enclitic derived from *tur-* 'stand', e.g. *Kiriptimen* 'I entered (reportedly, apparently, etc.)', *Tutuptu* '(S)he seized (reportedly, etc.)', *Kirmeptimen* 'I did not enter (reportedly, etc.)'.

A general present is formed with the suffix *-A* (always raised) and a pronominal marker, third person *-du*, e.g. *Kirimen* 'I enter, will enter', *Tutidu* '(S)he seize(s), will seize'. The corresponding past is expressed by means of

the complex suffix *-At-d*, e.g. *Kirettim* 'I used to enter', *Tuttigizlar* 'You used to seize'.

A focal present (progressive) is marked by converbial *-b*, which comes to be fricativised, and the enclitic *-atA* – a derivative of the verb *yat-* 'lie' in the present – plus personal markers, e.g. *Kiriwatimen* 'I am entering', *Tutuwatimiz* 'We are seizing'. The corresponding past utilises the sequence *-watatt-* plus cliticised personal markers, e.g. *Kiriwatattim* 'I was entering', *Kiriwatattuq* 'We were entering', *Tutiwatattiqlar* 'You were seizing'.

Conditional forms are derived by means of *-sA* + personal markers (3p. sg. and pl. \emptyset , 1p.pl. *-K*), e.g. *kirsem* 'if I enter', *kirsek* 'if we come', *tutsaq* 'if thou seizest', *tutsigizlar* 'if you seize'. Participial *-GAn* and the conditional form of the verb *bol-* 'be, become' are used to express the past conditional, e.g. *kirgen bolsam* 'if I had entered', *tutqan bolsaq* 'if we had seized'.

Obligation is expressed by means of *-A* and the cliticised sequence *-diyandi* + personal markers, or *-diyandu* for the third person, developed from *turyan turur*, e.g. *Kiridiyandimen* 'I am supposed to enter', *Tutidiyandu* '(S)he is supposed to seize'.

The sequence *-mAKçi* expresses intention, e.g. *Kirmekçimen* 'I intend to enter', *Tutmaqçi(du)* '(S)he intend(s) to seize'.

The so-called aorist, regularised to *-(A)r*, is used to express modal nuances, e.g. a suppositional future, e.g. *Kirermen* 'I might enter', *Tutarsiler* 'You might seize', *Oqurmiz* 'We might read'. Negation is indicated with *-mAs*, e.g. *Kirmesmen* 'I might not enter', *Tutmas* '(S)he might not seize'.

A familiar or abrupt imperative, used to address one or more persons, is expressed with \emptyset , i.e. the verb stem alone, e.g. *Kir* 'Enter!', *Tut* 'Seize!' Less abrupt, urging variants are formed with *-GIn(A)*, e.g. *Kirgin* 'Go on and enter!', *Tutqin* 'Go on and seize!', *Kirgine* 'Please, do enter!', *Tutqina* 'Please, do seize!' Polite imperatives are formed with *-ñ(lAr)*, e.g. *Kirig* 'Please enter!' (one addressee), *Kirigler* 'Please enter!' (more than one addressee). The conditional marker *-sA* followed by emphatic *la* expresses polite request addressed to one or more persons, e.g. *Kirsila?* 'Will you not enter?' The optative or voluntative is expressed by means of *-Ay* and *-Ay(li)*, e.g. *Kirey* 'Let me enter!, I shall enter!', *Tutaymu?* 'Shall I seize?', *Kireyli* 'Let us seize!', *Tutaylimu?* 'Shall we seize?' The enclitic *-sun* yields invocatory forms, e.g. *Kirsun* 'May (s)he enter!', *Tutmisun* 'May (s)he not seize!' Desiderative or invocatory forms may also be formed with *-GAy* + personal markers, e.g. *İşletkey* 'If only (s)he would work!', *Tutqaymen* 'If only I would seize!', *Baraysizler* 'If only you would go!' (polite address to more than one), *Kirmigeysiler* 'If only you would not enter!' (more than one addressee).

Syntax

Adjectival and adverbial phrases precede the segments they modify, e.g. *Yağşi oquyuçi yağşi oquydu* ('good student good studies-he') 'A/the good student

studies well', *Yèŋi mektepte yaxşı oquymen* ('new school-at good study-I') 'I study well at the new school'. Relative clauses are confined to the adjectival slot, e.g. *Bu mektepte eŋ yaxşı oquyan oquyučiniŋ ismi Tömür* ('this school-at most good studying student's name-his Timur') 'The name of the student who studies best at this school is Timur'.

A predicative phrase may consist of a noun-based nominal phrase, e.g. *Tömür oquyuči* ('Timur student') 'Timur is a student', or of a verb-based nominal phrase, e.g. *Nime qiliwatisiz?* ('What doing-are-you?') 'What are you doing?'

Thanks to extensive morphological marking of the various parts of speech, sentence structure is very flexible. The basic Uyghur sentence structure is subject + predicate. In a predicate phrase, a verb follows an object, and an indirect object precedes a direct object, e.g. *Tömürge kitapni berdim* ('Timur-to book-accusative gave-I') 'I gave the book to Timur'.

Probably as a result of non-Altaic influence, filling all subject slots and thus using pronouns consistently is fast becoming mandatory. This has led to a type of pronominal doubling, e.g. *Men uyyurmen* 'I am an Uyghur', *Siz yéziwatamsiz?* 'Are you writing?'

Virtually any phrase may serve as a topic at the head of a sentence, e.g. *Xemit, yaxşı oqutquči(dir)* 'As for Hamid, he is a good instructor'. A topicalised subject is set off by means of a pause. Topicalisation can be marked by the conditional copula particle *bolsa*, e.g. *Bügün bolsa, hawa bek yaxşı* 'As for today, the weather is very good'. A constituent may also take the post-predicate position, usually set off by means of a pause, *Yaxşı oqutquči(dir), Xemit* 'A good instructor he is, Hamid'.

Lexicon

Speakers of Uyghur have had close contacts with speakers of numerous other languages for several centuries. Of particular importance were contacts with speakers of other Turkic languages, especially Kazakh and Uzbek, all of which have influenced the Uyghur lexicon. Most Iranian, i.e. Farsi and Tajik, loanwords, many of which are of Arabic origin, seem to have reached Uyghur via the urban varieties of Uzbek and via the common Central Asian heritage of Chaghatay and Iranian literature. Many Arabic words may have reached Uyghur directly through Islamic literature.

In more recent times Russian and Chinese influences have been particularly strong. Russian-derived words are numerous in all Uyghur varieties but are particularly numerous in the varieties used outside Xinjiang. They include internationalisms, e.g. *aptomobil* 'automobile' < Russian *avtomobil*, *kastum* 'Western-style outfit' < Russian *kostjum* 'costume, suit', and also several German-derived words, e.g. *pertuk* 'apron' < Russian *fartuk* < Polish *fartuch* < Middle High German *vortuoch*, *ma(s)ištap* 'scale' < Russian *masštab* < German *Maßstab*. The majority of Chinese loanwords are recent borrowings

from Standard Mandarin, and most of them denote technological, political, bureaucratic and military innovations in China, but not all of them are necessarily adopted into the everyday language. Most of these loanwords are absent from the Uyghur varieties outside Xinjiang, where Russian loanwords tend to take their place. Older Chinese loanwords are found in all varieties of Uyghur. These reached the Uyghur language through contacts with the western Mandarin dialects of Xinjiang, including those used by the Muslim Dungan (Hui) people. Most of these loanwords denote items that were introduced by Chinese immigrants, e.g. *joza* 'table' < Mandarin *juōzi*, *manta* 'stuffed bun' < *mántou* 'steamed bun', *koy* 'largest monetary denomination' < *kuài*.

Dialects

Most Uyghur-speaking communities in Xinjiang, the main Uyghur settlement area, are essentially oases that are spread over a vast area. In the north they are grouped in a line along the Tianshan Mountains all the way from the Xinjiang–Kazakhstan border in the west to the Xinjiang–Gansu border in the east. In the south they lie in an elliptic formation around the Tarim basin and the Taklamakan desert. Prior to the introduction of high-speed public transport, electronic media and uniform education in Standard Uyghur, communication between the communities was infrequent and poor. As a result, there is a great deal of dialectal variation in Uyghur. There have been several proposals with regard to dialect division. According to the view that has been officially adopted in China, there are three major dialect groups.

The central group is divided into a northern subgroup with the dialects of Urumchi (*Ürümči*), Ili, formerly Taranchi, Turfan (*Turpan*), Hami (*Qumul*), and a southern subgroup with the dialects of Kashghar (*Qešqe*), Dolan, spoken by the Dolans of the Kashghar–Aqsu area, often considered a Kashghar variety, Atush (*Atuš*), Tarim, Mughal (*Muŋal*), spoken by the Mughal who migrated to areas along the Yarkand river, Kuchar (*Kuča*).

The Khotan dialect group consists of the dialects of Khotan (*Xoten*), Cherchen (*Čerčen*), Charqliq (*Čaqiliq*).

The Lopnur group consists of the Lopnur dialect.

The Kelpin (*Kelpin*) dialect seems to be a special case: it is spoken in a central dialect area but appears to have a Khotan-type substratum.

There are considerable lexical, phonological, morphological and syntactic differences between the main dialect groups, if all of them can indeed be regarded as being varieties of the same language. Dialectological research has so far been rather superficial, and most of the findings have been presented in Chinese and Uyghur publications strictly within the framework of the official dialect division theory.

Standard Xinjiang Uyghur, a type of regional 'high' language, and the language variety described above, is based on the northern 'central' dialect of

Urumchi, Xinjiang's capital. As a result of its use in region-wide education and communication, northern 'central' influences on other dialects throughout Xinjiang have been increasing in recent decades.

The Standard Uyghur language that was developed in the Soviet Union and is now used particularly in Kazakhstan is based on the Ili dialect and thus does not differ greatly from the standard language used in Xinjiang. The most notable differences between the two standard languages are orthographic and lexical. Lexical differences are found mostly in the area of loanwords. While both varieties share an inventory of older Chinese- and Russian-derived terms, most recent loanwords are Chinese in Xinjiang Standard Uyghur and Russian in the standard language used west of the border.

Apparently, no specific written standard language has been developed in the Uyghur enclaves of Uzbekistan's Ferghana Valley where a specific dialect group has developed.

Further Reading

- Hahn, R. F. (1991) *Spoken Uyghur*, Seattle: University of Washington Press.
- Jarring, G. (1933) *Studien zu einer osttürkischen Lautlehre*, Lund: Borelius; Leipzig: Otto Harrassowitz.
- (1946–51) *Materials to the Knowledge of Eastern Turki* 1–4, Lund: Lunds Universitets årsskrift.
- Kajdarov, A. T. (1966) 'Ujgurskij (novoujgurskij) jazyk', in V. V. Vinogradov (ed.) *Jazyki narodov SSSR: Tjurkskie jazyki* 2, Moscow: Nauka.
- Nadzhip, E. N. (1971) *Modern Uigur*, Moscow: Nauka.
- Pritsak, O. (1959) 'Das Neuuigurische', in J. Deny *et al.* (eds) *Philologiae Turcicae Fundamenta* 1, 525–63.
- Sadvakasov, G. S. (ed.) (1989) *Stroj ujgurskogo jazyka*, Almuta: Akademija Nauk Kazaxskoj SSR.
- Schwarz, H. G. (1992) *An Uyghur–English Dictionary*, Bellingham: Western Washington University Press.
- Wei, C. Y. (1989) 'An introduction to the modern Uyghur literary language and its dialects', *Wiener Zeitschrift für die Kunde des Morgenlandes*, 79: 235–49.

24 Yellow Uyghur and Salar

Reinhard F. Hahn

Yellow Uyghur

Introduction

The Turkic language referred to here as Yellow Uyghur – *Sariy Uiyur* or *Sariy Yuuyur* – is known in China as 'Western Yughur' (*Xibu Yuguyu*). It is used by Yughur people [sə'rəy jo'γur] who live in the western part of Sunan county (see pp. 9–10). Yughurs living in the eastern parts of the same county use Eastern Yughur or Ingar [əŋ'gər], a Mongolic language. These Yughurs call themselves [jəra jəkər] 'Yellow Uyghur', while their Turkic-speaking relatives refer to them as [χa'ra jo'kor] 'Black Uyghur'. Mandarin Chinese is the language of the Yughurs who live in the townships of Huangnibao and Qiantan of the Minghua district. Chinese serves as the lingua franca between the Turkic-, Mongolic- and Chinese-speaking Yughurs.

The Yughurs are believed to be descendants of Orkhon Valley Uyghurs who in the ninth century moved to what is now western China. Originally consisting predominantly of nomadic herders, this minority group has been absorbing various influences from their Chinese, Mongolian and Tibetan neighbours and officially converted to Buddhism, though certain elements of their previously practised shamanist religion still survive.

Probably owing to its relatively limited geographical spread, Yellow Uyghur has only insignificant dialectal variation. Modern Yellow Uyghur is not an officially written language. Chinese is used in written communication by almost all Yellow Uyghur people. At this time, Yellow Uyghur is one of the least studied Turkic languages. There are some significant inconsistencies and discrepancies in the few descriptions that have been published, and a number of important aspects still need to be investigated, especially those pertaining to phonology.

Until recently it used to be widely believed in the west that Yellow Uyghur was very closely related to Modern Uyghur, and it was assumed by some to be an isolated dialect of Modern Uyghur. Yellow Uyghur ought to be considered to belong to the northeastern group of the Turkic languages and

to be most closely related to Shor and some other Turkic varieties used in South Siberia. This classification is typified by the form [a'zaq] 'leg', 'foot', in which * δ has become z rather than y : cf. Modern Uyghur *ayaq*, Shor *azaq*.

Sound System

Table 24.1 Yellow Uyghur vowels

	Front		Back	
	unrounded	rounded	unrounded	rounded
high	i	ü	ï	u
non-high	e	ö	a	o

Whether or not vowel length is phonemic remains to be determined. Chinese sources show no vowel length distinction.

According to Chinese sources, the phoneme /i/ tends to be realised as central [ə], as is /i/ in certain environments. The high unrounded vowels tend to assimilate to preceding consonants. They are apicalised after dental fricatives, e.g. *kisi* [k^hə'zɪ] 'person', *qozi* [qo'zɪ] 'lamb'. They also assimilate to preceding retroflex fricatives and affricates, e.g. *malji* [mol'tʃɪ] 'herder', *šīyan* [ʃɪ'ʁan] 'mouse', 'rat'.

Sound harmony appears to be fairly defective, at least on the surface. It applies consistently to low suffix vowels, e.g. *at* + -GA [a^ht'qa] 'to the horse', *kelin* + -GA [kelən'ge] 'to the daughter-in-law', *qus* + -GA [qus'qa] 'to the bird', *gümüš* + -GA [kimus'ke] 'to the silver', but *at* + -ni [a^ht'nə] 'the horse' (acc.) versus *kelin* + -ni [kelən'nə] 'the daughter-in-law' (acc.).

Yellow Uyghur has fortis (strong)–lenis (weak) distinction among the consonants even in native words. However, all stops and affricates are realised as voiceless, and fortis are expressed by way of aspiration. Voiced lenes occur only between vowels. Fortis consonants tend to be pre-aspirated; e.g. *od* [ot] 'fire' vs. *ot* [o^ht] 'grass'. Pre-aspiration applies across sonorants; e.g. *ultur* [u^htur] 'front'. This process, which tends to be defined as an 'aspirated

Table 24.2 Yellow Uyghur consonants

	Labial	Alveo-dental	Alveo-palatal	Retroflex	Velar	Uvular	Laryngeal
stop	p ^h , p	t ^h , t			k ^h , k	q ^h , q	ʔ
affricate			tʃ ^h , tʃ	tʃ ^h , tʃ			
fricative	f, v	s, z	ʃ	ʂ, ʐ		χ, ʁ	h
nasal	m	n			ŋ	ɴ	
liquid		l, r					
glide			j				

vowel' phenomenon, may well be related to what have been regarded as glottalised vowels in Tuva, e.g. [ot] 'fire' vs. [oʔt] 'grass' (see pp. 31, 98, 404). What makes the Yellow Uyghur case even more intriguing is that pre-aspiration may come to precede and devoice a vowel, e.g. *at* [a^ht] > [hət] 'horse', *eteg* [e^htek] > [hɛ'dek] 'front panel in a jacket'.

Under the influence of Amdo Tibetan, Yellow Uyghur has come to permit clustered syllabic onsets, e.g. *namtar* [(r)nam't^har] 'tale'. This may even occur in native words, e.g. in the particle [tur] ~ [tro].

Other Features

There are some significant morphological simplifications in Yellow Uyghur. For example, possession marking does not distinguish between singular and plural, and in casual speech only the third person is distinguished: -(s)ə vs. -(ə)ŋ for all other persons.

The Yellow Uyghur lexicon retains a number of items that are derived from Old Turkic, e.g. *mən* 'broth' < *bün*, and *kem* 'illness' < *kem*. It has been greatly enriched by borrowings from Sinitic, Mongolic and Amdo Tibetan varieties.

Salar

Introduction

The Salar are one of China's official ethnic minorities. Being predominantly devout Hanafi Muslims, they live on the easternmost periphery of the Turkic-speaking Islamic world, in an area where Sinitic, Mongolic, Tibetan and Turkic cultural and linguistic spheres converge (see p. 10).

Early written records about the Salar are very scarce. According to their own oral history, the Salar emigrated to their current areas from the Samarkand area in present-day Uzbekistan some time between 1368 and 1393, at the very beginning of China's Ming dynasty (1368–1644). Since that time their language has been variously influenced by the surrounding non-Turkic ones. The results include the use of Chinese devices such as [ʃɪ] 'be' < *shi*, and [vulə] 'in order to' < *weile*, noun + numeral construction rather than the typically Turkic numeral + noun construction, and the Mongolic device of expressing possession by way of dative forms, e.g. [aŋa avo ʔɪfki vara] 'to-him son two exist' = 'he has two sons'; cf. Modern Uyghur *Uniŋ ikki oylı bar*. Eastern influences also account for the presence of a retroflex series in the Salar consonant inventory.

However, underneath these strata of assumedly more recent eastern influences we detect elements and traits that are shared by Turkic languages spoken farther to the west. One of these is the phonological process of liquid assimilation ('deletion'), e.g. *Varyar* [va(ɪ)ʁa(ɪ)] '(S)he will go', *Gelmiš* ['ke:mɪʃ] '(S)he (reportedly) came'. This is a typical feature of Modern Uyghur and also occurs sporadically elsewhere, including the Salar dialect group of

Turkmen. There are several elements and features in Salar that are usually associated with the Oghuz or southwestern languages and are unknown or rare in other Turkic languages. These include the lexical items [el] 'hand', 'arm' (cf. Turkish *el* 'hand'), [səχ] ~ [səχ] 'right (side)' (cf. Turkish *sağ*), [oɪɪ] 'noon' (cf. Turkish *öğle*), [peɪ] 'thus' (cf. Turkmen *beyle*, Turkish *böyle*), [se(ɪ)tʃɪ] 'sparrow' (cf. Turkmen *serçe*, Turkish *serçe*), [o'xen] 'lung' (cf. Turkmen *öyken*), and the specific use of the indirective suffix *-miş*.

Furthermore, like the southwestern Turkic languages and unlike Modern Uyghur and related languages, Salar is a *sari*-type language: [sa'rɪ] 'yellow', cf. Turkish *sarı*, but Modern Uyghur *seriq*. It distinguishes initial fortis and lenes and has a tendency toward southwestern-type initial *v*-, e.g. [va:] 'go!', cf. Turkish *var*, but Modern Uyghur *bar*.

These findings tempt one to give credence to Salar oral history and to assume that Salar originated as an Oghuz Turkic language and during the course of its speakers' gradual eastward migration acquired various influences from southeastern- and northwestern-type Turkic languages as well as from non-Turkic ones. It had been assumed earlier that Salar is an isolated dialect of Modern Uyghur, mostly on the basis of phonological features such as liquid assimilation and vowel raising.

Salar is not a written language. Its speakers tend to use Chinese in written communication. However, there have been reports about the discovery of old documents written in Salar or in a Salar-influenced form of Chaghatay.

There is some dialectal, mostly phonological variation in Salar. The prevalent view in China is that there are the Salar varieties of Jiezi, Qingshui, Baihuang, Gandu, Dahejia and Ghulja.

Sound System

Table 24.3 Salar vowels

	Front		Back	
	unrounded	rounded	unrounded	rounded
high	i	ü	ɨ	u
non-high	e	ö	a	o

Whether or not Salar has vowel length distinction or only conditioned vowel lengthening is not yet clear.

There is a strong tendency towards neutralisation of front and back vowels, mostly in the form of medialisation. The back phoneme /i/ tends to be realised as [ɨ]. The high unrounded vowels tend to assimilate to preceding consonants: they are apicalised after dental fricatives, e.g. [sɨ'nɨχ] 'bone', and they also assimilate to preceding retroflex segments, e.g. [tʃɨzɨ] '(measuring) rule'.

Table 24.4 Salar consonants

	Labial	Alveo-dental	Alveo-palatal	Retroflex Velar	Uvular	Laryngeal
stop	p ^h , p	t ^h , t			k ^h , k	q ^h , q
affricate			tʃ ^h , tʃ	tʂ ^h , tʂ		ʔ
fricative	f, v	s, z	ʃ	ʂ, ʐ	x, ɣ	χ, ʁ
nasal	m	n			ŋ	ɴ
liquid		l				
glide			j			

Vowels after a glottal stop and before voiceless consonants (including lenes) tend to be devoiced, and high unrounded vowels tend to be fricativised when they follow glottal stops, e.g. [sq^h] 'sell!', [ʔɛʃɛx] 'donkey', [ʔɪʃt] 'dog'. Non-high vowels in non-initial open syllables tend to be raised, e.g. *balalar* [balə'la:] 'children'.

Sound harmony does apply in Salar, although this tends to be somewhat unstable and mostly obscured due to vowel neutralisation, e.g. *varyur* [va:ɹu:] 'will go' vs. *vergür* [ve:ɹu:] 'will give', *kiši* + *-lar* [k^hɨʃ'la:] ~ [k^hɨʃ'la:] 'people', cf. Modern Uyghur *kişiler*.

Salar observes fortis vs. lenis distinctions among the consonants in all positions. However, all stops and affricates are realised as voiceless, and fortis are expressed by way of aspiration. Voiced lenes occur only before nasals. Retroflex consonants occur only in loanwords, primarily in those of Chinese (Mandarin) origin.

Velar and uvular stops tend to be fricativised in syllabic rhymes, e.g. *ešek* [ʔɛʃ'ɛx] 'donkey', *toq* [t^hoχ] 'chicken'.

Alveodental stops undergo palatalisation before /i/, e.g. *digna-* [tʃɪŋ'na] 'listen', *etim* [ʔɛ'tʃɨm] 'my flesh'.

Other Features

There are various types of simplification in Salar morphology, for example non-distinction between singular and plural in personal marking.

The Salar lexicon abounds with loanwords from neighbouring languages. Most of these are of Chinese origin. Furthermore, there are numerous Arabic-, Iranian- and Tibetan-derived loanwords.

Further Reading

Yellow Uyghur

Geng, Sh. M., and Lei, X. Ch. (1985) *Xibu Yugu-yu jianzhi*, Beijing: Minzu Chubanshe.

Malov, S. E. (1957) *Jazyk želtyx ujugurov: slovar' i grammatika*, Almuta: Izdatel'stvo Akademii Nauka Kazaxskoj SSR.

Tenišev, E. R. (1966) *Jazyk želtyx ujugurov*, Moscow: Nauka.

——— (1976) *Stroj saryg-jugurskogo jazyka*, Moscow: Nauka.

Thomsen, K. (1959) 'Die Sprache der Gelben Uiguren und das Salarische', in J. Deny *et al.* (eds) *Philologiae Turcicae Fundamenta* 1, Wiesbaden: Steiner, 564–8.

Salar

Drimba, V. (1968) 'Sur la classification de la langue salare', *Ural-Altäische Jahrbücher* 40: 200–13.

Dwyer, A. (1996) *Salar phonology*, Ph.D. thesis, University of Washington.

Hahn, R. F. (1988) 'Notes on the origin and development of the Salar language', *Acta Orientalia Academiae Scientiarum Hungaricae* 17: 235–75.

Kakuk, S. (1961) 'Textes salars', *Acta Orientalia Academiae Scientiarum Hungaricae* 13: 95–117.

Poppe, N. (1953) 'Remarks on the Salar language', *Harvard Journal of Asiatic Studies* 16: 438–77.

Tenišev, Ė. R. (1963) *Salarskij jazyk*, Moscow: Izdatel'stvo vostočnoj literatury.

——— (1964) *Salarskie teksty*, Moscow: Nauka.

——— (1976) *Stroj salarskogo jazyka*, Moscow: Nauka.

Thomsen, K. (1959) 'Die Sprache der Gelben Uiguren und das Salarische', in J. Deny *et al.* (eds) *Philologiae Turcicae Fundamenta* 1, Wiesbaden: Steiner, 564–8.

25 South Siberian Turkic

Claus Schönig

Introduction

South Siberian Turkic is a relatively young regional group of Turkic. It can be divided into four main branches: Altay Turkic, Yenisey Turkic, Sayan Turkic and Chulym Turkic. Though these languages exhibit numerous common traits, they differ considerably in detail. The description given below can only outline some of their peculiarities.

The first three branches are represented by five literary languages: Altay, the Yenisey Turkic languages Khakas and Shor, and the Sayan Turkic languages Tuvan and Tofa. All are written in Cyrillic alphabets. Altay, Khakas, Shor and Tuvan were written in Latin alphabets during a short period in the 1920s and 1930s. Before the construction of the modern South Siberian literary languages, Mongolian was used for literary purposes in the area.

The Altay-kiži, Teleut and Telengit dialects constitute the southern group of Altay Turkic. The northern group consists of the Tuba-kiži, Qumanda-kiži and Qū-kiži dialects. A written Altay language, called Oyrot up to 1948, was established in 1922 on the basis of the Altay-kiži dialect. A pre-revolutionary predecessor, used for Orthodox missions, was based on the Teleut dialect.

Khakas was created in 1926, on the basis of the Saɣay and Qača dialects, as a literary language for Turkic tribes speaking different dialects, Saɣay, Qača, Qızıl, Shor and the now extinct Beltir and Qoybal dialects. Its predecessor was, like pre-revolutionary Altay Turkic, written with a missionary alphabet. The closely related Shor language has been written since 1927. Its base is the Mrass dialect, which differs considerably in some respects from the Kondoma dialect.

Literary Tuvan, developed around 1930, is based on the central dialect of Ottuq-Daş and Bayan-Qol in the Kem (*Uluɣ-Xem*) valley. It was at first written in a Latin script and, from 1941 onwards, in a Cyrillic one. Besides the central dialect there are two other main dialect groups, Western Tuvan and Southeast Tuvan, partly spoken in Mongolia and China. Another Sayan Turkic idiom, Toja, is sometimes treated as a Northeast Tuvan dialect. At the end of the 1980s, Tofa, formerly called Karagas, became a literary language written in a Cyrillic alphabet.

Chulym Turkic, which has no written form, consists of three dialects, one of them being Küerik.

Classificatory Features

As far as the classificatory features are concerned, the four main branches can be characterised by the development of Old Turkic intervocalic *-*δ*-, which has become -*y*- in Altay Turkic, -*z*- in Yenisey Turkic, -*y*- or -*z*- in Chulym Turkic, and -*d*- in Sayan Turkic. Altay Turkic has close ties to Kirghiz. Its northern dialects lean towards Yenisey Turkic, which is somehow related to the Kipchak languages of the Volga–Ural region. Both Altay and Yenisey Turkic are closely related to Siberian Tatar, mainly to its eastern dialects. Sayan Turkic exhibits features making it closer to Yakut, with which it must have had long-standing regional contacts. The other South Siberian languages also show connections to this most remote Turkic branch. Tofa can be regarded as a transitional language between Sayan Turkic and Yakut. Chulym Turkic occupies a position between Altay and Yenisey Turkic. Parts of it have been influenced by the language of Uyghur fur traders from the beginning of the seventeenth century onwards. The literary languages have undergone strong Mongolian and Russian influence. There is also a Russian influence on the syntax.

Sound Systems

Phonetics and Phonology

South Siberian Turkic languages generally display the symmetrical basic set of eight short vowel phonemes, /a/, /e/, /i/, /i/, /o/, /ö/, /u/, /ü/. Khakas also exhibits a reduced /i/ going back to **i*, and an /i/, written и, in suffixes neutral with respect to the front vs. back opposition. Tofa has a similar distinction as well as an opposition /e/ vs. /è/.

All South Siberian languages have long vowels, normally as products of contractions. The written languages express them graphically with two vowel signs. Altay Turkic has word-final long vowels going back to the loss of -*G*, -*b*, -*d* and -*y*. Even intervocalic -*ŋ*- can get lost due to contraction. The same is true of Yenisey Turkic, where -*n*- can also be lost between vowels. Sayan Turkic also exhibits loss of intervocalic -*l*- and -*r*-, mainly in monosyllabic verb stems, e.g. *bōr* < **bolur* 'becomes', *kīr* < **kirer* 'enters'. A typical product of contraction in Khakas is the long neutral *ī*, e.g. *kīk* < *keyik* 'deer'. Chulym Turkic exhibits rising diphthongs as contraction products.

There are also, particularly in Sayan Turkic, glottalised vowels in first syllables, e.g. *aʔ*, as reflexes of old short vowels combined with strong consonants (fortes, see pp. 31, 98). The phonetic expression of this feature may be inspired by a South Samoyedic substratum. Glottalisation is not consistently expressed in the orthography.

In Tuvan and Tofa, vowels are sometimes nasalised under the influence of neighbouring nasal consonants.

As for the consonants, all South Siberian languages possess labial, dental and

velar obstruents: voiceless stops, voiced stops and fricatives. The phoneme /k/ has the front variant *k* and the back variants *q* and *χ*. In Turkic words, back /k/ is realised as *q* in Altay Turkic, and as *χ* in Khakas (except in the ethnonym *χакас*). The phoneme /g/ has the back variant *γ*, whereas the front variant is *g* or a more front fricative. There are also the glide *y*, the sibilants *s* and *š*, and the affricate *č*. In standard Khakas, however, *š* and *č* are missing in words of Turkic origin, their predecessors having developed into *s*. The liquids /l/ and /r/ and the nasals /m/ and /ŋ/ do not occur initially in native words. As in Old Turkic, **nēme* 'what, something' is the only native word in which the dental nasal /n/ may occur initially. The Old Turkic palatalised nasal *ñ* has normally changed to *y*, but is sometimes represented by *n* in Sayan Turkic (pp. 105–6). In the spoken language, it may be represented by a nasalised *y*, which, however, also occurs in words that do not go back to forms with *ñ*.

The tendency towards voiceless initial obstruents is strongest in Altay and Yenisey Turkic. The Khakas and Shor orthographies only show initial voiceless obstruents in native words, whereas the Altay orthography exhibits initial *b*-, even if the pronunciation is *p*-. Sayan Turkic does not show this tendency: even the orthography exhibits initial *b*- and *d*-. Except Tofa, none of the literary languages display initial *g*- in native words. Tuvan has numerous cases of initial *χ*- both in front and back initial syllables. No rule has been found for this distribution. The same holds true for the distribution of initial *h*-, *k*- and *q*- in Tofa.

Diachronic Remarks

Sayan Turkic *š* goes back to Old Turkic *š* and *č*. The latter is still represented by affricates only in Altay and Shor. Even in Altay, the initial *č* tends to become *š*. In Shor, the old *č* has become *j* between vowels, but has coincided with *š* in initial and final positions (as in some Khakas varieties). Old Turkic *y* has been preserved word-finally and between vowels, whereas it has affricative realisations word-initially, e.g. Altay *j* ~ *d*', Yenisey and Sayan Turkic *č*. In Yenisey Turkic and Tofa, if the word contains a nasal consonant, initial *y* is represented by *n* and *ñ*, respectively. In some cases, mainly in Tofa, this nasalisation even occurs in words that do not contain a nasal consonant, but a velar one. In all South Siberian languages, old *b* is represented by *m* if the word contains a nasal consonant, sometimes even if the word does not contain a nasal or velar one. In Tofa, old *n* is sometimes represented by a nasalised *y* between vowels. Old final -*n* is sporadically represented by *ŋ*, mainly in Yenisey Turkic, Tofa, Chulym, and in some Altay case suffixes.

Phonotactics and Morphophonology

All South Siberian languages exhibit front vs. back sound harmony in words of native Turkic origin, whereas loanwords may remain non-harmonic. The neutral Khakas *ī* occurs after front and back syllables, e.g. *parīr* 'goes, will go'. In Tofa, certain suffixes contain a non-harmonic *é* ~ *i* under the influence of a neighbouring *y*.

Labial harmony is developed in different ways. In Altay Turkic, high rounded vowels of the first syllable influence subsequent high vowels of the lexical stem, whereas low vowels are only sporadically affected. After *o* and *ö* of the first syllable, low vowels are rounded in certain derivational and grammatical suffixes. In written Khakas, labial harmony affects high vowels in lexical stems, but not in suffixes. In Tuvan and Tofa, rounded vowels affect high vowels in all suffixes.

South Siberian obstruents tend towards word-final devoicing and intervocalic voicing or loss. Voiceless consonants occurring between vowels normally go back to consonant clusters. Tuvan *p* and *b* are even fricativised into β . There are, however, exceptions. Altay \check{c} is not voiced between vowels. Tofa is still more resistant against voicing tendencies. Consonants occurring after glottalised vowels are not voiced between vowels. The obstruents *s* and *t* remain as such, *p* becomes a bilabial Φ , *k* and *q* develop into *h*, and \check{s} changes into a palatalised *h'*.

Furthermore, in Sayan Turkic, word-initial consonants may be weakened if a suffix with initial vowel is added to the word, e.g. Tuvan χip - 'burn' \rightarrow *qivar* 'burns', Tofa *taš* 'stone' \rightarrow *daji* 'his stone'.

The South Siberian Turkic languages have phonotactic rules that regulate the realisation of neighbouring consonants within primary stems and at suffix borders. Certain consonants are assimilated to the preceding sound, e.g. with respect to the distinction voiced vs. unvoiced. Suffix-initial *l* and *n* are preserved after vowels and changed into *t* after voiceless consonants. Their realisations after other stem consonants are shown in Table 25.1. Khakas and Tofa follow the same rules, while the other languages apply individual systems. After nasals, there is a tendency towards progressive nasal assimilation. Altay is the only language that dissimilates *l* and *n* to *d* in this position, but its genitive suffix *-Niŋ* preserves its *n*. The tendency to dissimilate the last member of a cluster of two non-obstruents to an obstruent is strongest in Altay, though it is also observed in Chulym, Shor and Tuvan.

Certain suffixes have initial *b* or β after vowels, *b* after voiced consonants, *p* after voiceless consonants, and *m* after nasal consonants, e.g. the Tuvan negative suffix *-BA*.

Table 25.1 Suffix-initial consonants of the plural and accusative suffixes

	Stem-final consonant	Chulym	Altay	Khakas	Shor	Tuvan	Tofa
-LAr	-y, -r	l	l	l	l	l	l
	-l	d	d	l	l	d	l
	nasal	n	d	n	n	n	n
-NI	-y, -r	n	d	n	d	n	n
	-l	d	d	n	d	d	n
	nasal	n	d	n	n	n	n

Syllables are usually open or closed by one consonant. Two consonants may occur syllable-finally only in the case of *r* plus an obstruent, e.g. *tört* 'four'. A final cluster *-rK* is mostly dissolved by inserting a high vowel, e.g. *çiriç*, *qiriç* 'forty' < *qirq*. Sayan Turkic maintains *-rK* between vowels, e.g. Tuvan *qorqar* 'fears', *böryü* 'his cap', but tends to change it into **-rt* otherwise, e.g. **qorq- > qort-* 'fear' and **börk > bört* 'cap'. Dropping of high vowels in non-first syllables mainly occurs in Yenisey and Sayan Turkic, particularly in front of suffixes. In Khakas, loss of high vowels in the first syllable may produce word-initial consonant clusters, e.g. *tlo* 'hummock'.

Morphology

Nouns and Adjectives

Word Formation

Among derivational suffixes forming denominal nouns there is a wide range of diminutive suffixes. One type includes Altay *-čAK*, Khakas *-(I)JAK*, Sayan Turkic *-čIK*, e.g. Altay *ulčaq* 'little boy', Tuvan *çölčük*, Tofa *höljük* 'little lake'. The diminutive in *-Aš*, e.g. Tuvan *bičigeš* 'booklet', may combine with *-čIK* to form *-čIGAš*, e.g. Shor *quščayaš* 'birdie'. Other forms go back to *-GAč* or *-AK*, e.g. Tofa *qušqaš* 'birdie', Khakas *ineyek* 'grandma'. Old Turkic *qina* has survived in Tofa as a non-harmonic particle, e.g. *qis qina* 'little girl'.

The suffix *-čI* is used to form agent nouns, e.g. Altay, Tuvan *aŋči*, Khakas *aŋji*, Tofa *aŋši* 'hunter' \leftarrow *aŋ* 'game, beast'. The suffix *-LIK*, which forms abstract nouns in most other Turkic languages, is not used much, except in Altay, e.g. *bayliq* 'wealth' \leftarrow *bay* 'rich'.

Among suffixes forming deverbal nouns, *-t* is relatively frequent, e.g. Shor *šiyüt* 'exit' \leftarrow *šiq-* 'go out', Tuvan *olut* 'seat' \leftarrow *olur-* 'be seated'.

The suffix *-l* seems to be copied from Mongolic, e.g. Altay, Tuvan *töröl* 'relatives' \leftarrow *törö-* 'give birth ~ be born'. The shape of the actor noun suffix *-AččI* or *-ÄččI*, also present in Yakut, seems to be influenced by Mongolic as well.

The suffixes forming denominal adjectives include an impressive number of detentive suffixes which often modify colour adjectives, e.g. Altay *-ZU*, *-ZImAK*, *-(Il)tlm*, Shor *-(Il)Dlrlm*, *-ZlBAn*, Tuvan *-ZImAr*, *-lIdlr*.

The relational suffix *-KI*, e.g. Altay *qišqi* 'belonging to the winter', is widely used. Another suffix forming possessive adjectives with the meaning 'belonging to', Altay *-Nyl*, Khakas *-Nī*, Tuvan *-Nī*, etc., goes back to a combination of the genitive suffix **-nIŋ* and *KI*, e.g. Altay *aŋniyi*, Khakas *aŋni* 'belonging to beasts'.

The old suffix **-lIG* is represented as *-LU* in Altay, e.g. *attu* 'having a horse, horseman' (similar to Kirghiz) and as *-LIG* in the other languages, e.g. Khakas, Shor *attiy*, Tuvan, Tofa *aŋttiy*. As in all northeast Turkic languages, South Siberian Turkic has lost the privative suffix **-sIz*.

There are also suffixes copied from Russian such as Khakas *-ay* ~ *-oy*, e.g. *pedagogičeskay* 'pedagogical'.

Comparison

Comparison is expressed syntactically, mainly by means of the ablative, e.g. Khakas *Čir aydaŋ uluy* 'The earth is bigger than the moon'. Altay Turkic also uses the converb *körö* ← *kör-* 'see', e.g. *Ülyä körö qis d'aqšī* 'The girl is better than the boy'.

A superlative is formed with particles meaning 'most, very' such as Tofa *eŋ*, *dam*, Tuvan *eŋ*, *tergīn* (also combined), Khakas *iŋ*, *tīŋ*, Altay *eŋ*, *sürekey*, *sürēn*, *tīŋ*, e.g. Tofa *dam semis* 'very fat, fattest', Tuvan *eŋ uluy* 'very big, biggest', Khakas *tīŋ pözīk* 'very high, highest'. Superlative expressions are also formed with genitive and ablative constructions, e.g. Khakas *čaxsīdaŋ čaxsī* 'best' ← *čaxsī* 'good', Tuvan *ekiniŋ ekizi* 'best' ← *eki* 'good', Shor *qistiŋ čaqšizi* 'the best girl', Altay *ūldaŋ körsü ūl* 'the most intelligent boy'.

Intensive forms of adjectives may be created by means of partial reduplication, e.g. Tofa, Tuvan, Shor, Altay *qap-qara* 'pitch-black' ← *qara* 'black'.

Declension

The South Siberian languages have six cases in common: nominative in \emptyset , genitive in *-NIŋ*, dative in *-GA*, accusative in *-nI*, locative in *-DA* and ablative with forms such as Sayan Turkic *-DAn*, Altay and Yenisey Turkic *-DAŋ*, Chulym *-DAŋ* and *-DIn*. After stems ending in vowels, the dative suffixes have long vowels resulting from contraction.

In addition to its other uses, the dative may also serve to denote location. In Tofa, the locative is also used as a partitive.

All standard languages except Altay exhibit additional cases. Sayan Turkic and Khakas have a directive, e.g. Tuvan *-čA*, *-DİvA*, Tofa *-šA*, Khakas *-Zar*. Yenisey and Chulym Turkic possess an instrumental or comitative case, e.g. Khakas *-nAŋ*, Shor *-BA*, Chulym *-(BA)IA*. Since the initial *n* of the Khakas suffix has no phonotactic variants, this suffix is identical with the ablative in words ending in nasals, e.g. *aŋnaŋ* 'from the beast' or 'with the beast'.

The singular possessive suffixes are *-(I)m*, *-(I)ŋ* and *-(z)I*, Tofa *-(s)I*. The plural forms are the following:

Table 25.2 Plural possessive suffixes

	Chulym	Altay	Khakas	Shor	Tuvan	Tofa
1p.	-(I)vIs	-(I)bIs	-(I)bIs	-(I)bIs	-(I)vIs	-(I)bIs
2p.	-(I)ŋnAr	-(I)GAr	-(I)ŋAr	-LArIŋ	-(I)ŋAr	-(I)ŋAr
3p.	-(z)I	-(z)I	-(z)I	-LArI	-(z)I	-(s)I

The replacement of the Common Turkic pronominal **(I)z* by the nominal plural **-LAr* in the second person is a main distinctive feature of South

Siberian Turkic. The initial position of *-LAr* in the second- and third-person Shor forms is unique in the Turkic family. Shor is also the only language of the area that marks plurality in the third person.

Some special case forms are used after possessive suffixes. A 'pronominal *n*' generally appears after third-person suffixes. The genitive suffix is *-Iŋ*, e.g. *atiniŋ* 'of his horse'. The dative suffix is *-A* in Khakas and Altay, *-GA* in Tuvan, Tofa and Shor. The latter two may also fuse with the possessive and dative markers into *-IŋA*, which in Tofa is homophonous with the second-person singular dative form. The accusative suffix is *-n*, e.g. *atin* 'his horse', in some Chulym dialects *-nI*. The Altay and Yenisey Turkic ablative after *-n* is *-nAŋ*, in Khakas identical with the instrumental. The instrumental in *-BA* lacks 'pronominal *n*', e.g. Shor *ačcaziba* 'with his/her money'.

There is less variation in the other persons. Tuvan and Shor suffixes are regular. As for the dative, the Khakas first- and second-person dative marker is *-A*, e.g. *atima* 'to my horse', *atiŋa* 'to your horse'. Both *-GA* and *-A* occur in the first person in Tofa and Altay, in Tofa also in the second person. Tofa also has the first- and second-person accusative form *-In* besides the regular *-nI*.

Pronouns

The first- and second-person personal pronouns are 1p.sg. Altay, Shor, Tuvan, Tofa *men*, Khakas *min*, 2p.sg. Altay, Shor, Tuvan, Tofa *sen*, Khakas *sin*, 1p.pl. Altay *bis(ter)*, Shor *pis*, Khakas *pīs*, Tuvan *bis*, Tofa *biʔs*, 2p.pl. Altay *sler*, Shor, Tuvan *siler*, Tofa *sīler*, Khakas *sīrer*. An example of the irregular declension of singular pronouns is given in Table 25.3. The plural forms are treated like nouns. The third-person pronoun is identical with the demonstrative pronoun *ol*. In Shor, however, the personal pronoun *ol* has the plural *ilar*, while the demonstrative *ol* has *anar*.

The basic Turkic demonstrative pronouns are represented by *ol* for distance, and for nearness Altay *bu*, Khakas *pu*, Shor *po*, Tuvan, Tofa *bo*. Some languages exhibit additional demonstratives, e.g. Shor, Khakas *tīgī*,

Table 25.3 Declension of first-person singular pronouns

	Chulym	Altay	Khakas	Shor	Tuvan	Tofa
nom.	men	men	min	men	men	men
gen.	mēŋ	meniŋ	minīŋ	mēŋ	mēŋ	mīŋ, mim
acc.	mēni	meni	minī	mēni	meni	meni
dat.	mā	mege, mā, mē	mayā	mayā, mā	meŋē, mē	meŋē, mē
loc.	mendē	mende	minde	mende	mende	mēnde
abl.	mendin	meneŋ	minneŋ	meneŋ	menden	mēndēn
instrumental	meŋvele		minneŋ	mēŋme		
directive			minzer		menče	

Table 25.4 Declension of the 3p.sg. pronoun *ol*

	Chulym	Altay	Khakas	Shor	Tuvan	Tofa
nom.	ol	ol	ol	ol	ol	ol
gen.	anıñ	onıñ	anıñ	āñ	ōñ	onuñ
acc.	anı	onı	anı	anı	onu	onu
dat.	ere, aya	oyo, aya, o	ayā	aya, ā	añā, ā	aña, ā
loc.	anda	ondo, anda	anda	anda	ında	indi
abl.	andin	onoñ	annañ	anañ	ōn	unun, ından
instrumental	anıñvila		annañ	āñma		
directive			anıñzar		olče	

Tuvan *dō*, Tofa *tē*. Examples of irregularities in declension are given in Table 25.4. The reflexive pronouns go back to Old Turkic *boδ* 'body, shape', e.g. Altay *boy*, Shor, Khakas *pos*, Tuvan, Tofa *bot* plus possessive suffixes, in Tofa with an additional *bile*, e.g. Altay *boyım*, Khakas *pozım*, Tofa *bodum bile* 'myself'.

Interrogative pronouns are Altay, Shor, Khakas *kem*, Tuvan *qim*, Tofa *qum* 'who', Altay *ne*, Khakas *nime*, Shor *nō*, Tuvan *čū*, Tofa *čū* 'what'.

Indefinite pronouns in Altay, Khakas, Shor and Tofa consist of combinations of interrogatives and the particle *da*, in Altay and Khakas also *la*, e.g. Tofa *qum da* 'whoever'. Tuvan uses interrogatives followed by *bir*, e.g. *qim bir* 'whoever'. Yenisey Turkic places *bir* in front of the interrogative, e.g. Khakas *pır nime*, Shor *pir nebe* 'anything'.

South Siberian languages have numerous quantifiers meaning 'all', e.g. Altay *ončozi*, *bastıra*, Khakas *pray(zi)*, *polyan(i) na*, *tiksı*, Shor *parčan*, *parčın*, *tōza*, *tügeze*, *tekši*, Tuvan *tödü*, *büyü*, *çamiq*, *dōza*, *amitan*, Tofa *tödö*.

A typical South Siberian phenomenon is the existence of pronominal verbs such as Tuvan *mınča-* 'act like this', *ınçalap* 'acting like that', Altay *qayıp* 'acting how'.

Numerals

The first decade of cardinal numerals is almost identical with those of other Turkic languages and exhibits mainly phonetical variations (see Table 25.5).

The analytical way of expressing tens, consisting of numerals of the first decade plus the word for 'ten', is partly applied already from 'thirty' onwards. Most other Turkic languages use this form only for 'eighty' and 'ninety'. For 'hundred', the Common Turkic **yüz* is used in its various phonetic variants. 'Thousand' is expressed by Chulym *mıñ*, Altay, Khakas, Shor, Tuvan *muñ*, whereas Tofa uses the loanword *tısičci* (< Russian тысяча).

Ordinals are derived from cardinals by means of Tuvan *-KI*, Tofa *-(I)skI*, and *-(I)ñI* in the other languages.

Collective numerals are derived with *-(U)IA* ~ *-(U)IAZI* and *-LAp* in Altay, e.g. *onu*, *onula(zi)*, *ondop* 'ten together'. The remaining languages have

Table 25.5 Cardinal numbers from 1 to 10

Numeral	Altay	Tuvan	Chulym	Shor	Khakas	Tofa
1	bir	bir	pir	pir	pır	birē
2	eki	iyi	igi	iygi	ıki	i?hi
3	üç	üş	üç	üş	üş	üş
4	tört	dört	tört	tört	tört	dört
5	bes	beş	pes	peş	pis	beş
6	altı	aldı	altı	altı	altı	a?ti
7	d'eti	çedi	yedi	cetti	çitı	cēdi
8	segis	ses	segis	segis	sigis	se?hes
9	togus	tos	togus	togus	toyis	to?os
10	on	on	on	on	on	on

Table 25.6 Tens of cardinal numerals

	Chulym	Altay	Khakas	Shor	Tuvan	Tofa
20	yerve	d'irme, irme	çibirgi	çegirbe	çerbi	i?hyon
30	odus	odus, otus	otıs	odus	ußen	ußen
40	qırıq	qırıq, törtön	çiriç	qırıq	dörten	dörten
50	elig	bezen	ilg	elig	bezen	bejen
60	alton	altan	alton	alton	aldan	a?lton
70	yedon	d'eten	çiton	çetton	çeden	çeden
80	sėekson	segizen	sigizon	segizon	sezen	se?hezon
90	toqson	toyuzon	toyizon	toyuzon	tozan	to?hozon

suffixes with initial vowels which absorb stem-final vowels: Khakas *-OlAñ*, Tuvan *-AlĀn*, *-AldIrzi*, Tofa *-Ān*, *-ĀIn*, e.g. Khakas *ıköleñ*, Shor *iygele*, Tuvan *iyelēn*, *iyeldirzi*, Tofa *i?hēn*, *i?hēIn* 'two together'.

Distributive numerals are derived with *-DAñ* in Altay and Yenisey Turkic. Khakas also uses *-LArDAñ*, and, like Tofa, *-LArLAp*. Khakas and Tofa also exhibit the suffix *-(l)Ar*, Tofa sometimes with the addition of *-nl*. Tuvan uses reduplicated numerals, often followed by the suffix *-LAp* or the word *qıldır*, e.g. *iyi-iyi(lep)*, *iyi-iyi qıldır* 'by two, two each'.

There are interrogative quantifiers of several kinds, e.g. Altay *qanča*, Shor *qanče*, Khakas *ninje*, Tuvan *qaş*, Tofa *qa?ş* 'how many'.

Verbs

Denominal Verb Derivation

Suffixes forming verbs from nominal stems are mostly similar to those found in other Turkic languages, e.g. *-A*, *-Ar*, *-LA*, *-ZIrA*, *-ZIn*.

In periphrastic derivation of verbs from nouns, intransitive verbs are formed with a nominal plus a verb for 'become, be', Altay, Tuvan, Tofa *bol-*,

Chulym, Khakas, Shor *pol-*. For transitives, verbs denoting 'do, make' are used, variants of **qil-* in Sayan Turkic and of **et-* in the other languages. Tofa also uses *qin-*, mostly in onomatopoeic verbs.

Deverbal Verb Derivation

South Siberian has four kinds of voice markers, passive, middle, cooperative-reciprocal and causative suffixes. They combine with each other according to specific rules.

The passive suffix *-(I)l* also occurs after stems in *-l*, e.g. *alil-* 'be taken'. The middle suffix, which may also have passive meaning, is *-(I)n*. Sayan Turkic also displays the compound forms *-DIIn* and *-ttIn*. The cooperative-reciprocal suffix is *-(I)š*, in Khakas *-(I)s*. The causative suffixes *-t*, *-DIr* and *-GIs* are common to all South Siberian languages. The causative suffix *-(I)s* does not exist in Khakas, while *-KIr-* is infrequent in Tuvan and Altay. The causative forms can also be used to express passive meaning (see p. 56).

One of the most important actionality suffixes, Tuvan *-GAIA*, Tofa *-GIIA*, Khakas *-GIA* etc., expresses frequentativity, often combined with detensive nuances, e.g. Tuvan *nomčuyala-* 'read (a little) here and there'.

Actional modifications can also be expressed with combinations of *-A* and *-(I)p* converbs with auxiliary verbs. More than one auxiliary verb may be used. Some constructions are used for phase specification. Thus, transformativity (p. 42) is specified with **-(I)p iđ-* 'send', e.g. Altay *-(I)p iy-*, Khakas, Shor *-IvIs-*, Tuvan, Tofa *-IvIt-*, e.g. *biživit-* 'write down'. The most important auxiliaries go back to the verbs *oltur-* 'sit (down)', *tur-* 'stand (up)', *yat-* 'lie (down), lie' and *yürü- ~ yori-* 'go'. Combinations with such verbs have also been used to renew the focal present tense (see pp. 43, 114–15). The converb suffix can, mainly in Yenisey Turkic, be omitted after stem-final consonants, e.g. Khakas *pas sal tur-* 'write down' (several times) < **pasip salip tur-* ← *pas-* 'write', *sal-* 'put', *tur-* 'stand'.

Negation is normally expressed by *-BA*, though some verb forms have special negations. Possibility is expressed by various combinations of converbs and auxiliary verbs, e.g. Altay, Tuvan *-A al-*, Altay, Khakas, Shor, Chulym *-p al-*, Tuvan, Tofa *-p bil-*, Altay, Khakas, Tuvan, Tofa *-p bol-*, Shor *-p pol-*. To express impossibility, different forms containing the negation suffix are used, e.g. Chulym *-p alva-*, Shor *-p polba-*. There are also combinations with participles such as Khakas *-JAğ polba-*.

Finite Verb Forms

Finite verbs take on different personal markers. One set is almost identical with the possessive suffixes and is used in the simple past and the conditional. Altay Turkic sometimes exhibits *-K* in the first person plural. It also uses possessive forms as short forms in other paradigms. This is also the case with the *-GAN* perfect in Tofa.

A second set contains, in the first two persons, personal suffixes going back to postposed personal pronouns. The Sayan Turkic suffixes are still

identical with the pronouns. The Khakas suffixes are 1p.sg. *-m* (after vowels) and *-Bin* after consonants, 2p.sg. *-ZIŋ*, 3p.sg. *∅*, 1p.pl. *-BIs*, 2p.pl. *-ZAr*, 3p.pl. *-LAr*. Shor has a 1p.sg. variant *-Bim*, but no *-m* and no plural marker in the third person. Chulym has first- and second-person forms with low vowel variants, e.g. *-mAn*. Altay differs from Khakas mainly by 2p.sg. *-ZIn* and 2p.pl. *-(I)GAR*. In Khakas, Altay, Shor and Chulym, a suffix-final *n* may be lost due to contraction in front of the suffixes *-m* and *-ŋ*, e.g. *Alyanım* 'I have taken' > *Alyam*.

The finite tense, aspect and mood forms include a set that is common to most Turkic languages, e.g. simple pasts in *-DI*, post-terminals (perfects) in *-GAN*, optatives in *-GAY*, and so-called aorists (present-futures) in *-(V)r*, negated *-BAs*. Yenisey and Sayan Turkic also exhibit narrative pasts in *-JIK*.

The literary forms possess focal presents as the result of renewal processes (see pp. 43, 114–15). The Tuvan ones consist of the converb in *-(I)p* plus one of the auxiliaries *tur(u)*, *olur*, *čor(u)* and *čidir*. The other languages display similar forms. There are also less focal present markers such as Khakas, Shor *-(p)čA*, Altay *-At*, Khakas, Shor *-A Dir*, Tofa *-AdIr*. The corresponding imperfects are, e.g., Altay *-AtI*, Khakas *-(p)čAtKAN*. Other combinations with past copula particles may create various additional items, e.g. pluperfects from *-GAN*.

Habitual presents include Altay *-AtAn*, habitual perfects Khakas, Shor *-JAğ*, inferential presents Shor *-pčattir*, Tuvan *-AdIr*, and inferential pasts Tuvan *-ptIr*.

Non-finite Verb Forms

The forms in *-(V)r* and *-GAN* are also used as intraterminal and post-terminal participles, respectively. The present (intraterminal) participles also include the more focal items such as Tofa *-A turu*, *-A oluru* etc., Tuvan *-p turar*, *-p olurar* etc., Khakas *-(p)čAtKAN*, Shor *-(p)čItKAN* and Altay *-AtAn*. There is also a typical South Siberian participle of future possibility in Altay *-GAdIy*, Khakas *-GAdIğ*, Shor *-GAdAg* and Tuvan *-GIdAg*. Participles in *-GAIAK* express an event that has not yet taken place (participium nondum facti 'participle of the not yet done'), e.g. Tuvan *kelgelek kiži* 'a person who has not come yet'. Both types are missing in Tofa. Both can also function as finite predicates, e.g. Tuvan *Kelgideğ* 'Maybe (s)he will come', *Kelgelek* '(S)he has not come yet'.

The *-(I)p* converb functions as in other non-Oghuz languages. The *-A* converb, which has only survived in auxiliaries and complex verb forms, still displays forms with high vowels in Sayan Turkic. Its variant after vowels is *-y*, which in Khakas fuses with the final stem vowel to the neutral *i*. These converbs are negated in Altay with *-BAy*, Khakas *-Bin*, Shor *-BĀn*, Tuvan *-VAyn*, Tofa *-Beyn(ša)*.

Tuvan *-BIšĀn* and Tofa *-BIšĀğA* are intraterminal converbs denoting events that have started in the past and still going on at a given orientation point. The Tofa form also functions as a finite item. Anteriority is expressed

by Sayan Turkic *-GAš*, Altay *-GAžIn*, Khakas *-AbAs* (originating from the Qača dialect). Terminative converbs ('until') are formed with *-GIžA* in Sayan Turkic, and *-GANčA* in the other languages. Abtemporal converbs ('since') exist in Yenisey Turkic, e.g. Khakas *-(G)AII*, Shor *-AIA*, and Altay, e.g. *-AIA*, *-GAII*.

Conditionals are formed with *-ZA*. Tuvan first- and second-person markers are formed according to the pattern *-ZI* + personal marker + *-ZA*, e.g. *barzimza* 'if I go'; cf. third-person forms such as *barza* 'if (s)he goes'. Counterfactual forms, mostly occurring after conditional phrases, are Altay *-(V)r edi* and *-GAy edi*, Khakas *-(V)rčIK*, Shor *-(V)r edi*, Tuvan *-(V)r iyik*, Tofa *-GAy ik*.

Adverbs

Though adjectives can normally be used as adverbials without any marker, there are some suffixes that derive adverbs, e.g. Tofa *ekkiđi* 'well' ← *ekki* 'good'. The suffixes are often fossilised case forms, e.g. Tuvan *qizhin* 'in winter' ← *qış* 'winter'.

A number of interrogative adverbs go back to case forms of pronouns meaning 'which', e.g. Altay *qandiy*, Shor *qay(i)*, Khakas *χayzi*, Khakas *qayda*, Tuvan *qayda* 'where', Tuvan *qayin* 'from where'.

Other Morphological Categories

The numerous postpositions of the South Siberian languages are considerably different. Certain postpositions, e.g. those going back to **soŋ* 'later, after', play important roles in forming subordinate clauses with participles as predicates.

Conjunctions of Turkic origin include Khakas *paza*, Tuvan *baza*, *bolyaš* 'and', Tuvan *azi* 'or'. Some conjunctions are Russian loanwords, e.g. *a* 'but'. In many cases, the way of using the conjunctions is copied from Russian patterns.

South Siberian languages have numerous particles, e.g. the limitative particle *IA* ~ *LA*. Sentence questions are marked with *BA*, Sayan Turkic *be*.

Syntax

Nominal Phrases

Unmarked nouns can have singular or collective meaning. The use of the plural suffix *-LAR* is relatively restricted.

The lack of a privative suffix **-sIz* is compensated for by constructions in which the noun referring to the lacking entity combines with an adjective meaning 'no-existing', e.g. Altay *tuzazi joq*, 'useless', Khakas *χoryis čox*, Tuvan *qoryuš čox* 'fearless'. Altay Turkic adds, like Yakut, a third-person singular possessive suffix to the noun. This method is also found in other languages, e.g. Khakas *χani čox* 'without blood', Tuvan *düp čox*, *düvü čox* 'without ground'.

Compounds can be formed with or without a third-person singular possessive suffix, e.g. Tuvan *čā čil balı* 'New Year's ball' ← *čā čil* 'New Year', *bal* 'ball' + possessive suffix, Altay Turkic *Altay kiži* 'Altay-people' ← *Altay* 'Altay mountains', *kiži* 'person'. There are also coordinative juxtapositions such as Altay, Shor *ada-ene* 'parents' ('father-mother').

The Sentence

The standard languages do not exhibit any major differences from other Turkic languages with respect to clause structure, word order, patterns of complex clauses, etc. Records of spoken language may reveal considerable divergences from the standard.

Subordinated clauses are mainly formed with converbs and participles. Participles as non-finite predicates are used for relative, complementary and adverbial clauses.

A relative clause whose head is not coreferential with the first actant may use a possessive suffix as subject marker. Yenisey Turkic and Tofa do not use this option. Altay Turkic attaches the possessive suffix to the head of the relative clause, and the subject of the relative clause is usually not genitive-marked. Tuvan displays both types. It also has cases with a genitive-marked subject and a possessive suffix added to the participle. Khakas also constructs relative clauses introduced by forms of the pronoun *χay(zi)* 'which', probably constructions copied from Russian.

Lexicon

The South Siberian languages have been exposed to a good deal of foreign lexical influence. One Mongolic layer can be connected to Classical Mongolian, and a west Mongolian layer goes back to the Dzungar empire (fifteenth to seventeenth centuries). Tuvan exhibits strong Khalkha influences dating back to the period of the Altan Khans. There are also borrowings from Buryat, particularly in Tofa. Mongolic cultural loanwords play a role comparable to that of Arabic-Persian loanwords in Islamic Turkic languages. Large parts of the lexicon, particularly political, technical, and scientific terms, are copied from Russian, playing a role comparable to that of western loanwords in Turkish. South Samoyedic and Yeniseyic (Ket) elements go back to the non-Turkic origin of some tribes. Iranian influences are also observed, e.g. Khakas *nan* 'bread' ← Persian *nān*. Even some Arabic words have found their way into this area, mainly into Altay Turkic, e.g. *urmat* 'reputation' ← Arabic *ħurmat*.

Further Reading

Babuškin, G. F. and Donidze, G. I. (1966) 'Šorskiy jazyk', in N. A. Baskakov (ed.) *Jazyki narodov SSSR 2, Tjurkskie jazyki*, Moscow: Nauka, 467-81.

Baskakov, N. A. (1966) 'Altajskij jazyk', in N. A. Baskakov (ed.) *Jazyki narodov*

- SSSR 2, *Tjurkskie jazyki*, Moscow: Nauka, 506–22.
- (1975) *Grammatika xakasskogo jazyka*, Moscow: Nauka.
- and Toščakova, T. M. (1947) *Ojrotsko-russkij slovar'*, Moscow: Gosudarstvennoe izdatel'stvo inostrannyx i nacional'nyx slovar'ej. (Contains a short Altay grammar.)
- and Inkižekova-Grekul, A. I. (1953) *Xakassko-russkij slovar'*, Moscow: Gosudarstvennoe izdatel'stvo inostrannyx i nacional'nyx slovar'ej. (Contains a short Khakas grammar.)
- Castrén, M. A. (1857) *Versuch einer koibalischen und karagassischen Sprachlehre*, St Petersburg: Kaiserliche Akademie der Wissenschaften.
- Dul'zon, A. P. (1966) 'Čulymsko-tjurkskij jazyk', in N. A. Baskakov (ed.) *Jazyki narodov SSSR*, 2, *Tjurkskie jazyki*, Moscow: Nauka, 446–66.
- Dyrenkova, N. P. (1940) *Grammatika ojrotskogo jazyka*, Moscow, Leningrad: Izdatel'stvo Akademii Nauk SSR.
- (1941) *Grammatika šorskogo jazyka*, Moscow, Leningrad: Izdatel'stvo Akademii Nauk SSR.
- Isxakov, F. G. and Pal'mbax, A. A. (1961) *Grammatika tuvinskogo jazyka. Fonetika i morfolologija*, Moscow: Izdatel'stvo vostočnoj literatury.
- Karpov, V. G. (1966) 'Xakasskij jazyk', in N. A. Baskakov (ed.) *Jazyki narodov SSSR*, 2, *Tjurkskie jazyki*, Moscow: Nauka, 428–45.
- Menges, K-H. (1955–6) 'The South-Siberian Turkic languages', *Central Asiatic Journal* 1: 107–36 and 2: 161–75.
- (1963) 'Die sibirischen Türksprachen', in B. Spuler (ed.) *Handbuch der Orientalistik 5. Turkologie*, Leiden, Cologne: Brill, 72–138.
- (1959) 'Das Sojonische und Karagassische', in J. Deny et al. (eds) *Philologiae Turcicae Fundamenta* 1, Wiesbaden: Steiner, 640–70.
- Pritsak, O. (1959a) 'Das Altaitürkische', in J. Deny et al. (eds) *Philologiae Turcicae Fundamenta* 1, Wiesbaden: Steiner, 568–98.
- (1959b) 'Das Abakan- und Čulymtürkische und das Schorische', in J. Deny et al. (eds) *Philologiae Turcicae Fundamenta* 1, Wiesbaden: Steiner, 598–640.
- Rassadin, V. I. (1971) *Fonetika i leksika tofalarskogo jazyka*, Ulan-Ude: Burjatskoe knižnoe izdatel'stvo. (Contains a Karagas glossary.)
- (1978) *Morfolologija tofalarskogo jazyka v sravnitel'nom osveščanii*, Moscow: Nauka.
- Sat, Š. Č. (1966) 'Tuvinskij jazyk', in N. A. Baskakov (ed.) *Jazyki narodov SSSR*, 2, *Tjurkskie jazyki*, Moscow: Nauka, 387–402.
- Schönig, C. (1991) 'Das Lenatürkische und die sprachlichen Merkmale des nordöstlichen türkischen Areals', in B. Brendemoen (ed.) *Altaica Osloensia*, Oslo: Universitetsforlaget, 263–85.
- Tenišev, Ė. R. (ed.) (1968) *Tuvinsko-russkij slovar'*, Moscow: Sovetskaja Ėnciklopedija.

26 Yakut

Marek Stachowski and Astrid Menz¹

Introduction

Yakut exhibits a set of important classificatory features that distinguish it from all other Turkic languages, e.g. the shift **-δ- > -t-*, e.g. *ataχ* 'leg', preservation of original long vowels, sporadic *-t* for *-š* and *-z*, e.g. in the suffixes *-BIt* and *-BAI*, corresponding to *-mlš* and *-mAz*, loss of the locative, and emergence of new cases. Besides these features there is a set of other sound changes such as **y- > s-*, **s- > θ-*, e.g. *en* 'you' < **sen*.

These facts as well as numerous loanwords make a Yakut text unintelligible to speakers of other Turkic languages. Yakut has a rather mixed character. At least two migration waves of genuine Turkic groups seem to have left their traces in its phonetics. There is considerable older Mongolic and later Russian influence, and a still little explored impact of Tungusic and Yeniseian substrate languages.

The scholarly investigation of Yakut begins with Otto Böhtlingk's grammar (1851). Another valuable early piece of work is Edward Piekarski's Yakut dictionary (1907–30).

Sound System

Phonetics and Phonology

The vowel system consists of eight short vowels, eight long vowels and four diphthongs. It works according to three phonological oppositions, back vs. front, rounded vs. unrounded, low vs. high, the last one concerning monophthongs only. The first element of a diphthong is high and phonologically dominant. The second element of the diphthong *ia* is pronounced as a sound between *a* and *e*. The consonants *χ* and *k* are allophones, the former occurring before *a* and *o* and after low vowels (monophthongs and diphthongs), while the latter occurs in all other positions. In the modern

¹ M. Stachowski has contributed the following sections to this chapter: Introduction, Sound System, Morphology, Dialects. Syntax and Lexicon are by A. Menz.

Table 26.1 Vowels

	Monophthongs				Diphthongs	
	low		high		unrounded	rounded
	unrounded	rounded	unrounded	rounded		
back	a, ā	o, ȯ	ı, ȳ	u, ū	ia	uo
front	e, ē	ö, ȳ	i, ȳ	ü, ū	ie	üö

Table 26.2 Consonants

	Labial	Alveodental	Palatal	Velar	Uvular	Glottal
stop	b, p	d, t		g, k		
fricative		s			ɣ, χ	h
affricate			j, č			
nasal	m	n	ń	ŋ		
oral glide			y			
nasal glide			ȳ			
liquid		r, l, ł				

alphabet, they are written with different letters. Similarly, *ł* occurs in back words and *l* in front words only. In the modern alphabet as well as in the present chapter both are transcribed as *l*. The phonetic value of the phoneme /j/ varies between the palatal *d'*, and the affricate *j*.

As far as diachronic changes are concerned, the sibilants **s*, **z*, **š* and **č* have become *s*. The new sounds *j* and *č* do not occur in native Turkic words. Except in some Russian loanwords, *s* becomes *h* between vowels, e.g. *kus* 'duck' → *kuha* 'his/her duck'. The nasal glide *ȳ* has developed from **ń*.

The final syllable of a Yakut word is normally accented. Exceptions include compounds, e.g. *bihikki* 'both of us' ← *bihigi* 'we' + *ikki* 'two', pronominal forms with the emphatic element *sV-*, e.g. *súbu* 'just this' ← *bu* 'this', *sámanna* 'just here' ← *manna* 'here', some Russian loanwords, e.g. *náka* 'however' ← *odnáko*, and some other words with unexplained accent shift, e.g. *túra* 'not at all'. The accent may in a few cases have distinctive function, e.g. *áta* 'although' vs. *átá* 'his/her/its name', *aráha* 'obviously' vs. *arāhá* 'its variety'.

Words used as exclamations may be pronounced with a special intonation plus lengthening of vowels of the final syllable, low ones becoming long monophthongs and high ones diphthongs, e.g. *Doyōr* 'oh, friend!' ← *doyor* 'friend', *Doyoruom* 'Oh, my friend!' ← *doyorum* 'my friend', *Sordōχpuon* 'Oh, I am so unhappy!' ← *sordōχ* 'unhappy' + *-puon* < *-pun* < *-Bln* 'I am'.

Morphophonology

Vowels and consonants are distributed according to the front vs. back harmony. Native Yakut words contain either back or front sounds. Accordingly, suffixes have back and front variants, e.g. *-a* and *-e* in *suluha* 'his/her star' ← *sulus* 'star', *silihe* 'its root' ← *silis* 'root'.

One labial harmony rule predicts that a stem with a high rounded vowel is followed by a suffix with a high rounded or a low unrounded monophthong or a rounded diphthong, e.g. *kūs* 'force' → accusative *kūhū*, possessive *kūhe* [force:POSS3SG.], *ūr-* 'lay, put' → *ūruom* [lay:PROSP.1SG.]. A stem with a high unrounded vowel can only be followed by a suffix with an unrounded vowel, e.g. *kīs* 'girl' → *kīhī* [girl:ACC.], *kīha* [girl:POSS3SG.], *īr-* 'get furious' → *īriem* [get furious:PROSP.1SG.]. A second rule predicts that a stem with a low rounded vowel is followed by a suffix with a rounded monophthong or diphthong vowel, e.g. *bolot* 'sword' → *bolotu* [sword:ACC.], *boloto* [sword:POSS3SG.], *kötör* 'bird' → *kötörü* [bird:ACC.], *kötörö* [bird:POSS3SG.], *toyonnō-* 'interpret, explain' → *toyonnuo* [interpret:PROSP.3SG.], *kömölös-* 'help' → *kömölöhüö* [help:PROSP.3SG.]. The quality of diphthongs with respect to sound harmony is determined by their dominant first component, which is high. They are thus not affected by the second rule, e.g. *suol* 'way' → *suola* [way:POSS3SG.], not **suolo*, *tüös* 'breast' → *tüöhe* [breast:POSS3SG.], not **tüöhö*.

Table 26.3 summarises the distribution of sound harmony variants. Consonants undergo progressive and regressive assimilation at morpheme borders, e.g. *tl* > *tt*, *tb* > *pp*, *tɣ* > *kk*, *nl* > *nn*, *nb* > *mm*, *nɣ* > *ŋŋ*, *χl* > *χt*, *χb* > *χp*, *χɣ* > *χχ*. Examples are *at* + *-LAr* > *attar* [horse:PL.], *at* + *-Blt* > *appit* [horse:POSS1PL.], *at* + *-Glt* > *akküt* [horse:POSS2PL.], *oron* 'bed' + *-LAr* > *oronnor* [bed:PL.], *oron* + *-Blt* > *orommut* [bed:POSS1PL.], *oron* + *-Glt* > *oronggut* [bed:POSS2PL.], *bölöχ* 'group' + *-LAr* > *bölöχtör* [group:PL.], *bölöχ* + *-Blt* > *bölöχpüt* [group:POSS1PL.], *bölöχ* + *-Glt* > *bölöχχüt* [group:POSS2PL.].

Table 26.3 Sound harmony

in the syllable preceding the suffix	The vowel	
	in the first syllable of the suffix	
/a/, /ɨ/, /iə/	/a/ or /ɨ/ or /iə/	
/e/, /i/, /ie/	/e/ or /i/ or /ie/	
/u/, /uo/	/a/ or /u/, or /uo/	
/ü/, /üö/	/e/ or /ü/ or /üö/	
/o/	/o/ or /u/ or /uo/	
/ö/	/ö/ or /ü/ or /üö/	

In the following presentation, suffixes will be given in standardised forms, in which capital letters indicate regular morphophonological variants.

Table 26.4 Consonant and vowel variants in suffixes

Standardised suffix consonant	After		voiceless consonant		nasal	/χ/	
	high vowel	low vowel	/l/	/r/			
L		l	l	d	t	n	t
B		b	b	b	p	m	p
T		t	l	d	t	n	t
G	g	γ	g	g	k	ŋ	χ

Standardised suffix vowel	After front		After back	
	rounded vowel	unrounded vowel	rounded vowel	unrounded vowel
A	ö	e	o	a
I	ü	i	u	i

The consonants /l/ and /r/ can only be followed by one voiced consonant in intervocalic position. Otherwise the cluster of three consonants is reduced, e.g. *olor-* 'sit' + *-Bl* > *Olorbut* [sit:PST.3SG.] '(S)he sat', but **Olor-t-but* > *olorput* [sit:CAUS.:PST.3SG.] '(S)he set'.

Two rules are crucial for the correct construction of declension and conjugation forms. First, short high vowels of the stem-final syllable are unstable in some words, e.g. *tayis-* (< **taqış-*) 'go out' → *taxsar* (< **taqšar*) 'goes out', *χaris* 'caution, care' → *χarsa* 'his/her caution', but *suruk* 'letter' → *suruga* 'his/her letter' and *sulus* 'star' → *suluha* ~ *sulsa* 'his/her star'. Secondly, diphthongs and long vowels of the stem-final syllable are dropped before suffix-initial diphthongs or long high monophthongs, e.g. *taptā-* 'love' + 1sg. prospective *-iam* → *Taptiam* 'I shall love', *kepsē-* 'talk' + nominal suffix *-ī* → *kepsī* 'talking, talk'.

Orthography

The first complete alphabet of Yakut was created by the German scholar Otto Böhtlingk and published in his grammar of 1851. A new alphabet based on the International Phonetic Alphabet and composed by the Yakut linguist S. A. Novgorodov was introduced in 1922. Seven years later it was replaced by a new Latin alphabet, which was in use until the Cyrillic alphabet was introduced in 1939.

The history of the written norm, albeit not very long, is rather stormy. Not only alphabets and orthographic rules, but also opinions on what is correct have changed several times in the last 70 years. One controversy concerned the application of morphological and phonetic principles, i.e. whether one should write *Sanābatpyn* or *Sanābappyn* [think:NEG.PRES.1SG.] 'I do not think', *kuoratka* or *kuorakka* [city:DAT.] 'in the city', etc. Another conflict concerned the graphic adaptation of new Russian loanwords, whether one

should write *socializm* or *sissiyelīsim* 'socialism', *programma* or *bīragrāmma* 'programme', etc. There was also disagreement about the choice between *a-i* and *o-u* in words such as *χamis* ~ *χomus* 'Jew's harp' and *alīs* ~ *olus* 'very'. The present-day norm is a result of compromises between various tendencies.

Morphology

Nouns and Adjectives

Declension

The plural suffix is *-LAR*, e.g. *tabalar* 'reindeers', *saptar* 'threads', *kustar* 'ducks', *bölöχtör* 'groups', *sireyder* 'faces', *büörder* 'livers', *kīmnar* 'sparks', *oronnor* 'beds', *tīgnar* 'squirrels'.

The lack of the genitive and the locative are striking features of the case system. The old locative suffix has become a partitive object case, and its original function has been taken over by the dative. The dative suffix is *-GA*, but *kihi* 'man' and *kim* 'who' have *-ieχe*, i.e. *kihiēχe*, *kimieχe*. The accusative is formed with *-(n)I*. The accusative in *-LAR* is used as coordinative form. The partitive is an object case formed with the old locative suffix *-TA*. The initial consonant of the ablative suffix *-(t)An* is invariable, e.g. *oron* 'bed' → *oronton*, not **oronnon*. The instrumental is formed with *-(I)nAn*, the comparative with *-TĀγAr*, and the comitative with *-LĪn*, after kinship terms with *-nĀn*.

As for older and unproductive case forms, the old equative in *-čA* now expresses approximative numerals and pronominal adjectives, e.g. *uonča* 'about ten', *uon* 'ten', *oččo* 'such, one like that' ← *ol* 'that'. The old locative function of *-TA* is still observed in adjectives in *-TĀγI*, e.g. *küöl-lēyi* 'being in a lake', in local expressions such as *manna* 'here' ← *bu* 'this', and in the temporal suffix *-ina*, e.g. *bārbina* 'while I was (there), in my presence' ← *bār* 'existing', *suoχχuna* 'while you were not (there), in your absence' ← *suoχ* 'non-existing'.

Table 26.5 Case suffixes

Case	Suffix	Examples			
nom.	-∅	eye 'peace'	uot 'fire'	oχ 'arrow'	oχ-tor 'arrows'
acc.	-(n)I	eyeni	uotu	oγu	oχ-tor-u
dat.	-GA	eyeye	uokka	oχko	oχ-tor-go
partitive	-TA	eyete	uotta	oχto	oχ-tor-do
abl.	-(t)An	eyetten	uottan	oχton	oχ-tor-ton
instrumental	-(I)nAn	eyenen	uotunan	oγunan	oχ-tor-unan
comitative	-LĪn	eyelīn	uottūn	oχtūn	oχ-tor-dūn
comparative	-TĀγAr	eyetēyer	uottāyar	oχtōyor	oχ-tor-dōyor

The possessive suffixes are 1sg. *-(I)m*, 2sg. *-(I)ŋ*, 3sg. *-(t)A*, 1pl. *-BI*, 2pl. *-GI*, 3pl. *-LArA*, e.g. *ebe* 'grandmother' → *ebem*, *ebeŋ*, *ebete*, *ebebit*, *ebeyit*, *ebelere* 'my [etc.] grandmother', *küön* 'breast' → *küönüm*, *küönüŋ*, *küöne*, *küömmüt*, *küöŋgüt*, *küönnere* 'my [etc.] breast'. There is no difference between *attara* [horse:PL.:POSS3SG.] 'his/her horses' and *attara* [horse:POSS3PL.] 'their horse' or 'their horses'. A form **at-tar-dara* [horse:PL.:POSS3PL.] 'their horses' does not exist. Some Russian loanwords ending in two consonants receive an additional vowel before the third-person possessive suffix, e.g. *očerk* 'outline' → possessive *očerkata* 'its outline', not **očerga*.

Though possessive declensional suffixes consist of a possessive and a declensional suffix, the suffixes of oblique cases are not quite identical to those of the nonpossessive declension. The possessive suffixes followed by a case suffix are 1sg. *-BI*, dative *-BA*, 2sg. *-GI*, dative *-GA*, 3sg. *-(t)I*, 1pl. *-BI*, 2pl. *-GI*, 3pl. *-LArI*. The case suffixes in the possessive declension are dative 1sg. and 2sg. *-r*, 3sg. and all the plural forms *-gAr*, accusative *-n*, partitive *-nA*, ablative *-tAn*, instrumental *-nAn*, comitative *-nĭn/-nĀn*, comparative *-nĀγAr*. Examples: dative *at* 'horse' → *appar* < **atbar* 'to my horse', *akkar* < **atγar* 'to your horse', accusative *kömö* 'help, aid' → 3sg. *kömötün*, 1pl. *kömöbütün*, partitive *tabax* 'tobacco' → 2pl. *tabaxxıtına*, 3pl. *tabaxtarına*, ablative *üle* 'work' → 1sg. *ülebitten*, 2sg. *üleyitten*, instrumental *kūs* 'strength, force' → 3sg. *kühünen*, 1pl. *küspütünen*, comitative *töbö* 'head' → 2pl. *töböyütünün*, 3pl. *töbölorünün*.

Pronouns

The personal pronouns are *min* 'I', *en* 'you', *kini* 'he, she, it', *bihigi* 'we', *ehigi* 'you', *kiniler* 'they'. The pronouns *min* and *en* have the oblique stems *mīgi-* and *eīgi-*, e.g. accusative *migin*, *eigin*, *kinini*, ablative *mīgitten*, *eīgitten*, *bihigitten*. The dative suffix of all persons except 3pl. is *-iexe*, e.g. 3sg. *kiniexe*, 2pl. *ehiexe*, but 3pl. *kinilerge*. The dative stems of the first- and second-person singular are *m-* and *eŷ-*, i.e. 1sg. *miexe*, 2sg. *eŷiexe*. Stems of all non-third person pronouns receive an additional *n* in the comitative and comparative, 1sg. *mīginnĭn*, *mīginnēyer*, 2g. *eŷiginnĭn*, *eŷiginnēyer*, 1pl. *bihiginnĭn*, *bihiginnēyer*, 2pl. *ehiginnĭn*, *ehiginnēyer*; cf. 3sg. *kinilĭn*, *kinitēyer*, 3pl. *kinilerdĭn*, *kinilerdeyer*. The accusative suffix is *-n* in 1sg. and 2sg., *-ni* in 3sg., 1pl. and 2pl., *-i* in 3pl.

Collective pronouns are *bihikki* 'we [= you, he . . .] + me' and *ehikki* 'he, they . . . + you [sg.]', both declined as nouns, e.g. *en bihikkineyer* 'more than you [sg.] and me/us', *Uybān ehikkige* 'to Ivan and you'.

The demonstrative pronouns are *bu* 'this', plural *balar*, *iti* 'that', plural *itiler*, *ol* 'that [remote]', plural *olor*. Their emphatic forms are *subu* 'exactly this', plural *sabalar*, *siti* 'exactly that', plural *sitiler*, *sol* 'exactly that [remote]', plural *solor*. The singular forms have different oblique stems: *bu* → *man-*, *subu* → *saman-*, *iti* → *itin-*, *siti* → *sitin-*, *ol* → *on-*, *sol* → *son-*. All singular forms have a locative: *manna* 'here', *samanna* 'exactly here', *itinne*

'there', *sitinne* 'exactly there', *onno* 'there [remote]', *sonno* 'exactly there [remote]'. The dative suffix is *-IAXA* in the singular and *-GA* in the plural. The accusative suffix is *-I*, not **-n* or **-nl*.

Adjectival forms are based on the oblique stem + *-nlk*, i.e. *mannik* 'one like this' ← *bu*, *itinnik* 'one like that' ← *iti*, *onnuk* 'one like that [remote]' ← *ol*. Numeral derivations of pronouns are *bačča* 'as much as this' ← *bu*, *itečče* 'as much as that' ← *iti*, *oččo* 'as much as that [remote]' ← *ol*.

Interrogative pronouns and derivations from them are *kim* 'who', plural *kinneχ*, *tuox* 'what', plural *tuoxtar*, *χas* and *töhö* 'how much', *χaydayχ* 'how', *χaya* 'which', *χannik?* 'what kind of . . .?', *χanna* 'where, where to', *χantan* 'from where', *toyo* 'why', *χahan* 'when'. The accusative and instrumental of *tuox* have double forms, *tugultuoyu*, *tugunan/tuoyunan*. The dative is now *tuoxχa*, whereas the old monophthong form *toyo* has become a different word meaning 'why'.

Indefinite pronouns are formed with the particle *ere*, if specific entities are referred to, and with *eme* in other cases, e.g. *kim ere* 'somebody', *kim eme* 'somebody, anybody'. Reflexive pronouns are formed with *beye* + possessive suffixes, e.g. *beyem* 'myself', *beyete* 'himself', reciprocal pronouns with *beye* + *beye* with possessive and declensional suffixes, e.g. *beye-beyebitin* [each other:POSS1PL.:ACC.], *beye-beyeyitiger* [each other:POSS2PL.:DAT.].

Generalising pronouns and derivations from them are formed with interrogative pronouns + *bayarar*, e.g. *kim bayarar* 'whoever (s)he may be, every', *kimim bayarar* 'each of my (men, children, etc.)', *kinneχ bayarar* 'any group of people', *χantan bayarar* 'wherever it may be from, from anywhere', *χanna bayarar* 'everywhere'. Indefinite pronouns + the particles *da* and *dayani* are used in both affirmative and negative sentences, e.g. *kim da* or *kim dayani*, e.g. *Kimten da kini eŷigin taptür* [who:ABL. also (s)he you:ACC. love:PRES.3SG.] '(S)he loves you more than anyone else', *Kimi da kini taptābat* [who:ACC. also (s)he love:NEG.PRES.3SG.] '(S)he does not love anyone'.

The possessive pronouns, containing the possessive suffix *-e*, are *miene* 'mine', *eŷiene* 'yours', *kiniene* 'his, hers, its', *bihiene* 'ours', *ehiene* 'yours', *kinienere* 'theirs', with plural forms such as *miennere*, *eŷiennere*, e.g. *Iti mas bihiene* 'This tree is ours', *Iti mastar bihiennere* 'These trees are ours'.

Numerals

Cardinal numbers include *bir* 'one', *ikki* 'two', *üs* 'three', *tüört* 'four', *bies* 'five', *alta* 'six', *sette* 'seven', *ayis* 'eight', *toyus* 'nine', *uon* 'ten', *sūrbe* 'twenty', *otut* 'thirty', *sūs* 'hundred' and *tihĭnča* 'thousand'. The cardinals from eleven to nineteen are formed with *uon* 'ten' + a digit, e.g. *uon tüört* 'fourteen', *uon sette* 'seventeen'. The tens from 40 to 90 are formed with a digit + *uon* 'ten', e.g. *tüört uon* 'forty', *sette uon* 'seventy'. Multiplicative cardinals are formed with cardinal + *-TA*, e.g. *bırde* 'once', *ikkite* 'twice', *uonna* 'ten times'.

Ordinals are formed from cardinals with *-(I)s*. Low final stem vowels and high vowels of the stem syllable are dropped, e.g. *alta* 'six' → *altis* 'sixth', *ayis* 'eight' → *aχsis* 'eighth', and diphthongs become short low monophthongs, e.g. *bies* 'five' → *behis* 'fifth'. The ordinal form of *bīr* 'one' is used in compounds only, e.g. *sūrbe bīris* 'twenty-first'; otherwise words such as *mannaygi*, *bastaki* or *bastin* 'first' are used. The tens 20 to 90 have no ordinal forms, e.g. *sūrbe* (not **sūrbes*) *uruok* 'lesson twenty'. Multiplicative ordinals consist of ordinals + *-(I)n*, e.g. *ikkis* 'second' ← *ikkihin* 'the second time'.

Approximative numbers are expressed with *χas* 'how much?', e.g. *ūs χas* 'about three', or, with cardinals of tens, the suffix *-(č)čA(kA)*, e.g. *sūrbečče(ke)* 'about twenty'. Multiplicative approximatives are formed with *-TA*, e.g. *üste χasta* 'about three times', *sūsčete* 'about a hundred times'. Distributives are formed with *-Lī*, e.g. *uonnū (kinige)* 'ten (books) for each', approximative distributives with the approximative + *-Lī*, e.g. *üstū χastī (kinige)* 'about three (books) for each', *sūsčelī (kinige)* 'about a hundred (books) for each'.

Collective nominals are formed with *-IA*, e.g. *Bihigi onuobut* [we ten:COLL.:1PL.] 'we are ten persons', *Behie keličpit* [five:COLL. come:PROSP.1PL.] 'Five of us will come [together]', or with *-IAn* (only with the numerals 2 to 10), e.g. *Ikkien bu jieye olorollor* [two:COLL. this house:DAT. live:PRES.:3PL.] 'They both live in this house'. Higher collectives are formed with the converb *buolan* preceded by the cardinal, e.g. *Uon ikki buolan keličpit* 'Twelve of us will be coming [together]'. Limitatives are formed with *-IAyAχ*, e.g. *Manna behieyex kibi ülelir* [here five:LIM. person work:PRES.3SG.] 'Only five persons work here'. Collective limitatives add the suffix *-In*, e.g. *settieyeyin* 'only seven of (us ...)'. Multiplicative limitatives add *-TA*, e.g. *ikkieyexte* 'only twice'.

Fractional numbers consist of the denominator + *gimmüt* (literally 'done') + the numerator with a third-person singular possessive suffix, e.g. *ūs gimmüt bire* 'a third', *sir üs gimmüt ikkite* 'two thirds of the world', *Sirim bies gimmüt bire oyuruot buoluo* 'A fifth of my field will be garden'. The word *ağara* 'half' may be used instead of *ikkie gimmüt bire* 'a half'. With words designating paired body parts, e.g. *ilī* 'hands' (not **hand*), it refers to a single entity, e.g. *ağar ilī* 'one hand', *ağar ililēχ* 'having one hand'.

Verbs

Deverbal Verb Derivation

As for voice forms, passives are formed with *-n* and *-(I)ln*, the former added to vowel stems, with shortening of the vowel, e.g. *ahā* 'eat' → *ahan-*, and the latter to consonant stems, e.g. *tik* 'sew' → *tigilin-*. With monosyllabic stems in *-y*, the suffix *-lln* or the short form *-lln-* are possible, in the latter case the stem final *-y* is dropped, e.g. *bāy* 'bind' → *bāyilin-*, *bālin-*.

Polysyllabic stems in *-y* only take the short form *-lln-*, while the stem final *-y* is dropped, e.g. *χomuy* 'tidy up' → *χomulun-*. When followed by a vowel, *-(I)ln* becomes *-(I)ll-*, e.g. *Jieler tutullallar* 'Houses are built' ← *tutulun-* ← *tut-* 'build'. The *-n* forms, which may also have reflexive meaning, need *-lln* to express passive meaning unequivocally, e.g. *bihā* 'save' → *bihan-* ~ *bihanilin-* 'be saved'. The reflexive, also expressing 'do for oneself', is formed with *-(I)n*, with shortening of stem-final long vowels including diphthongs, e.g. *χolbō* 'unite' → *χolbon-* 'get united'.

Causatives are formed with three suffixes. Thus, *-t* is added to stems ending in vowels (which are shortened), e.g. *sanā* 'think' → *sanat-* 'make somebody think', *die* 'say' → *det-*. However, *ia* 'milk' and *sie* 'eat' have the causatives *iat-* and *siet-*. Consonant stems take *-TAr*, e.g. *tönün-* 'come back' → *tönünner-* 'make somebody come back', *tik* 'sew' → *tikter-* 'get sewed, get somebody to sew', and *-(I)Ar* in some forms, e.g. *tüs-* 'fall' → *tüher-* 'drop', *tur-* 'stand' → *туруor-* 'set'. One stem can bear two causative suffixes, e.g. *öl-* 'die' → *ölör-* 'kill' → *ölört-* 'have somebody killed'.

Cooperative-reciprocal stems are formed with *-(I)s*, after some consonant stems *-sIs*, e.g. *üelē* 'work' → *üleles-* 'work together', *bil-* 'know' → *bilis-* 'know each other', *kör-* 'see' → *körsüs-* 'see each other, meet'. Examples: *Moskvaya körsüstüler* [MOSCOW:DAT. see:REC.:PST.3PL.] 'They met in Moscow', *Kiniler miğin kitta barsallar* [they I:ACC. with go:REC.:PRES.:3PL.] 'They all go with me'. Singular forms mean 'help to do, participate in doing', e.g. *Kolχozka ülelēhebin* [kolkhoz:DAT. work:REC.:PRES.:1SG.] 'I work [participate in working] in a kolkhoz', *Kini bihieχe ot tieyiste* [(s)he we:DAT. hay bring:REC.:PST.] '(S)he helped us to bring hay home'.

There are also suffixes signalling actional modification. A multiplicative stem is formed with *-(I)tAlĀ*, e.g. *Čāskilari süytalā* [cup:PL.ACC. wash: MULT.IMP2SG.] 'Wash the cups one after another/every single cup!'; cf. *Čāskilari süy* [cup:PL.ACC. wash:IMP2SG.] 'Wash the cups!' Intensive forms in *-MAχtĀ* may signal accelerated actions, e.g. *ülelēmeχtē* 'work quickly' etc.

Finite Verb Forms

The verbal system is extremely rich. In the following, examples of the most important forms will be given. The present is formed with 1sg. *-A-bln*, negative *-BAp-pIn*, 2sg. *-A-GIn*, negative *-BAk-klIn*, 3sg. *-Ar*, negative *-BAt*, 1pl. *-A-blIt*, negative *-BAp-plIt*, 2pl. *-A-GIt*, negative *-BAk-klIt*, 3pl. *-Al-lAr*, negative *-BAt-tAr*. Stem-final vowels merge with the initial suffix vowel into a long high vowel *ī*, e.g. *ahā* 'eat' → *ahībin* 'I eat'. An imperfect is constructed with the intraterminal aorist participle + past forms of the copula, e.g. *Keler ete* [come:R.PART. COP.:PST.] '(S)he was coming', *Kepsir etigit* [tell: R.PART. COP.:PST.:2PL.] 'You were telling', or with the aorist participle + possessive suffixes, e.g. *Kelerim* [come:R.PART.:POSS1SG.] 'I was coming', *Kepsirgit* [tell: R.PART.:POSS2PL.] 'You (plural) were telling'.

The simple past takes the suffixes 1sg. *-TIm*, 2sg. *-TIŋ*, 3sg. *-TA*, 1pl.

-TibIt, 2pl. *-Tigt*, 3pl. *-tIAR*. A post-terminal past, a kind of perfect with inferential nuances, is expressed with *-BIt*, negative *-BAAtAχ*, plus possessive suffixes, e.g. 3sg. *kelbite*, negative *kelbeteye*, 2pl. *kelbikkIt*, negative *kelbetexxit* ← *kel-* 'come'. A resultative post-terminal is formed with *-BIt*, negative *-BAAtAχ*, plus pronominal personal markers, 3sg. *kelbit*, negative *kelbetex*, 2pl. *kelbikkIt*, negative *kelbetexxit*, e.g. *Kini ületin oqorbotoχ* [(S)he work:POSS3SG.ACC. DO:NEG.RES.] '(S)he did not do his/her work'. There is no formal difference between the two post-terminals in the first- and second-person plural. A pluperfect is formed with *-BIt* + past copula, e.g. *Kelbit ete* '(S)he had come'. There is also an 'episodic' perfect constructed with *-BIt* + *-TAX* + pronominal personal markers, e.g. 3sg. *Kelbittēχ* [COME:BIT.PART.:TAX] 'Once (s)he came', *Kepsēbetexxētēχxit* [tell:NEG.BIT.PART.:TAX:2PL.] 'Once you did not tell'. A corresponding pluperfect is formed with *-BIt* + *-TAX* + past copula, e.g. *Kelbittēχ ete* '(S)he had come once'.

A prospective (future) is expressed with 1sg. *-IAY-Im* ~ *-IAM*, 2sg. *-IAYIŋ* ~ *-IAŋ*, 3sg. *-IAY-A* ~ *-IA*, 1pl. *-IAX-plt*, 2pl. *-IAX-χIt*, 3pl. *-IAX-tArA*. Negative forms are based on the short form + *suoya*, e.g. *Min üleliem suoya* 'I shall not work'.

The imperative is formed with 1sg. *-Īm*, negative *-Im-Īm*, 2sg. *-θ*, negative *-ImA*, 3sg. *-TIn*, negative *-BATIn*, 1pl. inclusive *-IAX*, negative *-ImIAX*, 1pl. exclusive *-IAYIŋ*, negative *-ImIAYIŋ*, 2pl. *-Iŋ*, negative *-ImAŋ*, 3pl. *-TInnAr*, negative *-BATInnAr*. The future imperative ('do later') is used only in the second person: 2sg. affirmative *-Ar* ~ *-Arly* ~ *-Arly*, negative *-Im* + affirmative, 2pl. *-ArIŋ* ~ *-ArIŋItly* ~ *-ArIŋItly* ~ *-ArIŋItInly* ~ *-ArIŋItInly*, negative *-Im* + affirmative.

Conditionals are formed with the irreal condition suffix *-TAr*, negative *-BAAtAr*, followed by pronominal personal markers, or with the real condition suffix *-TAX*, negative *-BAAtAX*, followed by the personal markers 1sg. *-pInA*, 2sg. *-χInA*, 3sg. *-InA*, 1pl. *-pltnA*, 2pl. *-χIttnA*, 3pl. *-tArInA*. A presumptive is formed with *-Īhl*, negative *-(I)mĪhl*, + pronominal personal markers, e.g. *Kinitten iyitihibin* [(S)he:ABL. ask:PRESUMP.1SG.] 'I guess I should ask him/her'. A present necessitative consists of the aorist participle + *-LAX* + pronominal personal markers, e.g. *Eterdēχpin* [say:R.PART.:NEC.:1SG.] 'I have to/must/should say'. The past is formed with past forms of the copula, e.g. *Min učūtalbar kinige biererdēχ etim* [I teacher:POSS1SG.:DAT. book give:R.PART.NEC. COP.:PST.1SG.] 'I had to give a book to my teacher'. A prospective necessitative is formed with the prospective participle + *-LAX* + pronominal personal markers, e.g. *Ehigi kīspar plat'e tigiexxētēχxit* [you daughter:POSS1SG.:DAT. dress sew:PROSP.PART.:NEC.2PL.] 'You (plural) will have to sew a dress for my daughter'. A voluntative-potential is formed in the first- and second-person singular and plural with *-(Ā)yA*, negative *-(I)m-ĀyA*, + pronominal personal markers, e.g. *Barāyabīn* [go:VOL.:1SG.] 'I would like to go, I can/may go', in the third-person singular with *-(Ā)rAy*, negative *-(I)m-ĀrAy*, and in the third-person plural with *-(Ā)yAl-IAR*, negative *-(I)m-ĀyAlIAR*, e.g.

Ahāray [eat:VOL.3SG.] '(S)he would like to eat' or '(S)he can/may eat', *Ahāmāyallar* [eat:NEG.:VOL.3PL.] 'They would not like to eat' or 'They cannot/may not eat'.

Non-finite Verb Forms

The aorist participle is formed with *-Ar* after consonants and *-Īr* after vowels, negative *-BAr*, e.g. *ÜelĪr ahĪr, ülelēbet ahābat* [work:R.PART. eat:PRS. work:NEG.R.PART. eat:NEG.PRS.] 'The one who works eats, the one who does not work does not eat'. A post-terminal (perfect) participle is formed with *-BIt*, negative: *-BAAtAχ*, e.g. *Uybān suruybut suruga* [Ivan write:BIT.PART. letter:POSS3SG.] 'the letter that Ivan has written'. A prospective (future) participle is formed with *-IAX*, negative *-ImIAX*, e.g. *üörenieχ* 'learnable'. A habitual participle or actor noun is formed with *-ĀččI*, e.g. *manna ünēččĪ mastar* [here grow:HAB.PART. tree:PL.] 'trees that grow here', *manna sayilāččĪ oγolor* [here spend the summer:HAB.PART. child:PL.] 'children who spend every summer here', *inaχ iaččĪ* [cow milker] 'milker'. Finally, a participle expressing that an event has not yet taken place (participium nondum facti) is formed with the converb in *-A* + *ilik*, e.g. *üörene ilik oγo* [learn:A.CONV. ILIK child] 'a child who does not yet go to school'.

The converb in *-An* has a vague meaning similar to 'and', e.g. *Uol isten turar* [boy listen:CONV. stand:PRS.] 'The boy is standing and listening'. The *-A* converb, formed with *-A* after consonants and merging with stem-final vowels into *-Ī*, expresses an adverbial modification, e.g. *küle-küle et-* [laugh:CONV. laugh:CONV. say] 'say something laughing'. Both converbs are negated with *-(I)mInA*.

The converb in *-Bakka* means 'without doing', e.g. *tugu da sajarbakka* [what:ACC. also say:BAKKA.CONV.] 'without saying a word'. The one in *-ĀrI* means 'in order to', e.g. *KĪhĪn bulārĪ barda* [daughter:POSS3SG.ACC. find:ARI.CONV. go:PST.] '(S)he went to find his/her daughter'. *-Āt* indicates 'as soon as', e.g. *Min χospor kĪrēt ardaχ tüste* [I room:POSS1SG.DAT. enter:AT.CONV. rain fall:PST.] 'As soon as I came into my room, it began to rain'. *-BIččA* signals causality, e.g. *Itini kutammüčča ettim* [that:ACC. be frightened:BIččA.CONV. say:PST.1SG.] 'I said that because I was frightened'.

Converbs in *-A* or *-An* of lexical verbs combined with auxiliary verbs such as *tur-* 'stand', *olor-* 'sit', *sĪrĪt-* 'move' and *sĪt-* 'lie' denote various actional modifications, e.g. *SamĪr tūhe turar* 'It is continuously raining'. The fossilised verbal element *kulu* is used as a particle after the *-An* converb in imperatives, e.g. *Āyan kulu!* 'Read (for me)!'

Syntax

Nominal Phrases

The noun head of a nominal phrase may be preceded by a demonstrative pronoun or a quantifier such as *barĪ* 'all', *bütĪn* 'whole'. Note that *bĪr*, which

is essentially a numeral, occurs much less frequently than indefinite articles in most other Turkic languages. Adjectives, participles and relative clauses may be used attributively. Determiners and modifiers precede the head and do not agree with it in case and number. The quantifier *bari* may follow the head, e.g. *tabalar bari* [reindeer:PL. all] 'all reindeers'.

Relative clauses based on participles, most prominently in *-Bl* and *-Ar*, precede their head noun, whereas, unlike in Turkish, the first actant of the relative clause is marked on the head noun, e.g. *körbüt sirbit* [see:PART. place:POSS1PL.] 'the place that we see', *olorbut jiem* [live:PART. house:POSS1SG.] 'the house I live in'. Since Yakut lacks a genitive case, the subject of a relative clause is unmarked for case.

One type of nominal compounds consists of the juxtaposition of two nouns. The first noun mostly denotes a material, e.g. *mas jie* [tree house] 'log cabin' *timir suol* [iron way] 'railway track'. Another type consists of two nouns, the last one being marked with the possessive suffix, e.g. *kihi āta* [person name:POSS3SG.] 'personal name', *χaraχ āta* [eye water:POSS3SG.] 'tear'. Interestingly enough, if the first noun is a plural-marked or collective noun, the second one agrees with it even if it expresses a singular concept, e.g. *bu oyolor kinigelere* [this child:PL. book:PL.:POSS3SG.] 'a book/books of these children', *tabahittar künnere* [reindeer breeder:PL. day:PL.:POSS3SG.] 'the day of reindeer-breeders', *jon sanālara* [people idea:PL.POSS3SG.] 'national idea' (or *jon sanāta*).

A construction containing two nouns, of which the second one bears a possessive suffix, can function as a possessive nominal construction, e.g. *učūtal jiete* [teacher house:POSS3SG.] 'the house of the teacher'. A noun with a possessive suffix can be preceded by a personal pronoun, e.g. *min ebem* 'my grandmother', *ehigi akkit* 'your (plural) horse', but not by a possessive pronoun. In possessive constructions with more than two constituents, the first one is unmarked, whereas the following ones are each marked with *-(t)In* (a remnant of the genitive) and the last one bears the possessive suffix, e.g. *kini ayatın χarayın āta* [(s)he father:TIN eye:TIN water:POSS3SG.] 'his/her father's tears', literally 'the water of the eyes of his/her father'; cf. *kini ayata* 'his/her father'. If a complex possessive construction contains a compound, the first element of the compound is not marked with *-(t)In*, e.g. *ehem* 'my grandfather' + *balik ilime* 'fishing net' → *ehem balik ilime*. *Ehem baligın ilime* would mean 'a net of my grandfather's fish'. The first constituent may even be a third-person singular personal pronoun, e.g. *kini ehetin orono* [(s)he grandfather:TIN bed:POSS3SG.] 'his/her grandfather's bed'. The pronoun may be omitted, but the *-(t)In* marker on the second element remains, e.g. *ehetin orono* with the same meaning.

Case Syntax

The dative case can be assigned to indirect objects, directive and local complements, temporal expressions, etc., e.g. *oskuolaya bar-* 'go to school',

oskuolaya buol- 'be in the school', *Samīr ūs χonukka tüste* 'It rained for three days and three nights', *ūs χonukka* '[during] three days'. The accusative marks direct objects, complements expressing distance and duration, e.g. *Ūs kilometri bardibit* [three kilometre:ACC. walk:PST.:1SG.] 'We have walked three kilometres', *ikki künü* 'for two days'. The partitive object case indicates that just a part of an object is affected, e.g. *Kiliepte ayal* 'Get some bread!'. The corresponding expression *Kiliebi ayal* with the object in the accusative means 'Get the [whole loaf of] bread!' The partitive is only used in imperative or necessitative expressions, e.g. *Kiliepte ayaliāχχa nāda* [bread:ACC.PART. bring:PROSP.PART.:DAT. necessary] 'One has to bring some bread'. The coordinative accusative in *-LArI* marks a complementary direct object, e.g. *Kini atin siaryalari süterde* [(s)he horse:POSS3SG.ACC. sleigh:COORD.ACC. lose:PST.] '(S)he has lost his/her horse and (at the same time/together with it) the sleigh'. In this example, *atin* 'his/her horse' is the main object and *siaryalari* the complementary one. The ablative expresses a starting point, a source, a time interval, a cause, etc., e.g. *Ol kihitten iyü* [this man:ABL. ask] 'Ask this man!', *kiehetten* 'from the evening on' or 'in the evening', *ardaχtan* 'because of the rain'. The instrumental marks different types of complements, e.g. *avtobuhunan kel-* 'come by bus', *ūs nedielenen* 'after three weeks', *iārīnan* 'because of illness', *oyūrunan is-* 'go through a forest', *ostuorahinan ülelē-* 'work as a guard', *direktorinan tal-* 'elect somebody director', *mahinan oŋor-* 'produce of wood', *ereyinen* 'hardly'. The comitative marks complementary participants, e.g. *Saχa nūččalīn kelliler* [Yakut Russian:COM. come:PST.3PL.] 'A Yakut and a Russian came', *Ebem oyolordūn kelle* [grandmother:POSS1SG. child:PL.COM. come:PST.] 'My grandmother came with [some] children'. The difference is signalled by the predicate, which is in the plural in the former, but in the singular in the latter case. The comitative suffix may be added to each element of a collective, e.g. *turāχīn, kustūn, χāstīn* 'crows and ducks and geese'. The suffix *-TĀγAr* is used in comparative constructions, e.g. *Χaχay ehētēyer kūstēχ* [lion bear:COM. strong] 'A lion is stronger than a bear'. The ablative may replace the comparative, as in the riddle *Tuoχtan dayanī kuččuguy, tuoχtāγar dayanī kūstēχ bār ühü* 'There is something that is smaller than anything else [and] stronger than anything else' (= *silim* 'glue').

Postpositional phrases include constructions based on items expressing spatial relationships. In such phrases, the nominal is unmarked, whereas the postposition bears possessive case forms, e.g. *jie ihiger* [house inside:POSS3SG.DAT.] 'in the house'. Certain other postpositions derived from verbal stems, e.g. *ila* 'from, since', *taχsa* 'more than', *sirsa* 'along', require case marking on the nominal.

Verbal Phrases

The dependent of the predicate core of a verbal phrase may be a nominal or an adverbial, e.g. *kinige āχ-* 'read a book', *Manī kör* 'Look at this!', *elbēxtik is-* 'drink much'.

Periphrastic verbs combine nominals with auxiliary verbs such as *gîn-* 'do', *il-* 'take', *bar-* 'go' *tut-* 'hold', e.g. *külû gîn-* 'ridicule'. The nominal may appear in various cases, e.g. in the dative with motion verbs: *külûge bar-* 'become a laughing-stock', *tolkuyga tûs-* 'come to mind, start thinking'. Russian verbs may be copied in stereotyped forms and inserted in a Yakut frame by means of pro-verbs such as *gîn-*, e.g. *svoni gîn-* 'make a telephone call' ← *zvonit*, *bîrasti gîn-* 'apologise' ← *prostit*'.

The Sentence

Constituent Order

The order of constituents in a neutral declarative sentence is SOV. Topics are sentence-initial e.g. *Ûtû bihiġi inaxtan ilabit* [milk:ACC. we cow:ABL. take:PRES.1PL.] 'We get the milk from cows', *Ûtten ürüġ as barîta keler* [milk:ABL. white food all:POSS3SG. come:PRES.3SG.] 'From milk come all dairy products'.

Pronouns are much more frequently used in Yakut than in most other Turkic languages, e.g. Turkish.

Interpretation in Terms of Definiteness/Specificity

An indefinite direct object is unmarked, whereas a definite or specific direct object is accusative-marked, e.g. *Bu sirge beleġ iyâmiya âspattar* [this place:DAT. gift hang:IMIYA.CONV. pass-by:NEG.PRES.3PL.] 'They do not pass this place without leaving [literally 'hanging'] a gift', *Kere beleġi iyâtîm* [nice gift:ACC. hang:PST.1SG.] 'I left a nice gift'. However, indefinite direct objects may also be marked with the accusative, which may be due to the influence of a contact language, e.g. *Bu fermaya kîrsani üösketeller* [this farm:DAT. polar fox:ACC. breed:PRES.3PL.] 'On this farm they breed polar foxes'.

Agreement Rules

Nouns within a possessive construction agree in number. Predicates agree in number with their subjects. Unlike most other Turkic languages, if the subject is a formal singular denoting a collective, e.g. *jon* 'people', its predicate is often plural-marked.

Expressions of 'Being', 'Having' etc.

'Be' in the present tense is expressed with the copula suffixes 1sg. *-Bln*, 2sg. *-Gln*, 3sg. *-Ø*. 1pl. *-LAR-blt*, 2pl. *-LAR-glt*, 3pl. *-Ø*, e.g. *Saxabîn* 'I am a Yakut', *Ergin* 'You are a man', *Kini uol* 'He is a boy', *Saxalarbit* 'We are Yakuts', *Erettergit* 'You are men', *Kiniler uolattar* 'They are boys'. The past copula particle is based on the simple past of the defective verb *e-* 'be', *etim*, *etiġ*, *ete*, *etibit*, *etigit*, *etiler* 'was (etc.)', e.g. *Min uçûtal etim* 'I was a teacher'. A second form of this verb is the indirective particle *ebit* ~ *ibit* (see p. 45).

'Have' is expressed with the adjective in *-LĀx*, e.g. *Min jielëxpîn* 'I have a house'. There is also a less used possessive construction with *bâr* 'existing',

in which the possessor is in the dative, while the possessed is unmarked, e.g. *Mieġe jie bâr* [I:DAT. house existing] 'I have a house'. If the possessed is plural-marked, the predicate agrees with it, e.g. *Jiege xostor bâllar* [house:DAT. room:PL. existing:PL.] 'The house has rooms'. Negative constructions are based on *suoġ* 'not existing', e.g. *Silġiga muos suoġ* [horse:DAT. horn not existing] 'Horses have no horns'. It is also possible to use a possessive pronoun predicatively, e.g. *Bu at miene* 'This horse is mine'.

Non-declarative Sentence Types

The question particle in yes/no questions is *duo* or *dû*. Unlike other Turkic languages, Yakut has an enclitic question particle *-ly* that combines with interrogative pronouns or adverbs and is usually attached to the predicate, e.g. *Bu kimiy?* [this who:Q.] 'Who is this?', *Bu kilaska xas tûnnûk bariy?* [this class:DAT. how many window existing:Q.] 'How many windows does this classroom have?'

The Syntax of Negation

Verb stems are negated with *-BA*. Constituents are negated with a postposed *suoġ* 'not existing', e.g. *üçüġey suoġ* 'not good, bad', *pioner suoġ oyolor* 'children who are not pioneer members'. Non-derived adjectives are negated with a third-person singular possessive suffix + *suoġ*, e.g. *kîrjayaha suoġ kîhi* 'a man who is not old' ← *kîrjayas kîhi* 'an old man', *En xorguha suoġxun* 'You are not a coward' ← *En xorguskun* 'You are a coward'. Adjectives derived with *-LĀx* are negated with the same construction, e.g. *sâta suoġ* [gun:POSS3SG. non-existing] 'unarmed' ← *sâlâx* 'armed'.

Types of Non-finite Clauses

Predicates of non-finite clauses are participles or converbs.

Complement clauses are usually based on the participles in *-Ar* and *-IAġ* + case markers. *Vanya utuya sîtarîn kördüm* [Vanya sleep:CONV lie:PART. POSS3SG.ACC see:PST.1SG.] 'I saw that Vanya slept'.

Various converbs form the basis for semantically different adverbial clause types. Yakut has a set of primary converbs, e.g. the ones in *-A*, *-An*, *-At*, and secondary forms constructed with participles + case markers. Purpose clauses are formed with the converb in *-Ārl*, causal clauses with the converb in *-BİċĈA*.

Irreal conditional clauses are based on the irreal conditional in *-TAr*. The main clause contains the prospective marker *-IAġ* + past copula, e.g. *Samîr tüspeter kîneye barîax etim* [rain fall:NEG.COND. cinema:DAT. go:PROSP. COP:PST.1SG.] 'I would go to the cinema if it were not raining'. Real conditions can be expressed with the real conditional in *-TAX*, e.g. *Ikkien sietexpitine, bîrâtîm açċik oloruo* [two:COLL. eat:COND.1PL. brother:POSS1SG. hungry sit:PROSP.] 'If we both eat up (this food), my brother will sit (stay) hungry'.

Coordination

Uonna 'and' is used to coordinate both clauses and constituents of phrases, e.g. *kinige uonna ɣarandās* 'book and pencil', *Jie sūyāčči sīhī ɣomuyar uonna bīlī sotor* [house cleaner dirt:ACC. clean:PRES.3SG. and dust:ACC. wipe: PRES.3SG.] 'The cleaner cleans the dirt and wipes the dust'. Two nouns can be coordinated by means of *ikki* 'two', which either follows the juxtaposed constituents or follows each of them, e.g. *en min ikki* 'you and I', *it ikki kuoska ikki* 'dog and cat'. For coordination with the comitative, see p. 429. The *-An* converb can also be used to coordinate two or more clauses, in which case the first actants (subjects) may be different, e.g. *Kelikke tiāhin isten, jietten Ōrūne bejeyen tayīsta*. [gate noise:POSS3SG.ACC. hear:CONV. house:ABL. Irina shuffle:CONV. go-out:PST.] 'When (they) heard the noise from the gate, Irina came shuffling out of the house', *Sayin āhan kūhūn kelbite* [summer go-past:CONV. autumn come:POSTT.PST] 'Summer went and autumn came'.

Diathetic Relations

The subject of a passive verb is in the nominative, whereas the agent, if expressed, is in the instrumental. Causatives of intransitive stems become transitives, e.g. *köt-* 'fly' → *kötör-* 'let fly'. The subject of a causative-marked transitive verb stem is in the nominative, the causee in the dative, and the direct object in the accusative, e.g. *Min manī köröbün*. [I this:ACC. see: PRES.1SG.] 'I see this', *Min manī ehieɣe kördöröbün* [I this:ACC. you:DAT. see:CAUS.:PRES.1SG.] 'I let you see (show you) this'.

Lexicon

The lexicon consists of native Turkic words as well as Mongolic, Russian and Tungusic loans. The basic stock, personal pronouns, numerals, designations of body-parts, etc., is of native origin. Most words of foreign origin are Mongolic borrowings. The older ones have undergone the same sound changes as Turkic words in Yakut, e.g. *suruk* 'letter' < *jiruy*. Russian is a much later source of loanwords. Some of them are fully assimilated, e.g. *učūtal* 'teacher' < *učitel'*, while others, mostly belonging to the domain of modern life and science, are at least spelled as in Russian. There are many Tungusic loanwords belonging to the domain of husbandry and everyday life, particularly in dialects. A considerable portion of the Yakut lexicon is of unknown origin, probably due to contact with Palaeosiberian languages.

Dialects

Yakut dialects are relatively weakly differentiated, so that speakers from quite distant territories can easily understand each other. There is a central group consisting of the Aldan, Eastern Lena and Western Lena dialects, and a peripheral group including the northeastern and the northwestern dialect. The latter are influenced by Tungusic, the northeastern one by Even (Lamut) and

the northwestern one by Evenki. The peripheral dialects often have stops instead of fricatives, e.g. *saka* 'Yakut' (cf. *saxa*), avoid regressive assimilation, e.g. *orto* 'middle' (not *otto*), have word accent on the first or second syllable and use the plural suffix *-AttAr*, e.g. *ialattar* < *ial* 'neighbour', *küöletter* ← *küöl* 'lake', and tend to lexicalise third-person singular possessive forms, e.g. *tumsa* 'beak, bill' (cf. *tumus*).

References and Further Reading

- Böhtlingk, O. (1851) *Über die Sprache der Jakuten*, St Petersburg. Reprinted in Indiana University Publications, Uralic and Altaic Series 35, The Hague: Mouton.
- Buder, A. (1989) *Aspekto-temporale Kategorien im Jakutischen*, Turcologica 9, Wiesbaden: Harrassowitz.
- Katużyński, S. (1961) *Mongolische Elemente in der jakutischen Sprache*, Warsaw: Państwowe Wydawnictwo Naukowe.
- Krueger, J. R. (1962) *Yakut Manual*, Indiana University Publications, Uralic and Altaic Studies, 21, The Hague: Mouton.
- Pekarskij, E. K. (1907–1930) *Slovar' jakutskogo jazyka*, St Petersburg, Leningrad: Akademija Nauk.
- Poppe, N. (1959) 'Das Jakutische', in J. Deny *et al.* (eds) *Philologiae Turcicae Fundamenta* 1, Wiesbaden: Steiner, 671–84.
- Stachowski, M. (1993) *Dolganischer Wortschatz*, Zeszyty Naukowe Uniwersytetu Jagiellońskiego MLXXXVI, Prace Językoznawcze, Zeszyt 114, Cracow: Uniwersytet Jagielloński.
- Ubrjatova, E. I. (ed.) (1982) *Grammatika sovremennogo jakutskogo literaturnogo jazyka*, Moscow: Nauka.

27 Chuvash

Larry Clark

Introduction

Chuvash is the national language of the Chuvash people who live in the Volga river region. The language is related most closely to Volga Bulghar and to other traces of Oghur languages. Its development from Proto-Turkic is complicated by a series of sound changes and replacements. Although Chuvash has retained most of the common Turkic morphosyntax and lexicon, these changes and especially those associated with the assimilation of Mari (Cheremis) tend to obscure its Turkic character.

Chuvash structure differs from other Turkic languages in several ways. For example, phonological oppositions are reduced in number as a result of the Oghur developments $-l < *-l, *-š,$ and $-r < *-r, *-z *-d$ (except $-y-r < *-d-r$). The devoicing of initial and final stops has neutralised the oppositions in these positions. As for the order of suffixation, the plural *-sem* follows possessive suffixes and precedes case markers, whereas in other Turkic languages the plural *-LAR* precedes possessive and case suffixes, e.g. *χērēmsenčēn = χēr+ēm+sem+ren* vs. Turkish *kız+lar+um+dan* 'from my daughters'. The personal pronouns *epĕ* 'I', *esĕ* 'you', *epir* 'we' and *esir* 'you' contain a deictic element *e-* ($< *a-$) and preserve the original Turkic stems **bi* and **si* (note Turkish *ben, sen,* but *biz, siz*). With second- and third-person imperatives, Chuvash uses the preposed negative particle *an*, e.g. *An pĭr!* 'Do not go!', whereas other Turkic languages use *-mA*, e.g. *Barma!* 'Do not go!'

Sound System

Phonetics and Phonology

The vowel system of Standard Chuvash consists of eight vowel phonemes (Table 27.1), plus /o/ in Russian words. The consonant system of Standard Chuvash consists of fourteen consonant phonemes (Table 27.2), plus /b/, /d/, /g/, /f/, /z/, /ʒ/, /tʃ/ in Russian words. In opposition to the tense vowels /a/ and /e/, the vowels /ä/ and /ĕ/ may be termed lax in the sense that the lips and tongue are completely relaxed in their articulation. Phonetically, /ä/ is a back-to-central vowel, and /ĕ/ a front-to-central vowel, and both are mid-to-low

Table 27.1 Vowels

	Front	Central	Back
high	i, ü	ĩ	u
mid	e	ĕ	ă
low			a

Table 27.2 Consonants

	Labial	Dental	Alveolar	Palatal	Velar
stop	p	t		č	k
fricative	v	s	ʃ	š	χ
nasal	m	n			
liquid		l	r		
glide				y	

(like *schwa*). The vowel /i/ is pronounced more centrally and with longer duration than Turkish /i/. The consonant /š/ is a voiceless dento-palatal fricative pronounced with the centre of the tongue raised towards the palate. In the standard language and Anatri dialect, /č/ is not an affricate like [tʃ], but a voiceless palatal stop pronounced with the tip of the tongue raised to the palate, thus, as an apical or retroflex *t* sound. Long consonants may occur in the stem of words, e.g. *allă* 'fifty', at the juncture of a stem with a suffix, e.g. *suxal-lă* 'bearded', and in certain suffixes, e.g. necessitative *-mAllA*. For lengthening of consonants, see below.

Two allophonic rules affect consonants. First, within a word and at the juncture between words, the voiceless consonants (except χ in most dialects) are pronounced lenis between vowels or after a sonorant, *m, n, l, r, y,* and before a vowel. This is not reflected in the writing system, e.g. *әпир* [eβir] 'we', *алпа* [alpa] 'by hand', *Мана пар!* [manapar] 'Give it to me!' Second, consonants are palatalised before and after front vowels (*пир* [p'ir] 'canvas'), but not in the environment of back vowels.

The vowels, except *i*, and most consonants may occur in any position of the word. Syllable types comprise: vowel, e.g. *a-lă* 'hand', consonant + vowel, e.g. *sü-re-* 'walk', vowel + consonant, e.g. *av-tan* 'rooster', consonant + vowel + consonant, e.g. *kil-* 'come', vowel + consonant + consonant, e.g. *alt-* 'dig', and consonant + vowel + consonant + consonant, e.g. *virt-* 'lie down'. In sequences of vowel + consonant + consonant, the first consonant is a sonorant *l* or *r*, and the second consonant is *t*.

Pronounced in isolation, the word receives accent on the final syllable, including suffixes in the standard language, e.g. *χulá* 'city', *χulaná* 'city'

(dative-accusative). If the final syllable of a word contains a reduced vowel *ă* or *ě*, accent falls on the preceding syllable that has a full vowel, e.g. *úrăχ* 'other'. If a word contains only these reduced vowels, accent falls on the first syllable, e.g. *χěrēměr* 'our daughter'. In compound words, a lighter accent falls on the last syllable of the first component and a stronger one on the last syllable containing a full vowel of the second component, e.g. *tāvútkětésłěχ* 'square'.

Morphophonology

Words are subject to front vs. back harmony, which requires that the vowels of a word, e.g. *timěr* 'iron', *laša* 'horse', and its suffixes, e.g. *timěrše* 'blacksmith', *lašapa* 'by horse', belong either to the front (*e, ě, i, ü*) or to the back (*a, ă, ĭ, u*) series of vowels. For palatalisation of consonants in front syllables, see above. Exceptions to the front-back harmony rule in Standard Chuvash include plural *-sem*, e.g. *ivălsem* 'sons'; third-person possessive *-ĕ/i*, e.g. *uri* 'his, her, its foot', attributive *-i*, e.g. *šurri* 'the white one', ordinal numeral *-měš*, e.g. *tāvattăměš* 'fourth', third-person singular personal marker *-ĕ*, e.g. *kayĕ* 'will go', *kayrĕ* 'went', adverb formants *-čen* (*-ččen*), *-len*, *-seren*, and kinship terms such as *atte* 'father', *anne* 'mother'.

With the addition of suffixes or in the juncture between words, the final vowel of a stem may change or disappear. Nouns ending in *-u* and *-ü* change these vowels to *-äv* and *-ĕv* with the possessive and some case suffixes, e.g. *pĕvĕ* 'her height' (*← pü*), *săvăn* 'the oil's' (*← šu*), *tăva* 'mountain' (dative-accusative) (*← tu*), while verbs ending in *-u* and *-ü* change these vowels to *-äv* and *-ĕv* with the present tense in *-(θ)At* and the participle in *-(θ)AkAn*, e.g. *săvat* 'he milks', *săvakan ěne* 'milk cow' (*← su-/säv-*). The final vowel of a verb stem drops with the future in *-(θ)Ă/-(θ)ĕ*, the present in *-(θ)At*, the subjunctive in *-(θ)Ăn*, etc., e.g. *Vulăp* 'I will read', *Vulatăm* 'I read', *Vulătăm* 'I would read' (*← vula-*). The final *-a* or *-e* of nouns and other words drops with the third-person possessive suffix *-ĕ/i*, e.g. *kĕneki* 'his/her book' (*← kĕneke*), *šaltti* 'inner' (*← šalta*).

Addition of tense, aspect, mood and person suffixes to verb stems produces several changes. The stems *kĕr-* 'enter', *kūr-* 'bring', *par-* 'give', *per-* 'throw, shoot', *pĭr-* 'go', *tăr-* 'stand', *šăr-* 'urinate', *χur-* 'place, put', *yar-* 'send', *yĕr-* 'weep' lose their final *r* when followed by the simple past in *-R*, the participle in *-nĂ* and the negative in *-mA*, e.g. *Kĕrĕ* '(S)he entered', *Kĕně* '(S)he has entered', *Kĕmerĕ* '(S)he did not enter', *Kĕmest* '(S)he does not enter'.

Final *-t/-t'* ~ *-d/-d'* of nouns becomes *-č* with the third-person possessive *-ĕ/i*, e.g. *yačĕ* 'his/her name' (*← yat*), *Čăvaš pečăčĕ* 'Chuvash press' (печать), *škul tetračĕ* 'school notebook' (тетрадь). The *-t* variant of the simple past and subjunctive suffixes of verbs ending in *-n*, *-l*, *-r* changes to *-č* with the third-person personal markers, e.g. *Purăncĕ* '(S)he lived',

Purăncĕĕ '(S)he would live', *Purăncĕš* 'They lived', *Purăncĕĕš* 'They would live'.

Consonant lengthening is a characteristic feature. It may occur in predicatively used adjectives, e.g. *Ku šĭn pĕčĕĕkĕ* 'This man is small' (cf. *pĕčĕk šĭn* 'small man') and in numerals for emphasis when they are used alone. Consonants also undergo lengthening with the addition of certain suffixes. This occurs when genitive, dative-accusative or possessive suffixes are added to a two-syllable word ending in *-ă/-ĕ*, e.g. *văyyăn*, *văyya*, *văyyăm* *← văyă* 'game', but not in words with more than two syllables or if the final vowel is preceded by two consonants, e.g. *kĕtüsĕn*, *kĕtüse* *← kĕtüsĕ* 'herdsman'. Lengthening also occurs when *-i* is added to an adjective or participle ending in a single consonant or consonant + *-ă/-ĕ*, e.g. *pĭsăkki* 'the big one' *← pĭsăk* 'big', *šĭrakanni* 'the one who writes' *← šĭrakan* 'who writes'. This is also the case when the adverb-forming suffix *-Ăn* is added to an adjective ending in a consonant or *-ă/-ĕ*, e.g. *ăššăn* 'warmly' *← ăšă* 'warm'.

Orthography

The alphabet is based on Cyrillic, but includes several letters for distinctively Chuvash sounds. Its initial formulation was due to Ivan Jakovlev (1848–1930), whose school primers of 1871–1873 systematised efforts to write Chuvash which had begun in the early eighteenth century. A final reform of the alphabet occurred in 1938. This alphabet indicates all phonemes of the sound system, but ignores the allophonic rules of lenis (weak) pronunciation and palatalisation. It also incorporates letters from the Russian alphabet which are irrelevant to Chuvash: *ю = yu*, *я = ya*, *е = ye* initially or after a vowel, but *e* otherwise.

Morphology

Nouns

Nouns may be simple words, e.g. *šĭv* 'water', *puš* 'head', words derived by lexical suffixes from other nouns, adjectives or verbs, e.g. *pulăšă* 'fisherman' *← pulă* 'fish', *văramăš* 'length' *← vărăm* 'long', *vĕreneken* 'learner', *vĕrenü* 'learning' *← vĕren-* 'learn', compounds, e.g. *šĕršĭv* 'country' ('land-water'), *yĭvăs-kurăk* 'plants' ('tree-grass'), or reduplications, e.g. *ača-pača* 'children' *← ača* 'child'.

Nouns are declined by adding plural, case and possessive (first-, second- and third-person, singular/plural) suffixes, attached to primary or derived stems in the order possessive + number + case, e.g. *kuš+lăχ+ămăr+sem+pe* [eye:DER:|PL.POSS:PL.INSTR] 'with our (pairs of) glasses'.

There are five different possessive suffixes (Table 27.3). Possessors in the third-person singular and plural are marked by the same suffix. To distinguish 'their X' from 'her, his, its X', possessive pronouns are used, e.g. *un(ăn) ači*

Table 27.3 Possessive suffixes

	Singular	Plural
1p.	-äm/-ëm (after consonants) -m (after vowels)	-ämär/-ëmër (after consonants) -mär/-mër (after vowels)
2p.	-u/-ü	-är/-ër (final vowel drops)
3p.		-ë (after consonants) -i (after vowels and some consonants)

'his/her child', *vësen ači* 'their child'. The third-person possessive *-ë/i* disappears before the dative-accusative case variant *-nA*, e.g. *yultašne* 'to his/her comrade' < **yultašëne*. In kinship terms for older relatives and some other words ending in a vowel and whose oblique stem is not *-Äv* (which takes *-ë*), *-U* is used as a second-person singular possessive suffix and *-Äšë* (variants *-šë/-š*) as a third-person possessive, e.g. *ašu* 'your father', *aššë* 'his, her, its, their father', *annü* 'your mother', *amäšë* 'his, her, its, their mother', *ëne pärušë* 'the cow's calf'.

Nouns may be marked for plural by *-sem* before the nominative and instrumental case suffixes, and by *-sen* before other case suffixes (Table 27.4). Five case relationships are indicated through the attachment of suffixes: possession by the genitive suffix, goal by the dative-accusative suffix, location by the locative suffix, source by the ablative suffix, and instrument by the instrumental suffix (Table 27.5). The nominative is suffixless. Declension with case endings applies to nouns, pronouns, numerals, the present and future participles, and word forms ending in *-i*, *-ë*, *-çi* and *-sker*.

Adjectives

Adjectives may be simple word forms, e.g. *sivë* 'cold', *çura* 'black', derivations from nouns, adjectives, adverbs and verbs, e.g. *väylä* 'strong' ← *väy* 'strength', *tävarsär* 'unsalted' ← *tävar* 'salt', *kilti* '(which is) in the home' ← *kil* 'home', *çuri* 'the black one' ← *çura* 'black', *çäräk* 'dry' ← *çär-*

Table 27.4 Case suffixes following plural forms

Cases	<i>ača</i> 'child'	<i>çër</i> 'daughter'
nom.	<i>açasem</i> 'children'	<i>çërsem</i> 'daughters'
gen.	<i>açasen(ën)</i> 'the children's'	<i>çërsen(ën)</i> 'daughters'
dat.-acc.	<i>açasene</i> '(to) the children'	<i>çërsene</i> '(to) the daughters'
loc.	<i>açasenče</i> 'on the children'	<i>çërsenče</i> 'on the daughters'
abl.	<i>açasenčen</i> 'from the children'	<i>çërsenčen</i> 'from the daughters'
instr.	<i>açasempe</i> 'with the children'	<i>çërsempe</i> 'with the daughters'

Table 27.5 Case suffixes

Word-final	Gen.	Dat.-acc.	Loc.	Abl.	Instr.
-a/-e	-nän/-nën ~ -n	-na/-ne			
-Cä/-Cë	-CCän/-CCën	-CCa/-CCe			
-i	-yën ~ -n	-ye			
-Cu/-Cü	-Cävän/-Cëvën	-Cäva/-Cëve			
-C	-än/-ën	-a/-e	-ra/-re -ta/-te	-ran/-ren -tan/ten	
-n, -l, -r					
2p. poss. -u/-ü		-na/-ne			
3p. poss. -ë/-i		-na/-ne	-inče	-inčen	-ëpe/-ipe
Plural -sem	-sen(ën)	-sene	-senče	-senčen	-sempe

C = consonant, CC = sequence of two consonants.

'dry', compounds, e.g. *šëvër sämsallä* 'sharp-nosed', or intensives formed by reduplication, e.g. *çëp-çërlë* 'bright red'.

Which of two entities possesses a greater amount of a quality may be indicated by the compared nominal + the standard of comparison in the ablative case + the adjective denoting the quality, e.g. *İltän këmëlten çaklä* 'Gold is more valuable than silver'. The adjective also may carry *-rAç* ~ *-tArAç*, e.g. *Payançi şumär ënerçinčen te väyläraç* 'Today's rain is (some-what) stronger than yesterday's'. The intensifying particles *çi*, *çat*, *vër* or *šëm* may be placed before an adjective to indicate the highest degree of a quality, e.g. *çi pisäk* 'biggest', *vër šënë* 'brand new'.

Pronouns

Pronouns include personal, demonstrative, reflexive, possessive, interrogative, collective, indefinite and negative pronouns. The element *e-* in the nominative of the first- and second-person pronouns goes back to a deictic element, e.g. *e-pë* '(here is/this one is) I' (Table 27.6).

The genitive forms are used as possessive pronouns, e.g. *man(än)* 'my', *pirën* 'our'. Addition of *-i* to either clipped or full forms of a possessive pronoun creates a noun meaning 'the one which belongs to me, you, etc.,

Table 27.6 Personal pronouns

	Singular			Plural		
	'I'	'you'	'she, he, it'	'we'	'you'	'they'
nom.	<i>epë</i>	<i>esë</i>	<i>väl</i>	<i>epir</i>	<i>esir</i>	<i>vësem</i>
gen.	<i>manän</i>	<i>sanän</i>	<i>unän</i>	<i>pirën</i>	<i>sirën</i>	<i>vësen(ën)</i>
dat.-acc.	<i>mana</i>	<i>sana</i>	<i>äna</i>	<i>pire</i>	<i>sire</i>	<i>vësene</i>
loc.	<i>manta</i>	<i>santa</i>	<i>unta</i>	<i>pirte</i>	<i>sirte</i>	<i>vësenče</i>
abl.	<i>mantan</i>	<i>santan</i>	<i>untan</i>	<i>pirten</i>	<i>sirten</i>	<i>vësenčen</i>
instr.	<i>manpa</i>	<i>sanpa</i>	<i>unpa</i>	<i>pirënpe</i>	<i>sirënpe</i>	<i>vësempe</i>

Table 27.7 Demonstrative pronouns

	Singular		Plural	
	'this'		'these'	
nom.	ku	śakă	kusem	śakssem
gen.	kunăn	śakăn	kusen	śaksen
dat.-acc.	kuna	śak(ă)na	kusene	śaksene
loc.	kunta	śakănta	kusenče	śaksenče
abl.	kuntan	śakăntan	kusenčen	śaksenčen
instr.	kunpa	śakănpa	kusempe	śakssempe
	'that'		'those'	
nom.	śavă	lešě	śavsem	lešsem
gen.	śavăn	lešěn	śavsen	lešsen
dat.-acc.	śavna	lešne	śavsene	lešsene
loc.	śavănta	lešěnče	śavsenče	lešsenče
abl.	śavăntan	lešěnčen	śavsenčen	lešsenčen
instr.	śavănpa	lešênpe	śavsempe	lešsempe

mine, yours, etc.', e.g. *manni* ~ *manănni* 'mine', *sanni* ~ *sanănni* 'yours'.

Demonstrative pronouns express three degrees of proximity: *ku* 'this' (visible and close), *śak(ă)* 'this (there), that' (visible and farther away), and *śav(ă)* 'that (there), that (spoken about)' (visible or invisible, but still farther away); see Table 27.7. The pronoun *lešě* 'that' also represents a more remote degree. The reflexive pronouns are coreferential with the subject of a sentence, e.g. *Xam kuśpa xam kurtăm* 'I (myself) saw with my own eyes' (Table 27.8). The common interrogative pronouns (and combinations with them) are *kam* 'who', *měn* and *měsker* 'what', *xaś(ě)* 'which', *miše* 'how much (in number)', *měn čuxlě* and *měn čul* 'how many (in quantity)', *měnlē* and *mělle yeple* 'what kind of', *xaśan* 'when', *ăšta* 'where', *měnsěn* 'why'.

Collective pronouns are universal quantifiers, e.g. *pur(te)* 'all, every', *purě* 'in all', *purěš*, *měnpur* 'all (without exception)'. Indefinite pronouns include *takam* 'someone', *tem(ěn)* ~ *teměsker* 'something', *taxśan* 'sometime' and *tašta* 'somewhere'. Negative pronouns are, e.g. *nikam* 'no one', *nim(ěn)* ~ *nimsker* 'nothing', *nixaśě* 'not one, any', *nixaśan* 'never', *ništa* 'nowhere', *nim(en)le* 'none'.

Numerals

Numerals may be divided into cardinal, ordinal, distributive, collective and approximative forms.

Cardinal numerals from one to nine and the decade ten each have three forms (Table 27.9). The basic form with a single consonant is used attributively when counting from one to ten items, e.g. *ikě śin* 'two men', and in the thousands, e.g. *ikě pin* 'two thousand'. The form used in isolation and in declension is an emphatic one with a lengthened consonant, e.g. *tăxxăr* 'a

Table 27.8 Reflexive pronouns

	Singular			Plural		
nom.	'I myself'	'you yourself'	'he himself', 'she herself', 'it itself'	'we ourselves'	'you yourselves'	'they themselves'
gen.	xam	xăv	xăy	xămăr	xăvăr	xăysem
dat.-acc.	xămăn	xăvăn	xăyăn	xămărăn	xăvărăn	xăysem(ăn)
loc.	xama	xăvna	xăyne	xămăra	xăvăra	xăysene
abl.	xamra	xăvănta	xăyănče	xămărta	xăvărta	xăysenče
instr.	xamran	xăvăntan	xăyănčen	xămărtan	xăvărta	xăysenčen
	xampa	xăvănpa	xăyănpe	xămărpa	xăvărpa	xăysempe

Table 27.9 Numerals

	Basic	Emphatic	Clipped		Basic	Emphatic
1	pěre	pěrre	pěr	20	śirēm	
2	ikkē	ikkē	ik	30	vātār	
3	višē	viššē	viś	40	χērēχ	
4	tāvata	tāvattā	tāvāt	50	alā	allā
5	pilēk	pillēk	pilēk	60	utmāl	
6	ultā	ulttā	ult	70	śitmēl	
7	śičē	śiččē	śič	80	sakārvunā	sakārvunnā
8	sakār	sakkār	sakār	90	tāχārvunā	tāχārvunnā
9	tāχār	tāχχār	tāχār	100	śēr	
10	vunā	vunnā	vun	1000	pin	

(unit of) nine', *allā ikkē* 'fifty-two', *ikk-žen*, *ikk-ē*. The third form is a clipped form with a single consonant and no final vowel used with 1 (*pěr*) to 10 (*vun*) for counting between decades and in the hundreds, e.g. *vunpěr* 'eleven', *ultśēr sakār vun ulttā* 'six hundred and eighty-six'.

Ordinal numerals are formed by attaching *-mēš* (dialects have *-māš*) to the emphatic forms of cardinal numerals 1 to 10 and to the basic forms of decade numerals 20 to 90 and higher numbers: *pērremēš* 'first', *ikkēmēš* 'second', *śiččēmēš* 'seventh', *vunnāmēš* 'tenth', *śirēmēmēš* 'twentieth', *śitmēlmēš* 'seventieth', *śērmēš* 'hundredth', *pinmēš* 'thousandth', etc.

Distributive numerals with the meaning '(number) each, in (number)s' are formed from the clipped forms of numerals with *-šAr*, e.g. *ikkēr* 'two each, in twos'. When numbers are distributed in the hundreds or thousands, this suffix is added to the first numeral, e.g. *tāvātšar śēr* 'four hundred each', *ikkēr pin* 'two thousand each'.

Collective numerals designate a certain number of items from a whole, and are formed from the emphatic forms with *-(Ā)n*, e.g. *ikkēn* 'two of them', *pillēkēn* 'five of them, the five together'. Forms with similar meaning can be formed with the possessive suffixes, in the first and second persons from the clipped forms and in the third person from the emphatic forms, e.g. *iksēmēr* 'we two, the two of us', *iksēr* 'you two, the two of you', *ikkēšē* 'those two, the two of them'.

Approximative numerals are formed by placing the numeral *pěr* 'one' before the number or with *-senče*, e.g. *Pěr śēr tenkē kirlē pulat' pulē* 'We shall need around one hundred roubles', *Vāl utmālsenče pulē* '(S)he is around sixty years old'.

Verbs

Verbs may be simple words, e.g. *kay-* 'go', *pul-* 'become', words derived by lexical suffixes from nouns, adjectives or verbs, e.g. *puśla-* 'begin' ← *puś* 'head', *vatał-* 'get old' ← *vata* 'old', or combinations of converb forms with

a second verb and of nouns with the auxiliary verb *tu-* 'do, make', e.g. *ilse kay-* 'take away' ('take and go'), *astu-* 'remember' ← *as* 'memory'. The stem of a verb is homonymous with the second-person singular imperative, e.g. *śi-* 'eat' - = *Śi* 'Eat!'

Voice, Possibility and Negation Suffixes

Verbs in the active voice lack a suffix, and may be inherently transitive, e.g. *vula-* 'read (something)', or intransitive, e.g. *kay-* 'go'. The reflexive voice is formed with *-(Ā)n*, e.g. *sāvān-* 'wash oneself' ← *śu-* 'wash (transitive)', but transitive verbs ending in *-t*, *-s*, *-ś*, *-r* form the reflexive with *-(Ā)l*, e.g. *uśāl-* 'open (intransitive)' ← *uśa-* 'open (transitive)'. Verbs with *-(Ā)n* sometimes function as passives, e.g. *śirān-* 'be written' ← *śir-* 'write'. Cooperative-reciprocals are formed with *-(Ā)ś*, e.g. *vārās-* or *vārs-* 'fight' ← *vār-* 'hit'.

Causatives are formed with *-tAr* (after consonants) ~ *-ttAr* (after vowels, with some exceptions) or its variants, e.g. *śirtar-* 'cause (someone) to write' ← *śir-* 'write', *vulattar-* 'cause (someone) to read' ← *vula-* 'read'. The variant *-t* is added to verbs ending in *-a*, *-e*, *-n*, *-r*, e.g. *χārat-* 'frighten' ← *χāra-* 'be afraid', *vērent-* 'teach' ← *vēren-* 'learn'. The variant *-Ar* is added to one-syllable verbs ending in *-t*, *-k*, *-s*, *-ś*, *-l*, e.g. *ūker-* 'drop' ← *ūk-* 'fall'.

Ability to perform an action is indicated by adding *-Ay* to the stem (final suffix vowel drops), e.g. *Yelena tuntikunniččen ēše tuχaymastāp terē* 'Yelena said she could not go to work until Monday'. Negation is expressed by adding a negative marker to *-Ay*.

The suffixes mentioned are attached in the order voice + ability + negation, e.g. *yurla+ttar+ay+ma-* 'not be able to make (someone) sing'.

Actional Modifications

There are various combinations of two verbs in which the first one is a non-finite form in *-sA*, *-A* or *-mA*, conveying the main action, and the second verb is an auxiliary verb, marking an actional modification.

The beginning of the action is indicated by *puśla-* 'begin' and *tūtān-* 'take up' following a main verb form in *-A* or *-mA*, e.g. *Epē čávašla χaśat vula (~vulama) puślarām* 'I have begun to read Chuvash newspapers', *Šumār śuma (~śava) tūtānčē* 'It started to rain'. The onset of a process may be expressed by the participle in *-AkAn* plus *pul-* 'become', e.g. *Kuś kurakan pulčē* 'His eyes started to see'.

Duration or regularity of action is indicated by *tār-* 'stand', *pīr-* 'go, arrive' and *yul-* 'stay', following converbs in *-sA*, e.g. *Epē Xīpar' χaśata vulasa tāratāp* 'I (regularly) read the Xīpar newspaper', *Urampa pēr śin čupsa pīrat* 'A person is running through the street'. Completion or cessation of action is indicated by *pēter-* 'end', *čārān-* 'stop' and *pāraχ-* 'throw, abandon' following a converb in *-sA*, e.g. *Epē śirsa pēterēm* 'I finished writing', *Ačasem kēr yenne pulla śūreme pāraχrēs* 'The children stopped going fishing by autumn'. The verbs *kay-* 'go', *śit-* 'reach' and *tuχ-* 'go out' express actions performed thoroughly or completely, e.g. *Esē ku kalava vulasa tuχrān-i?* 'Did

you finish reading this article?' The verb *lar-* 'sit' indicates the end of a process, e.g. *Śirta sünse larčĕ* 'The candle went out', and *χur-* 'place, put' an action completed before another action, e.g. *Uksĭne tĭlese χună* 'She has already paid her money'.

Finite and Non-finite Verb Forms

Verb forms having finite functions appear at the end of the sentence, whereas verb forms having non-finite functions do not. Most verb forms in finite use can be conjugated according to tense, aspect, mood, person and number. Converbs have non-finite functions, whereas participles may have both non-finite and finite functions.

Finite Verb Forms

Verbs are conjugated by adding grammatical suffixes to express tense–aspect, mood, person and number. Personal markers expressing person and number appear as the final element (Table 27.10).

Tense, Aspect and Mood

Chuvash expresses tense, aspect and mood through the addition of suffixes to the stems of verbs.

The present marks, as an intraterminal item, an event that is in progress, habitual or scheduled, e.g. *Śiru śiratăp* 'I am writing a letter', *Muskav patne sĭvlăš karapĕsem tătăšaχ vēšĕsĕ* 'Planes fly to Moscow every day', *Epĕ ĭran Muskava kayatăp* 'I will go to Moscow tomorrow'. The suffix is $-(\theta)At$ (final stem vowel drops (see also p. 436); negative $-mAst$), the third-person plural is $-As$ (negative $-mAs$).

The simple past indicates an event in the past, e.g. *Yultan tuχsan, ikĕ kilometra yaχăn šerempe utrămăr* 'When we left the village, we walked about two kilometres along the pasture'. The suffix is $-R$, i.e. mostly $-r$, $-t$ after stems ending in $-n$, $-l$, and $-r$, and $-č$ in the third person. The negative form is $-mAr$.

The perfect, a post-terminal with indirective nuances, marks an event that has taken place, often without being witnessed by the speaker. The suffix is $-nĂ$, negative $-mAn$; person and number are not determined by personal

Table 27.10 Personal markers of finite verbs

	Present	Future	Simple past, imperfect	Pluperfect, subjunctive	Imperative
1sg.	-ăp/-ĕp	-p	-ăm/-ĕm	-ăm/-ĕm	-am/-em
2sg.	-ăn/-ĕn	-n	-ăn/-ĕn	-ăn/-ĕn	∅
3sg.	[-']	∅	-ĕ	-čĕ	-tăr/-tĕr
1pl.	-păr/-pĕr	-măr/-mĕr	-ămăr/-ĕmĕr	-ămăr/-ĕmĕr	-ar/-er
2pl.	-ăr/-ĕr	-r	-ăr/-ĕr	-ăr/-ĕr	-ăr/-ĕr
3pl.	-sĕ	-s	-ĕs	-čĕs	-čĕăr/-čĕĕr

markers. The potential indirective nuances may be corroborated by adding particles such as *mĕn* 'apparently', e.g. *Atte kun sĭnĕn ĕnerĕχ pĕlnĕ mĕn* 'It turns out father already knew about this yesterday'.

There are two pluperfects, past post-terminals, indicating past events that had occurred prior to some point in the past. One, $-nĂ$ followed by $-čĕĕ$ 'was', is more focal (resultative), e.g. *Ĕslenĕčĕĕ* '(S)he had worked', *kaynăčĕĕ* 'was gone'. The other one, $-sA$ (negative $-mAsA$), followed by $-tt/-čĕĕ$ + personal markers 'I was (etc.)', is more event-oriented, e.g. *Epĕ χasat valli stat'a sĭrsattăm ta ilse kilme astuman* 'I had written the article for the newspaper, but I forgot to bring it'.

The imperfect, an intraterminal past, indicates a single, repeated or habitual past event envisaged in its course, e.g. *Pičĕ Muskavra purănatĕĕ* 'My older brother was living in Moscow', *Aĕa ĕuχ epĕ yumaχ itleme yuratattăm* 'I loved to listen to stories when I was a child'. The suffix is $-(\theta)Att$ with the variant $-(\theta)Atĕ$ in the third person (final stem vowel drops; negative $-mAst/-mAstĕ$). It consists of $-(\theta)At$ plus elements meaning 'I was (etc.)'. First- and second-person forms may combine with $-čĕĕ$ 'was', e.g. *Kayattămĕĕĕ* 'I went, was going, used to go'.

The prospective (future) expresses events that the speaker is not certain will occur, e.g. *Śil yăsărka pulsana ta, šavaχ kile sĭtĕpĕr* 'Although there are snowdrifts, we will manage to reach home'. The suffix is $-(\theta)Ă$ with the variant $-(\theta)ĕ$ in the third person (final stem vowels drop; negative $-mĂ/-mĕ$).

The optative–imperative paradigm includes forms with voluntative and imperative meaning. The first-person singular in $-(\theta)Am$ and plural in $-(\theta)Ar$ (final stem vowel drops) express a suggestion or willingness to perform an action that needs permission, e.g. *Epĕ te sĭrĕnpe pĭram-i?* 'May I go with you, too?' The negative is formed by postposing the word *mar*. The second-person form has imperative meaning, expressing, in the suffixless singular, a command or request for the performance of an action, e.g. *Śak sĭrăva vulasa par mana* 'Read this letter for me'. The plural suffix, which also can be used to address one person (politely), is $-(\theta)Ăr$ (final stem vowel drops), e.g. *Uttă seχĕtre puχăva pĭrăr* 'Please come to the meeting at 6 o'clock'. The word *an* is preposed to form the negative. The third-person form expresses a desire or instruction that an action be performed by a third party, e.g. *San'uk ku kĕnekene man valli te ĭltĕr-χa* 'Sanyuk should buy that book for me, too'. It is marked in the singular by $-tĂr$, in the plural by $-čĕĂr$ (after stems ending in vowels and $-n$, $-l$, $-r$), and $-čĂr$ (after other consonants). The word *an* is preposed to form the negative.

A subjunctive indicates counterfactuality, e.g. *Śĭrăttăm* 'I would write'. The suffix is $-(\theta)Ătt$ with the variant $-ĕčĕ$ in the third person (final stem vowel drops, negative $-mĂtt/-mĕčĕ$). The form consists of the future in $-(\theta)Ă/-(\theta)ĕ$ plus elements meaning 'I was (etc.)'. A past subjunctive is formed with $-nĂ$, followed by *pulătt-* and personal markers, e.g. *Epĕ sĭrnă pulăttăm* 'I would have written', *Epĕ văχătra instituta vērenme kĕnĕ pulsana, χalĕ ĕntĕ vērense*

tuxnă pulăttăm 'If I had entered the institute in time, I would have graduated by now'. Addition of *-ččě* 'was' expresses the wish that an action would have taken place, e.g. *Kaysa kurăttămččě Muskava!* 'How I wanted to see Moscow!'

The necessitative expresses obligation to perform an action and is marked by *-mAllA*. The form can be used impersonally or, for example, with a genitive-marked subject as in *Sanăn instituta kilmelle* 'You have to come to the institute'. The negative is expressed by postposing *mar*.

Non-finite Verb Forms

The participles include the present (intraterminal) participle in *-(θ)AkAn* (final stem vowel drops (see also p. 436), no negative), the past (post-terminal) participle in *-nĀ* (negative *-mAn*), and the future (prospective) participle in *-(θ)As* (final stem vowel drops, negative *-mAs*, *-As mar* or *-As šuk*).

They can be used as the non-finite verb of a relative clause attributed to a head, e.g. *Atner malta pīrakan, politikāpa interesleneken āslă-puślă šin šutlannă* 'Atner is considered a progressive and intelligent person who is interested in politics', *kilnĕ šin* 'the person who came', *kiles šin* 'the person who is to come'. Headless relative clauses based on *-(θ)AkAn* or *-(θ)As* may serve as complements (subject, object, etc.), the participle taking declensional suffixes, e.g. *Xirĕš kalakansem pulmarĕš* 'There was no one to object', *Kolxozra tărăšsa ěšlekensene premi parašĕš* 'They give prizes to those who work conscientiously on the kolkhoz', *Ača numay šiyes šuk* 'One child will not eat much', *Epĕ tusa kürenteresren ħaratăp* 'I am afraid I will offend my friend'. For the finite use of *-nĀ*, see p. 444; for its use with postpositions, see p. 447.

Verbal nouns derive from participles, *-(θ)AkAnni* ← *-(θ)AkAn + i*, *-ni* ← *-nĀ + i*, *-(θ)Assi* ← *-(θ)As + i*, *-mAlli* ← *-mAllA + i* (with affirmative and negative forms). They do not function as attributes, but as non-finite verb forms in complement clauses, e.g. *Mana anne ċirleni păšăryxantarat* 'That my mother has become ill worries me', *Epĕ Muskava kayassi šinċen šuxăšlatăp* 'I am thinking about going to Moscow'.

Converbs typically function as non-finite verb forms in adverbial clauses. The converb in *-sA*, negative *-mAsĀr*, may connect two events in a non-modifying, e.g. sequential, or modifying way, e.g. *Xĕvel ansa tĕttĕm puliċċĕcen tăxĕtăr šakănta* 'You wait here until the sun goes down and it is dark', *Nürĭĕ šantalăka tipĕtse, kăntăr yenċen āšă šil vēret* 'A warm wind blows from the south, drying out the wet air'. With the particle *ta*, it can mark clauses of concession. The negative form, often with the particle *aħ*, may denote 'without doing', e.g. *Xăyĕnċen iytmasăraħ kalama pĕlmestĕp* 'I cannot say without asking him/her'. The converb in *-sAn* ~ *-sAssĀn* (negative *-mAsAn*) marks clauses of time, e.g. *Văl kalama ċarănsan, āna iytusem paċĕš* 'When (s)he finished speaking, they asked him/her questions', and of condition, when the main verb is in the subjunctive, e.g. *Šinċeħ tuxĕtăr patne kaynă*

pulsan, ċĕrĕĭĕċċĕ-i ten man ača? 'If (s)he had gone to the doctor in time, maybe my child would have recovered?' For the use of the converb in *-(θ)A* 'doing' (final stem vowels drop; no negative), see p. 443.

The converb in *-(θ)iċċĕn* 'until, as long as (doing)' (final stem vowel drops, negative *-miċċĕn*) indicates limitation of an action, e.g. *kiliċċĕn* 'until coming'. It may combine with *-tArAħ* to express 'before doing', e.g. *Esĕ văl kayiċċĕntereħ kilsen avanċċĕ ĕntĕ* 'It would have been better if you had come before (s)he left'. The temporal converb in *-ArAħpA* consists of *-A + -rAħ + instrumental*, e.g. *Yal šivăħne šitereħpe ačasem tata tepĕr pĕċĕksĕ kĕper urlă kasrĕš* 'Just as they reached the outskirts of the village, the children crossed still another small bridge'. Other converbs include *-mAsseren* 'every time one does', e.g. *Xulana tuxmasseren kĕneke magazinne kĕretĕp* 'Every time I go to the city I go in the bookshop', *-nĀsem(ĕn)* 'to the degree that/as much as one does', e.g. *Petruħin Entriyalĕ patnelle šivħarnăšemĕn iĭlarăħ pălħansa piċĕ* 'The closer Petrukhin got to Entriyalĕ the more nervous he became', *-nĀrAnpA* 'since/from the time one does', e.g. *Epĕ āna kurnăranpa numay văħĕt irtrĕ* 'It has been a long time since I saw him/her', *-UšĀn* 'doing at the same time as', e.g. *Magazinten kilüšĕn kĕr, ċey ĕšĕpĕr* 'On your way from the store, drop in and we will have tea'. The participle in *-nĀ* may combine with postpositions to form converbs, e.g. *Atte ĕšren kilnĕ šĕre anne apat ħatĕrlerĕ* 'Mother prepared dinner by the time father came home from work'.

The so-called infinitive in *-mA* is a purposive adverbial form, indicating the goal or result of the action expressed in a second verb, e.g. *Škula kayma šutlamarĕ* '(S)he decided not to go to school'.

Adverbs

Adverbs may be simple words used adverbially, e.g. *kĕš* 'now', *pitĕ* 'very', or words derived from nouns, adjectives and pronouns by productive and less-productive case suffixes, e.g. *maltan* 'from the front, at first', *kunĕpe* 'all day', *uššăn* 'openly', *malalla* 'forward', *šullen* 'annually'. They may be modifiers of time, e.g. *payan* 'today', *ħalĕ (ħal')* 'now', *kĕš* 'soon', place and direction, e.g. *kunta* 'here', *unta* 'there', *lere* 'to there', degree, e.g. *saxal* 'few, little', *ċilay* 'a lot', *paċ(aħ)* 'completely', or manner, e.g. *pĕrle* 'together', *sasartăk* 'suddenly', *ħăvărt* 'quickly'. There are also reduplications such as *unta-kunta* 'here and there'.

Other Morphological Categories

Simple postpositions governing the nominative case include *kaša* 'during', *pek* 'like', *pirki* 'because of', *šinċen* 'about', *tavra* '(movement) around', *tărăħ* 'along', *valli* 'for', *vitĕr* 'through, across'. The dative-accusative is governed by *kura* 'considering', *may* 'in the direction of', *păħmasăraħ* 'in spite of', *pula* 'due to, because of', *šiti* (~ *šite*) 'up to, as far as', *ħirĕš* 'against, opposite'. The ablative is governed by *puślasa* 'beginning from' and *pušne* 'except, besides'.

Postpositions based on auxiliary nouns, typically expressing spatial

relationships, contain a possessive suffix plus case suffixes (dative-accusative, locative, ablative or instrumental), whereas a preceding modified nominal is in the genitive or the nominative, e.g. *man umra* 'in front of me', *sĕtel sine* 'on the table', *yal sumĕpe* 'around the village', *vĕrman ħuššipe* 'through the woods'. The spatial nouns include *āš(ĕ)* 'inside', *kut* 'under part, base', *pat* 'closeness', *sĭ* 'upper part, surface', *šum* 'environs', *tārā* 'top', *um* 'front part', *vĭrĕn* 'place', *ħĕr(ĕ)* 'edge, bank', *ħiś* 'back part', *ħušā* 'interval', *yen* 'side'. Certain suffixes of recent origin have developed from postposed words: *-ĕn* (*-ĕĕn* after vowels and *-n*, *-l*, *-r*) 'until', e.g. *ĭranĕĕn* 'until tomorrow', *tāħār seħetĕen* 'until nine o'clock', *-seren* 'every', e.g. *kilseren* 'every home', *šulseren* 'every year, annually', and *-šĀn* 'for', e.g. *Epir piśāk tuħāšlā tĭrpulšĕn ĕerešetpĕr* 'We are struggling for a high-yield harvest'.

Conjunctions such as *tA* 'and, also', *tata* 'and', *anĕaħ (ta)* 'but', *šapaħ (ta)* 'but' may coordinate two or more constituents, e.g. *Sak ħĕr aha šullĕ te kapār* 'This girl is tall and beautiful', *Epir ĕner sirĕn patāra pĭrsattāmār, anĕaħ esir kilte pulmarār* 'We came to your house yesterday, but you were not at home'. Subordinative conjunctions are rare, e.g. *tese* 'saying'.

Among the modal particles, *-ške* (~ *-eške/iś*) confirms, and *-Aħ, -mĕn* and *-ħa, -sker* ~ *-raħsker* intensify a statement.

Interjections stand in isolation or at the head of a sentence, followed by a pause. They include *Aħ!* 'Oh!', *Ey!* 'Hey!', *Aytāraħ!* 'O, my God!' and words of etiquette such as *Salam!* 'Hello!', *Sivā-i!* 'Hi!', *Kileħ!* 'Please!', *Tavtapuś!* ~ *Tavssi!* 'Thank you!', *Sivā pul(ār)!* ~ *Sivpul/sĭpul!* 'Goodbye!'

Syntax

Words may be connected in phrases and sentences through agreement in person and number, e.g. *Epĕ vularām* 'I read', juxtaposition, e.g. *šurā ĕĕpe* 'white shirt', or government, e.g. *Atte Šupaškarta purĕnĕĕ* 'My father lived in Cheboksary'.

Nominal Phrases

The components of a nominal phrase stand in an attribute-noun relationship where the attribute may be a noun, e.g. *ĕul śurt* 'stone house', adjective, e.g. *ilemlĕ ĕĕpe* 'beautiful dress', or other part of speech. Attributes are placed before the nouns they qualify, but do not agree in number, person or case with those nouns.

The Use of Cases

Nominals in the nominative stand as the subject, as the nominal predicate of a sentence, etc., e.g. *Aĕasem viĭ'asšĕ* 'The children are playing' or *Man piĕĕe saltak* 'My older brother is a soldier'.

The genitive marks the possessor in a possessive construction, in which the possessed carries a third-person possessive suffix *-ĕ/i*, e.g. *surĕħĕn tirĕ* 'the skin of a sheep'. This construction contrasts with the type of compounds in

which the attributed noun is in the nominative, e.g. *surĕħ tirĕ* 'skin of (any) sheep' = 'sheepskin(s)'. Examples illustrating the two types are *Pĕltĕr ilnĕ surĕħĕn tirĕ pitĕ layĕħ* 'The skin of the sheep purchased last year is very good', where the relative clause *pĕltĕr ilnĕ* is attributed to the noun meaning 'sheep', and *Surĕħ tirĕ ĕĕrĕk šĕletme kirle* 'A sheepskin is necessary for making a wool coat'.

The dative-accusative marks direct and indirect objects, with the distinction between these two functions apparent from the meaning of the verb. Its basic function is to express the goal or the recipient of the action described by the verb, e.g. *Epir irpe šupaškara śitrĕmĕr* 'We reached Cheboksary in the morning'.

The locative indicates location in time or space, e.g. *Atte ħĕĕpe ħulara vĕrense purĕnĕĕ* 'My father lived and studied in the city all winter', *Nikolay Ašmarin 1870 śulta śuralnĕ* 'Nikolay Ashmarin was born in (the year) 1870'.

The ablative case indicates the source, origin or starting point of an action, state or quality, e.g. *Aĕasem uyran ĕĕĕkem tatsa kilĕś* 'The children picked flowers and came home from the field', *yĭvāšran tunĕ pĭrt* 'house made of wood', *manran piśāk* 'bigger than me' (comparative construction).

The instrumental indicates an instrument or a comitative participant in an action, e.g. *Kolħozra tĭrpula mašinĕpa vĭrsa puštaraśšĕ* 'They harvest grain by machine on the kolkhoz', *Manpa pĕrle kilĕĕ* '(S)he came together with me'. It can also mark the place through which an action is carried out, e.g. *Urampa pĕr śin ĕupsa pĭrat* 'A person is running through the street', and the time or reason of an action, e.g. *Epir sivĕpe ĕĕtrĕrĕmĕr* 'We shivered from the cold'.

Verbal Phrases

In a phrase whose main component is a verb, the dependent component may be a noun, e.g. *ħulara purĕn-* 'live in a city', pronoun, e.g. *Kama kurtār?* 'Whom did you see?', adverb, e.g. *ħĭttĕn ũsĕr-* 'cough loudly', verbal form, e.g. *tārāšsa ĕśle-* 'work conscientiously', etc.

The Sentence

Constituent Order

The parts of a simple sentence typically occur in the order of subject-predicate, e.g. *Aĕasem viĭ'asšĕ* 'The children are playing', while the secondary parts are inserted within this framework, e.g. *Aĕasem uyra viĭ'asšĕ* 'The children are playing in the field'. A focused constituent typically takes the position directly before the verb.

Expression of 'Being' and 'Having'

Modal words and particles include the words *pur* 'there is/are' and *śuk* 'there is/are not' which affirm or negate the existence of a person or item, e.g. *Magazinta sĕt pur* 'There is milk at the store', *Magazinta sĕt śuk* 'There is no

milk at the store', or are used to express the notion 'to have' or 'not to have' in possessive constructions, e.g. *Manän kĕnekem pur* 'I have a book' (literally 'My book exists'). The third-person singular past suffix *-čĕ* (*-čĕčĕ* after vowels) may be used as a general past marker ('was') with non-verbs, e.g. *Šantaläk uyarčĕčĕ* 'The weather was clear'.

Negation

The postposed word *mar* '(it is) not' is used to negate constituents, e.g. *Pirĕn yaltan vărman ayakra mar* 'It is not far from our village to the woods'. Verb stems are, however, negated with *-mA*. A preposed *an* 'no, not' negates second- and third-person imperatives, e.g. *An kay* 'Do not leave!'

Interrogatives

Yes/no questions are asked by attaching one of the question particles *i*, *im*, *ši*, *šim*, each with a different nuance, e.g. *Väl ħulana kaynă-i?* 'Did (s)he go to the city?' (yes/no response), *Väl ħulana kaynă-im?* 'Did (s)he really go to the city?' (tone of uncertainty or surprise), *Väl ħulana kaynă-ši?* 'You mean (s)he went to the city?' (doubt, need for confirmation), *Väl ħulana kaynă-šim?* '(S)he did not go to the city, did (s)he?' (doubt, emotional). For interrogative pronouns and combinations with them, see p. 440.

Expressions of Modal Nuances

Modal nuances are also expressed with various periphrastic constructions. Predilection for performing an action ('like/love to do') is expressed by *-mA* followed by *yurat-* 'please, suit', e.g. *Esĕ mĕn šime yuratatän?* 'What do you like to eat?' Desire to perform an action may be expressed by the participle in *-(Ø)As* plus a form of *kil-* 'come', e.g. *Man(än) širas kilet* 'I want to write', by *-(Ø)As* plus a form of *te-* 'say', e.g. *Kinočĕen ěš puštaras tešĕĕ* 'They want to finish some work before the show', or by *-(Ø)AsšĀn* and *-(Ø)Asčĕ*, e.g. *Kăšal epĕ Muskava kayasšĀn* 'I want to go to Moscow this year', *Xăšan ta pulin Muskava kaysa kurasčĕ* 'I would like to visit Moscow someday'. Intention to perform an action is indicated by a converb in *-A* plus a form of *te-* 'say', e.g. *Epĕ šira tetĕp* 'I intend to write'. Need to perform an action is expressed by *-(Ø)As* plus *pulat* 'becomes', e.g. *Man(än) širas pulat* 'I need to write', *Pirĕn širas pulat* 'We need to write'. Need for something is expressed by a noun in the dative-accusative followed by *kirlĕ* 'necessary', e.g. *Mana ukša kirlĕ čuħne epĕ ūlaraħ ěšleme tărăšatăp* 'When I need money, I try to work more'. Possibility to perform an action may be indicated by *-mA* plus *pulat* 'becomes', e.g. *Manän kayma pulat* 'I can go (if I want to)', or by *-mA* plus *yurat* 'pleases, suits', e.g. *Sirten iytma yurat'-ši?* 'Can I (~ is it all right if I) ask you something?'

Complex Sentences

Coordinated sentences are connected by means of conjunctions (see p. 448) or by juxtaposition, e.g. *Epĕ apat ħatĕrlem, esĕ väl ħušăra ħvattere tirpeyle*

'I will prepare dinner and meanwhile you tidy up the flat'. Subordinate clauses precede the main predicate and are based on non-finite verb forms such as converbs and participles. For example, the converb in *-sAn* can express condition, e.g. *Atălpă ūl usal pulsan, epir turan kayăpăr* 'If the road along the Volga is bad, we will go by the mountains', or time, e.g. *Yaltan tuħsan, ikĕ kilometra yaħăn šerempe utrămăr* 'When we left the village, we walked about two kilometres along the pasture'. Concession may be expressed with *-sA* or *-sAn* plus the particle *tate*, or with the formations *pulin te* and *pulsan ta*, e.g. *Šil yăšărkka pulsan ta, šavaħ kile šitĕpĕr* 'Although there are snowdrifts, we will manage to reach home'.

Lexicon

The basic store of the Chuvash lexicon consists of words of the common vocabulary of the Turkic languages. Such words belong to parts of speech like pronouns and numerals, but also to every lexical domain, although the relationships may be obscured by the many sound changes in Chuvash. Replacements of vocabulary in these domains have occurred through internal and external processes. Formations such as *pitšămarti* 'cheek' ('egg of the face') have developed internally, whereas borrowing from Tatar, Mari and Russian has been the richest external source of replacements and additions.

Tatar has exerted a major influence on the lexicon, e.g. *yanaħ* 'chin' < *yaħak* 'cheek', *pišăk* 'big' < *bazik* 'thickset, large'. Through such heavy borrowing, even a few word-formation suffixes such as *-tAš* entered Chuvash from Tatar, e.g. *yultaš* 'comrade' < *yuldaš*. The majority of the Arabic and Persian loans entered through Tatar mediation, e.g. *čaršav* 'curtain' < Tatar *čaršaw* < Persian, *văħăt* 'time' < *vaqit* < Arabic. Most, if not all, of the Mongol words were mediated through Tatar after the thirteenth century, e.g. *yevčĕ* 'matchmaker' < Tatar *yawčĕ* < Mongol *jaγuči*, *pusaħa* 'threshold' < *busaγa* < Mongol *bosaγa*. There are more than a hundred Mari loanwords, especially in the Sundyr and other dialects, that reflect the assimilation by the Chuvash of a local Mari population, e.g. *lĕpĕ* 'butterfly' < *lipi*, *yantar* 'glass' < *yamdar*, *pürt* '(log) house' < *pört*. Pre-twentieth century Russian loans are fully assimilated, e.g. *apat* < Russian обедь 'dinner', *ħušat* < газета 'newspaper', *kĕneke* < книга 'book'. The newer, less assimilated layer is spelled as in Russian.

Dialects

Chuvash dialects are divided into the 'upper' or Viryal and the 'lower' or Anatri dialects, the latter forming the basis of the standard language. The 'transitional' dialects between them are often grouped together as the 'central' dialect. Those Chuvash settled outside the republic speak Anatri dialects. The small differences separating the Anatri and Viryal dialects include the following:

Anatri *u* vs. Viryal *o*, e.g. *pur* vs. *por* 'there is', *ulttā* vs. *olttā* 'six'. Anatri *Ā* vs. Viryal *Ā^o*, e.g. *šāltār* vs. *šā^oltā^or* 'star', *tělĕk* vs. *tĕ^olĕ^ok* 'dream'.

Anatri plural *-sem* vs. Viryal *-sAm*, e.g. *šinssem* vs. *šinsam* 'persons'.

Anatri genitive *-nĀn* vs. Viryal *-yān*, e.g. *lašānān* vs. *lašayān* 'the horse's'.

Anatri dative-accusative *-nA* vs. Viryal *-ya*, e.g. *lašana* vs. *lašaya*.

The converb *-sA* is used as the perfect in Viryal, corresponding to *-nĀ* in Anatri, e.g. *Vĕsem sire kurnā* vs. *Vāsam sire kursa* 'They saw you'.

References

- Andreev, I. A. (1966) 'Čuvaškij jazyk', in N. A. Baskakov (ed.) *Jazyki narodov SSSR* 2, *Tjurkskie jazyki*, Moscow: Nauka, 43–65.
- and I. P. Pavlov (eds) (1957) *Materialy po grammatike čuvaškogo jazyka*, 1. *Morfologija*, Čeboksary: Čuvašskoe gosudarstvennoe izdatel'stvo.
- Ašmarin, N. I. (1898) *Materialy dlja issledovanija čuvaškogo jazyka*, 1–2, Kazan.
- (1928–50) *Thesaurus Linguae Tschuvaschorum/Slovar' čuvaškogo jazyka*, 17 vols, Kazan-Čeboksary. Partially reprinted in Indiana University Publications, Uralic and Altaic Series 70:1–4, The Hague: Mouton.
- Benzing, J. (1959) 'Das Tschuwaschische', in J. Deny et al. (eds) *Philologiae Turcicae Fundamenta* 1, Wiesbaden: Steiner, 695–751.
- Egorov, V. G. (1964) *Ėtimologičeskij slovar' čuvaškogo jazyka*, Čeboksary: Čuvašskoe knižnoe izdatel'stvo.
- (1971) *Sovremennij čuvaškij literaturnyj jazyk v sravnitel'no-istoričeskom osveščeni*, Čeboksary: Čuvašskoe knižnoe izdatel'stvo. Second revised edition.
- Fedotov, M.R. (1992) *Materialy k istoriko-etimologičeskomu Slovarju čuvaškogo jazyka*, Čeboksary: Naučno-issledovatel'skij institut.
- Krueger, J. (1961) *Chuvash Manual. Introduction, Grammar, Reader, and Vocabulary*, Indiana University Publications, Uralic and Altaic Series 7, The Hague: Mouton.
- Levickaja, L. S. (1976) *Istoričeskaja morfologija čuvaškogo jazyka*, Moskva: Nauka.
- Róna-Tas, A. (ed.) (1982) *Chuvash Studies*, Budapest: Akadémiai Kiadó.
- Skvorcov, M. I. (ed.) (1982) *Čuvaško-russkij slovar'*, Moscow: Russkij jazyk.

Appendix

For the different scripts and writing systems used for Turkic Languages, reference is made to the following book: Peter T. Daniels and William Bright (1996) *The World's Writing Systems*, New York and Oxford: Oxford University Press. The chapters of special interest for Turkic languages are the following: György Kara 'Aramaic scripts for Altaic languages', pp. 536–58, Bernard Comrie 'Languages of eastern and southern Europe' (on Turkish pp. 682–4), Bernard Comrie 'Adaptations of the Cyrillic alphabet', pp. 700–26, Bernard Comrie 'Script reform in and after the Soviet Union', pp. 781–4. The bibliographies give further information on relevant literature.

For the older scripts dealt with in Chapter 6, 'Turkic writing systems', in the present volume, see András Róna-Tas (1991) *An Introduction to Turkology*, Szeged: University of Szeged.

A few examples of modern Turkic scripts are given below. Table 1 shows the Uzbek Cyrillic script, Table 2 the Kazakh Cyrillic script, and Table 3 a comparison of the Arabic script currently used for Uyghur with the Uyghur Cyrillic script.

Table 1 Uzbek script

Cyrillic	New Latin	Transliteration	Cyrillic	New Latin	Transliteration
а	a	a, á	о	o	á
б	b	b	п	p	p
в	v	w	р	r	r
г	g	g	с	s	s
Ғ	g'	γ	т	t	t
д	d	d	у	u	u, ü
е	e, [ye]	-e-, ye-	ў	o'	o, ö
ё	[yo]	yá	ф	f	f
ж	j	j, ž	х	x	χ
з	z	z	ҳ	h	h
и	i	i	ц	[ts]	ts
й	y	y	ч	ch	č
к	k	k	ш	sh	š
Қ	q	q	ь	'	'
л	l	l	э	e	e-
м	m	m	ю	[yu]	yu, yü
н	n	n	я	[ya]	ya, yá
нг ~ нҒ	ng	ŋ, ŋg ~ ŋγ			

Table 2 Kazakh script

а = a	о = o
ә = æ	ө = ö
б = b	п = p
г = g	р = r
Ғ = γ	с = s
д = d	т = t
е = e	y = (u)w or (ü)w
ж = ž	ұ = u
з = z	ү = ü
и = iy or iy	ф = r
й = y	х = χ
к = k	ҳ = h
Қ = q	ш = š
л = l	ы = i
м = m	і = i
н = n	ю = yuw or yüw
Ң = ŋ	я = ya or yæ

The remaining letters of the Russian alphabet are restricted to Russian loanwords, which are written according to their original orthography.

Table 3 Uyghur writing systems

1	2					3	4	1	2					3	4	
	2.1	2.2	2.3	2.4	2.5				2.1	2.2	2.3	2.4	2.5			
a	ا	ا	ا			Aa	Aa	q	ق	ق	ق	ق	ق	ق	К, к	К, к
ä	ا	ا	ا			Әә	Әә	k	ك	ك	ك	ك	ك	ك	Kk	Kk
b	ب	ب	ب	+	+	Bb	Бб	g	گ	گ	گ	گ	گ	گ	Gg	Гг
p	پ	پ	پ	+	+	Pp	Пп	ŋ	ڭ	ڭ	ڭ	ڭ	ڭ	ڭ	ng	Н, н
l	ل	ل	ل	+	+	Tt	Тт	l	ل	ل	ل	ل	ل	ل	Ll	Лл
j	ج	ج	ج	+	+	Jj	Ж, ж	m	م	م	م	م	م	م	Mm	Мм
č	چ	چ	چ	+	+	Qq	Чч	n	ن	ن	ن	ن	ن	ن	Nn	Нн
χ	خ	خ	خ	+	+	Hh	Хх	h	ھ	ھ	ھ	ھ	ھ	ھ	Hh	hh
d	د	د	د			Dd	Дд	o	و	و	و				Oo	Oo
r	ر	ر	ر			Rr	Рр	u	ۇ	ۇ	ۇ				Uu	Уу
z	ز	ز	ز			Zz	Зз	ö	ۆ	ۆ	ۆ				Öö	Өө
ž	ز	ز	ز			Zz	Ж, ж	ü	ۇ	ۇ	ۇ				Üü	Уу
s	س	س	س	س	س	Ss	Сс	w	ۇ	ۇ	ۇ				Ww	Вв
š	ش	ش	ش	ش	ش	Xx	Шш	e	ئې	ئې	ئې	+	+	Ee	Ee	
γ	غ	غ	غ	غ	غ	Юю	ҒҒ	i	ئى	ئى	ئى	+	+	Ii	Ии	
f	ف	ف	ف	ف	ف	Ff	Фф	y	ي	ي	ي	+	+	Yy	Йй	

1 Transliteration, 2 Kona Yeziq (2.1 listed, 2.2 isolated, 2.3 cluster final, 2.4 cluster medial, 2.5 cluster initial) 3 Yengi Yeziq, 4 Cyrillic

Index

The index covers themes, subjects and terms, but omits names of people.

- ablative case 39, 54, 111
- abtemporality 64
- accent 34–5, 111
- accordance 64
- accusative case 39, 111
- actancy patterns 52–3
- action clauses 60–61
- action nouns 46, 60
- actionality 41–2
- active participles 46
- adjective phrases 50–1
- adjectives 39–40
- adjunctors 48, 49
- adverbs 47–8
- Afghanistan 5, 8
- Afshar 82
 - people 5
- agglutination 36–7
- agreement 37, 49
- Akkadian 78
- allographs 126
- allophones 126
- Altaic hypothesis 77
- Altaic language family 16, 77
- Altay Turkic (Oyrot) 83, 87, 88, 403–15
 - adjectives 39
 - indirective forms 45
 - participium nondum facti 46
 - people 11
 - see also* South Siberian Turkic
- Altay-kiži 403
- Anatolia 236–40
- Anatolian Turkish 179, 181
- Anatri dialect 451–2
- Ancient Turkic 67–8
 - early dialects/isoglosses 68–9
- anteriority 64
- aoist 45, 116
- Arabic 3
 - loanwords 37
 - script 127, 131, 135–6, 158–9, 242–3
- Aramiac script 126
- Arghu Turkic 81, 82
- Armenian 3
 - script 135, 158–9
- Armeno-Kipchak
 - loanwords 165
 - texts 158
- aspect 41, 43, 114–16
- aspiration, notation xix
- auxiliary verbs 42
- Aynallu 82
 - people 5
- Azerbaijan 5, 29
- Azerbaijani 5, 82, 87, 248–60
 - adjectives 250–2
 - adverbs 256–7
 - comparison 251
 - conjunctions 257
 - consonant assimilations 34
 - consonant notation xxi–xxii
 - consonants 248–9
 - converbs 258–9
 - copulas 253
 - declension 251–2
 - dialects 260
 - equatives 57
 - first actant 258
 - intraterminals 44
 - Iranian 273, 276–80
 - lexicon 260
 - literary language 86, 248
 - modal forms 44
 - morphology 250–7

- morphophonology 249–50
 negative copulas 41
 nouns 250–2
 numerals 253
 participant reference 59
 particle *ki* 259–60
 particles 257
 people 5
 phonetics/phonology 248–9
 phonotactics 249–50
 post-terminals 44, 46
 postpositions 257
 primary stems 32
 pronouns 252
 relative clauses 61–2, 259–60
 second actants 53
 sentences 258–60
 sound harmony 250
 sound system 248–50
 syntax 257–60
 verbs 253–6
 vowels 30, 248
 word accent 249
 word formation 250–1
 writing systems 248
- Balkar 82, 87
 people 2, 6
 Bashkir 82, 87, 88, 283–300
 accent 284
 adjectival phrases 297
 adjectives 286
 adverbials 298
 adverbs 47, 296
 aorist 293–4
 categorical pasts 44
 compared with Tatar 284
 comparison 286
 conjunctions 296
 consonant assimilations 34
 consonants 283–4
 coordination 299
 declension 287
 demonstrative pronouns 40
 dialects 300
 enclitic particles 296
 genitive constructions 49
 interrogatives 299
 intraterminal converbs 47
 intraterminals 44, 295
 labial harmony 33
 lexicon 299–300
 morphology 286–96
- negation 299
 negative copulas 41
 nominal phrases 296–7
 nouns 286–7
 numerals 290
 orthography 285–6
 people 6–7
 personal markers on nominal
 predicates 41
 phonetics/phonology 283–4
 phonotactics/morphophonology
 284–5
 post-terminal converbs 47
 postpositions 48, 296, 298
 pronouns 40, 289–90
 prospective participles 46
 relative clauses 63
 sentences 298–9
 sound systems 283–6
 suffixes 284–5, 287–8
 syntax 296–9
 verbal phrases 297–8
 verbs 290–5
 vowel notation *xxi*
 vowels 31, 283
 word formation 286
- Bashkir Republic 8
 binomes 50
 Brahmi script 132–3, 138
 Bulgaria 4, 5, 23
 Bulghar tribal union 18, 22–3
see also Oghur
- cardinal numerals 40–1, 74
 case functions 54
 case suffixes 38–9, 111–12
 categorical pasts 44
 Caucasus 6
 causative suffixes 43
 causatives 55–6
 Proto-Turkic 75
 Central Oghuz 273–4
 Chaghatay 9, 166–77
 adjectives 170
 adverbial clauses 176–7
 adverbs 173
 conjunctions 174
 consonant notation *xxi*
 constituent clauses 176
 converbs 177
 copulas 171
 declension 170
 diachronic phonetics 169

- formation 166–7
 literary language 2, 87, 166–7
 morphology 170–4
 morphophonology 169–70
 nouns 170
 numerals 171
 origins 86
 particles 174
 passive 55, 175
 periods 167–8
 phonotactics 169–70
 possessive 174–5
 postpositions 174
 prepositions 174
 pronouns 170–1
 relative clauses 175–6
 sound system 169–70
 study of 168–9
 syntax 174–7
 verbal negation 175
 verbs 172–3
- Chechenia 6
 Cherkas 3
 Cherkessian people 2
 China 9–10, 79
 China Kirghiz 355
 Chulym Turkic 83, 403–15
 people 11
see also South Siberian Turkic
 Chuvash 7, 87, 88, 434–52
 accent 435
 action clauses 60
 action nouns 46
 adjectives 39, 438–9
 adjunctors 49
 adverbs 447
 aorist 45
 case suffixes 39, 434
 classification 82, 83
 clause junctions 66
 collective pronouns 440
 comparatives 57
 complex sentences 450–1
 conjunctions 48, 448
 consonant assimilations 34
 consonant lengthening 437
 consonants 71–2, 435
 converbs 446–7
 demonstrative pronouns 440
 dialects 451–2
 distributive numerals 41
 genitive constructions 49, 439
 imperative negation 58
 imperatives 434
 intensive forms 40
 interjections 448
 interrogatives 450
 intraterminals 44, 46
 lexicon 75, 79, 451
 Mari loanwords 451
 modal nuances 450
 morphology 437–48
 morphophonology 436–7
 necessative participles 46
 negation 450
 negative copulas 41
 nominal compounds 50
 nominal phrases 448–9
 non-modifying converb clauses 64
 nouns 437–8
 numerals 41, 74, 440–42
 ordinal numerals 41
 orthography 437
 particles 448
 passive 55
 past tense copula 41
 people 7–8, 434
 personal pronouns 74, 434
 phonetics/phonology 434–6
 plural suffixes 38, 51
 possessive pronouns 40, 439–40
 post-terminal (perfect) participles 46
 postpositions 48, 447–8
 pronouns 439–40
 prospective participles 46
 and Proto-Bulghar 81
 reflexive pronouns 40
 relative clauses 61–2
 second actants 53
 sentences 449–51
 sound system 434–7
 suffix notation *xx*, *xxii*
 suffixes 38, 73, 434
 superlatives 57
 syntax 448–51
 Tatar loanwords 451
 terminal converbs 47
 terminals 44
 verbal negation markers 57
 verbal phrases 42, 449
 verbs 442–7
 vowels 31, 434–5
 word structure 72
 classification of Turkic languages 82–3
 clause junctions 65–6
 collective numerals 41

comitative case 39
 Common Turkic 85
 branches 81–2
 split of Oghur (Bulghar) 81
 comparative case 39
 comparatives 39, 57
 compounds 50
 conditional clauses 64
 conjunctive adverbs (adjunctors) 49
 conjunctions 48–9
 conjunctors 48
 consonant assimilations 34
 consonant clusters 73
 consonant notations xix
 suffixes xxi–xxii
 consonants 31
 diachronic development 95–106
 diachronic suffix phonology 109–10
 Proto-Turkic 71–2, 95–9, 109–10
 contrast 64
 converb clauses 63–6
 converbs 46–7, 60, 117
 cooperative, Proto-Turkic 75
 cooperative-reciprocal suffixes 43, 54
 copula particles 45
 copula verb 42
 copulative forms 60
 copying *see* language contact, loanwords
 coreference 61–3
 Crimea 84
 Crimean Khans 27
 Crimean Tatar 9, 82, 87, 301–16
 accent 302
 actionality suffixes 42
 adjectives 305–7
 adverbs 314
 aorist 311
 comparison 306
 conjunctions 314
 consonants 301–2
 copulas 309
 declension 306–7
 diachronic changes 302
 dialects 316
 distributive numerals 41, 308
 equatives 54
 lexicon 316
 morphology 305–14
 nominal phrases 49, 314–15
 nouns 305–7
 numerals 41, 308–9
 ordinal numerals 41, 308
 orthography 304

particles 314
 people 8
 phonetics/phonology 301–2
 phonotactics/morphophonology 303–4
 postpositions 314
 pronouns 307–8
 sentences 315
 sound systems 301–4
 syntax 314–15
 verbs 309–13
 vowels 301
 word formation 305–6
 Cyprus 4
 Cyrillic script xviii, 136–7, 322, 454
 Dagestan 6
 dative case 39, 54, 111
 dative-accusative case 111
 declension 111–12
 demonstrative pronouns 40
 derivation notation xxii
 desideratives 42
 diachronic development 81–122
 dialect splitting 81–2
 dialectology *see* Turkish dialects
 diathetic patterns 54–6
 directive case 39
 discourse-pragmatic functions 58–60
 distributive numerals 41
 Dolgan 83
 people 11–12
 dynamic stress accent 34–5
 Early Chaghatay Turkic 86
 East European runiform script 127
 East Middle Turkic 85–6
 East Oghuz *see* Khorasan Turkic
 East Old Turkic 78, 79, 85
 inscriptions 81
 lexical stock 75
 East Turkic runiform script (runic) 127
 Eastern Turkestan 9
 echolaly 50
 enclitic copula particles 45
 enclitic particles 37
 equative case 39, 54
 equatives 57
 ethnic groups 3
 factive vs. non-factive 60–1
 finite clauses 57–60
 finite forms 43–5

finite thematic forms 43
 fini-transformatives 42
 first actants 52–6, 61–3
 focality 44
 focus 59
 frequentativity markers 42
 front vs. back harmony 31–4
 Fundamenta notation xviii
 future markers 116
 Fu-yü Turkic 2, 10, 83
 classification 83
 Gagauz 82, 87
 adverbs 47
 consonants 31
 people 5
 postpositions 48
 postpositive clauses 65
 Gansu province 10
 genitive case 39, 111
 genitive constructions 49–50
 Georgia 4
 Georgian 3
 Ghaznavids 24–6
 graphemes 126
 Greece 5
 Greek 3
 Halič 8
 head-final syntax 49
 Hebrew script 126–7, 136
 Hsiung-nu 17–18, 19
 Hualong Hui Autonomous County 10
 Hungarian 71–2, 74–5, 77, 95, 104–6
 Huns 17–18
 identity attribution 50, 61
 imperative negation 58
 imperatives 44
 Proto-Turkic 75
 imperfects 43–4
 impersonal predications 53
 indeclinable word classes 47–9
 indefinite article 38, 49
 Indic script 131
 indirective forms 45
 Indo-European languages 16
 and Proto-Turkic 78
 Indo-European influence 65–6
 initio-transformatives 42
 instrumental case 39
 instrumental converbs 64
 intensity markers 42

intensive forms 39–40
 intentionals 44
 interrogation markers 41, 57–8
 intersyllabic sound harmony 32–4
 intonation, diachronic phonology 111
 intrasyllabic sound harmony 31
 intraterminal converbs 47, 64
 intraterminal (present) participles 46
 intraterminality 42–4, 114–15
 Iran 5, 6, 28
 demography 276–7
 Iranian Turkic languages 78–9, 82,
 273–81
 Azerbaijani 273, 276–80
 Central Oghuz 273–4
 historical/demographic background
 276–7
 Khalaj *see* Khalaj
 Khorasan Turkic *see* Khorasan Turkic
 linguistic features 277–8
 Oghuz dialects 83, 274, 278–80
 South Oghuz 273–4, 276
 Turkmen 276–7
 Iraq 5
 Islam 1, 23–4
 izāfat, izafet 174–5, 200–1, 219, 244
 Judeo-Spanish 3
 juxtaposition 35–6
 Kabardinian people 2
 Karachay 82, 87
 adjective phrases 50–1
 nominal compounds 50
 people 2, 6
 reflexive pronouns 40
 verbal phrases 42
 Karachay-Balkar 88, 301–16
 accent 302
 adjectives 305–7
 adverbs 314
 aorist 311
 case suffixes 39, 306
 comparison 306
 conjunctions 314
 consonants 301–2
 copulas 309
 declension 306–7
 diachronic changes 302
 dialects 316
 lexicon 316
 morphology 305–14
 nominal compounds 50, 315

nominal phrases 314–15
 nouns 305–7
 numerals 51, 308–9
 orthography 304
 particles 314
 phonetics/phonology 301–2
 phonotactics/morphophonology 303–4
 postpositions 314
 pronouns 307–8
 sentences 315
 sound systems 301–4
 syntax 314–15
 verbs 309–13
 vowels 301
 word formation 305–6

Karachay-Cherkessian republic 6

Karagas *see* Tofa

Karaim 2, 82, 87, 301–16
 accent 302
 adjective phrases 50
 adjectives 305–7
 adverbs 314
 aorist 311
 comparison 306
 conjunctions 314
 consonants 31, 301–2
 copulas 309
 declension 306–7
 diachronic changes 302
 dialects 316
 lexicon 316
 morphology 305–14
 nominal phrases 314–15
 nouns 38, 305–7
 numerals 308–9
 orthography 127, 304
 particles 314
 people 8
 phonetics/phonology 301–2
 phonotactics/morphophonology 303–4
 postpositions 314
 pronouns 307–8
 sentences 315
 sound systems 301–4
 syntax 314–15
 verbs 309–13
 vowel notation xix
 vowels 301
 word formation 305–6

Karakalpak 9, 83, 87, 88, 318, 331

equatives 57
 orthography 331
 people 2, 10
 second actants 53

Karakalpakistan 10, 357

Karakhanid 85, 166
 adjunctors 49
 first actants 53
 middle voice 43
 nominal compounds 50
 people 24–6

Karluk 23–4

Kashkay (South Oghuz) 5, 82, 273–4, 276

Kazakh 9, 82, 87, 88, 318–31
 accent 320
 action clauses 60
 adjectives 322
 adjunctors 48, 49
 adverbs 327
 categorical pasts 44
 comparison 322
 conjunctions 327
 consonant assimilations 34
 consonant notation xxi
 consonants 319–20
 converbs 327, 329
 copulas 324
 declension 322
 demonstrative pronouns 40, 323
 dialects 330–1
 habitual past form 44
 indirective forms 45
 interrogative enclitics 328
 intraterminal participles 46
 intraterminals 43–4
 lexicon 330
 literary language 2, 87, 318
 morphology 322–8
 negative copulas 41
 nominal compounds 50
 nominal phrases 328
 nouns 322
 numerals 51, 323–4
 orthography 318, 321–2
 people 7, 10, 28, 319, 330
 personal markers on nominal predicates 41
 phonetics/phonology 319–20
 phonotactics/morphophonology 320–1
 possessive pronouns 40
 postpositions 327

primary stems 32
 pronouns 323
 relative clauses 61, 328
 script 454
 sentences 328–9
 sound system 319–22
 syllable structure 31
 syntax 328–9
 verbs 324–7
 vowels 319

Kazakhstan 5, 8, 9, 10, 12, 18, 29, 318–19

Kazan Tatar (Central dialect) 300

Khakas 11, 83, 87, 88, 403–15
 active participles 46
 classification 83
 collective numerals 41
 habitual past form 44
 intraterminal participles 46
 personal markers on nominal predicates 41
 primary stems 32

Khalaj 6, 81, 276, 280–1
 classification 82, 83
 consonants 71, 281
 diphthongs 70
 lexicon 280–1
 morphology 280
 numerals 51
 possessive constructions 56
 primary stems 32
 syntax 280
 vowel notation xix
 vowels 30, 280

Khazar Kaghate 22

Khorasan Turkic (East Oghuz) 82, 274–5, 276–7
 classification 275
 people 5

Khorezmian Turkic 86

Khotanese Brahmi script 133

Khwarezm 357

Khwarezmian Turkic 166–8

Kimeks 23–4

Kipchak 2, 9
 classification 82
 loanwords 79
 nations 6, 22
 origins 82
see also Middle Kipchak

Kipchak (northwestern) branch of Turkic 86

Kipchak Uzbek 83, 357

Kirghiz 83, 87, 88, 344–55
 accent 346
 action clauses 60
 adjectives 347–8
 adverbs 47, 352
 cardinal numerals 41
 case suffixes 39, 347–8
 categorical pasts 44
 Chinese 355
 classification 83
 collective numerals 41, 349
 conjunctions 252
 consonants 345–6
 converb clauses 63, 353–4
 converbs 353–4
 cooperative-reciprocal suffixes 43
 copulas 349
 declension 347–8
 demonstrative pronouns 40, 348
 dialects 355
 enclitic particles 352
 genitive constructions 49, 353
 habitual past form 44, 351
 intensive forms 39–40
 intraterminal participles 46, 351
 intraterminals 43–4
 labial (rounded vs. unrounded) harmony 33, 346
 lexicon 347, 354–5
 literary language 344
 modal forms 44, 351
 morphology 347–52
 negative copulas 41
 nominal compounds 50
 nominal phrases 353
 non-modifying converb clauses 64
 nouns 38, 347–8
 numerals 41, 348–9
 numerators 51
 ordinal numerals 41
 orthography 344
 participium nondum facti 46
 people 7, 10–11, 21–2, 28
 personal markers 45
 personal markers on nominal predicates 41
 phonetics/phonology 344–6
 phonotactics/morphophonology 346–7
 possessive constructions 56
 possessive pronouns 40
 possessive suffixes 51
 possibility markers 43

- postpositions 48, 352
 pronouns 348
 reflexive pronouns 40
 relative clauses 61–2, 63, 353
 second actants 53
 sentences 353–4
 sound system 344–7
 subjective predicatives 54
 suffixes 37, 347–8
 syntax 353–4
 third actants 53
 verbs 349–52
 vowels 344–5
 word formation 347
- Kirghizstan 8, 9, 10, 12, 29
 Kondoma 403
 Kosovo 4
 Kuman-Kipchak confederation 22
 Kumyk 6, 82, 87, 88, 301–16
 accent 302
 action clauses 60
 adjectives 305–7
 adjunctors 49
 adverbs 314
 aorist 311
 case suffixes 39
 comparison 306
 conjunctions 314
 consonants 301–2
 copulas 309
 declension 306–7
 diachronic changes 302
 dialects 316
 intraterminals 44
 lexicon 316
 morphology 305–14
 nominal phrases 314–15
 nouns 305–7
 numerals 51, 308–9
 orthography 304
 particles 314
 phonetics/phonology 301–2
 phonotactics/morphophonology 303–4
 post-terminals 44
 postpositions 314
 pronouns 307–8
 prospective participles 46
 relative clauses 62
 sentences 315
 sound systems 301–4
 syntax 314–15
- verbs 309–13
 vowels 301
 word formation 305–6
- Kurds 3
- labial harmony *see* sound harmony,
 rounded vs. unrounded harmony
- lambdaism 72, 105
- language contacts 78–9, 84, 91–2,
 110–11, 118, 120–2
- Latin script 136, 242–3
- Laz 3
- left-branching syntax 49
- lenis 32
- morphophonological variation 32
- notation xix
- lexicon 119–22
- Proto-Turkic 75–6
- limitation 64
- literary languages 2, 87–8
- Lithuania 8
- loanwords (lexical copies) 78–9, 122
- phonological adaptation 110–11
- pitch/stress accent 35
- Proto-Turkic 75
- syllable structure 31
- vowels 30
- locative case 39, 54, 111
- locative–ablative 111
- Macedonia 4
- Maimana 5
- Mamluk
- inscriptions 158
- lexicon 165
- Mandaean script 127
- Mandarin 379–80, 394–5
- Manichaean script 127, 129–31, 138
- manner adverbials 54
- means 64
- Meshkheti 4
- Middle Kipchak 158–65
- adjectives 160–61
- adverbs 164
- conjunctions 165
- declension 161
- lexicon 165
- morphology 160–65
- morphophonology 160
- nouns 160–61
- numerals 162
- orthography 159–60
- particles 165
- phonetics/phonology 159
- postpositions 165
- pronouns 161–2
- sound harmony 160
- sound system 159–60
- texts 158
- verbs 162–4
- word formation 160–61
- Middle Ottoman 181–2
- Middle Turkic 85
- writing systems 134–5
- middle voice 42–3
- Proto-Turkic 75
- Mishar Tatar (Western dialect) 300
- modal forms 44–5
- Modern Turkic writing systems 135–7
- modifying vs. non-modifying converb
 clauses 64–5
- Moldova 5
- Mongol invasions 26–7
- Mongolia 10, 21–2
- Mongolian 78
- consonants 71
- word structure 72
- Mongolic 76–8
- morphology 35–8, 111–17
- Proto-Turkic 73–5
- morphophonemes, notation xx
- morphophonological variation 31–4
- Mountain Bashkir (Eastern dialect) 300
- Mrass 403
- Nabataean script 127
- name forms xxii
- necessitative participles 46
- necessitives 44, 116
- negation 41, 43, 58
- Proto-Turkic 75
- negative converb markers 65
- negative copulas 41
- Nestorian script 134–5
- New Ottoman 181–2
- Noghay 83, 87, 88, 333–42
- accent 334
- adjectives 335–6
- comparison 335
- consonant notation xxii
- consonants 333–4
- converbs 341
- coordination 342
- declension 336
- dialects 342
- intraterminals 43–4, 339
- lexicon 342
- literature 333
- morphology 335–40
- morphophonology 334–5
- nominal phrases 340–1
- nouns 335–6
- numerals 337–8
- orthography 333, 335
- people 6, 27
- phonetics/phonology 333–4
- postpositions 340
- pronouns 336–7
- second actants 53
- sentences 341–2
- sound system 333–5
- syntax 340–42
- verbs 338–40
- vowels 333
- nomadism 16–17
- nomina loci 48
- nominal compounds 50
- nominal phrases 49–52
- nominal predicate personal markers 41
- nominal stems 37–8
- nominals 38–41
- Proto-Turkic 75
- nominative case 39
- non-finite clauses 60–6, 118–19
- non-finite forms 46–7
- non-propulsive relative clauses 63
- non-transformatives 42
- Northwestern Bashkir 300
- Nostratic hypothesis 77
- nouns 38–9
- case suffixes 39
- plural suffixes 38–9
- possessive suffixes 39
- Proto-Turkic 73
- numerals 40–1
- Proto-Turkic 74
- syntax 51–2
- numerators 51
- obligatives 44
- Oghur (Bulghar) branch of Turkic
- classification 82
- split from Common Turkic 81
- Oghuric Turkic people 18
- Oghuz dialects 274, 278–80
- classification 83
- Oghuz people 2, 5, 23–4
- Oghuz (southwestern) branch of Turkic
- 86

- classification 82
 loanwords 79
 origins 82
 Oghuz Uzbek 83, 357
 Old Anatolian Turkish 86
 Old Kirghiz inscriptions 85
 Old Ottoman 181–2
 Old Turkic 138–57
 adjectives 140–3
 adverbs 47–8, 148
 case suffixes 39, 142
 comparatives 141
 conjunctions 48, 148
 consonants 139–40
 converb clauses 153–4
 coordination 154–5
 declension 141–3
 dialects 157
 diathesis 155
 elatives 141
 equative 150
 inscriptions 138
 instrumental 150
 intensifiers 141
 interrogatives 152
 intraterminals 42
 lexicon 75, 156
 locative 150
 modal expressions 155
 morphology 140–9
 morphophonology 140
 negation 151–2
 nominal compounds 50
 nominal phrases 149–50
 nouns 140–3
 numerals 74, 144
 particles 148–9
 passive 55, 145
 period 85
 personal markers 75
 phonetics/phonology 138
 plural markers 141
 possession 152
 possessive markers 141–2
 postpositions 148
 pronouns 143–4
 reflexive markers 55
 register 157
 relative clauses 62, 154
 sentences 151–5
 similative 150
 sound system 138–40
 syntax 149–56
 text syntax 155–6
 verbal clauses 152
 verbal phrases 150–1
 verbs 144–8
 vowel notation xx
 vowels 138–9
 word formation 140–1
 word structure 72
 writing systems 126–34
 Old Uyghur 85
 adjunctors 49
 clause junctions 66
 collective numerals 41
 desideratives/similatives 42
 intensive forms 39
 manuscripts 138
 middle voice 42
 people 23
 scripts 20–1, 122–9, 138
 Old Uyghur Kaghhanate 21
 optatives 44, 111
 order of constituents *see* word order
 ordinal numerals 41, 74
 Orkhon Turkic
 adverbs 148
 inscriptions 20–1, 138
 suffixes 141
 verbal phrases 150
 verbs 147
 Orkhon Valley 19
 orthography xviii, 1, 20–1, 126–37
 Ottoman Turkish 86, 179–202
 action clauses 60
 action nouns 46
 adjectives 39
 adverbs 196–7
 Arabic verbal nouns 201
 consonants 184
 converbs 195, 198–9
 copula particles 192
 declension 188
 demonstrative pronouns 40
 determiners 190
 enclitic particles 37
 historical development 179–80
 literary language 2, 87, 88, 179–80
 loanwords 179–80
 morphology 187–97
 nominals 187–9
 orthography 182–3
 peoples 26–7, 28
 periodisation 181–2
 Persian-type clauses 199–200
 Persian-type nominal phrases 200–1
 personal markers 192
 phonetics 183–4
 phonological/morphophonological
 change 184–7
 phonology 183–4
 pronouns 189–90
 sound system 182–7
 syntax 197–202
 terminology 180–1
 Turkification of syntax 201–2
 verbs 190–96
 Oyro *see* Altay Turkic

 palatal harmony *see* sound harmony,
 front vs. back harmony
 palatalisation, notation xix
 Palmyran script 127
 parentheses xx
 participant reference 59–60
 participium nondum facti 46
 participles 46–7, 60–1, 116–17
 partitive case 111
 passive suffixes 42
 passives 55
 Proto-Turkic 75
 past tense copula 41
 Pechenegs 22
 perfects 44
 periodisation 84–7
 Persian 23, 167, 168–9, 276, 278–9
 Persian influence 65, 167, 199–201,
 278–81
 personal markers 41, 45, 47, 112
 Proto-Turkic 75
 personal pronouns 40, 112
 Proto-Turkic 74
 Phags-pa script 135
 phase structure 42
 phonemes 126
 transcription xviii
 phonology
 diachronic 88–111
 Proto-Turkic 69–72
 pitch accent 34–5
 plural suffixes 38, 51
 polite address 40, 51
 position 59
 possessive constructions 56–7
 possessive pronouns 40
 possessive suffixes 38–9, 51, 112
 possibility markers 41, 43
 post-terminal (perfect) participles 46
 post-terminal converbs 47
 post-terminality 42, 44, 45, 64, 115
 posteriority 64
 postpositions 48, 54, 112–13
 postpositive clauses 65
 postpositive markers 112–14
 postpredicative elements 58
 postverbs 42, 113–14, 115–16
 predications 52–7
 preference 64
 presents 43–4
 presumptives 45
 primary stems
 diachronic development 89–99
 irregularities 107
 morphophonological variation
 31–2
 prodeusive case 39
 progressives 43–4
 pronominal *n* 34, 112
 propulsive 63
 pronouns 40
 Proto-Turkic 74
 prosodic phenomena 34–6
 prospective (future) participles 46
 prospectives 45, 116
 Proto-Bulghar 81
 Proto-Kipchak 67
 Proto-Turkic 67–79
 Proto-Uralic 78
 purpose 64

 Qitan people 21–2
 Qizilsu 10

 reason 64
 reduced voice, notation xix
 reduced vowels 107–8
 notation xix
 Proto-Turkic 72–3
 reduplication 50
 reflexive markers 42–3, 54–5
 reflexive pronouns 40
 regional written languages 87–8
 regularity 36–7
 relative clauses 61–3
 rhotacism 71–2
 Romania 5
 rounded vs. unrounded harmony 33
 Runic script 127
 runiform alphabet 20–1, 127
 Russian 9
 Russian Federation 5

- Salar 83, 399–401
 consonants 401
 dialect 400
 first actants 52
 lexicon 400, 401
 liquid assimilation 399–400
 morphology 401
 people 10, 399
 sound system 400–1
 vowels 400–1
- Sanskrit script 131–2
- Sayan Turkic 83, 403–15
see also South Siberian Turkic
- second actants 52–3
- Seljuks 24–6
- sentence adverbials 58
- Serbo-Croat 3
- Shor 11, 83, 87, 88, 403–15
 conjunctions 48
- Siberian (northeastern) branch of Turkic 82
- simulative constructions 57
- simulative markers 42
- Slavic influence 65–6
- Soghdian
 alphabet system 21
 script 127, 128
- Sonqori 82
- sound harmony 31–4
 Proto-Turkic 73
- sound systems 30–35
- South Oghuz (Kashkay) 5, 82, 273–4, 276
- South Siberian people 11
- South Siberian Turkic 83, 403–15
 adjectives 407–9
 adverbs 414
 classification 404
 comparison 408
 conjunctions 414
 consonants 404–5
 declension 408–9
 diachronic remarks 405
 labial (rounded vs. unrounded)
 harmony 406
 lexicon 415
 morphology 407–14
 nouns 407–9
 numerals 410–11
 particles 414
 phonetics/phonology 404–5
 phonotactics/morphophonology 405–7
 postpositions 414
 pronouns 409–10
 sentences 415
 sound systems 404–7
 syllables 407
 syntax 414–15
 verbs 411–14
 vowels 404
- Soviet Union 1–2, 7, 29, 88
- specificity 59
- Stavropol 6
- stem formation 37–8
- Steppe Bashkir (Southwestern dialect) 300
- stress accent *see* dynamic stress accent
- subjective predicatives (subject adjuncts) 54
- subjunctors 48, 64
- subordinatives 48
- substitution 64
- suffixes 37
 actionality 42
 diachronic phonology 107–10
 morphophonological variation 32–4
 notations xx–xxii
 order of suffixes 37
 Proto-Turkic 73
- suffix *-ki* 39
- Sun-language theory 244
- superlatives 57
- syllable structure 31
- syntax 49–66, 118–19
- synthesis 35–6
- Syriac 3
- Syrian script 127, 138
- Tajik 9
 origins 82
 people 357
- Tajikistan 8, 10
- Tannu-Tuva 11
- Taranchi 83
- Tatar 82, 87, 283–300
 accent 284
 action clauses 60
 adjectival phrases 297
 adjectives 286
 adverbials 298
 adverbs 296
 aorist 45, 293–4
 classification 83
 comparatives 57
 compared with Bashkir 284
 comparison 286
 conjunctions 296
 consonants 283–4
 coordination 299
 declension 287
 dialects 300
 enclitic particles 296
 equatives 57
 focus 59
 genitive constructions 49
 interrogatives 299
 intraterminals 44, 295
 labial (rounded vs. unrounded)
 harmony 33
 lexicon 299–300
 literary language 87
 modal forms 44, 45
 morphology 286–96
 necessative participles 46
 negation 299
 negative copulas 41
 nominal phrases 296–7
 nouns 286–7
 numerals 51, 290
 numerators 51
 orthography 285–6
 palatal (front vs. back) harmony 33
 people 6–7
 phonetics/phonology 283–4
 phonotactics/morphophonology 284–5
 pitch/stress accent 35
 plural suffixes 38
 post-terminal (perfect) participles 46
 postpositions 296, 298
 primary stems 32
 pronouns 289–90
 relative clauses 61–2
 sentences 298–9
 sound systems 283–6
 suffixes 284–5, 287–8
 syntax 296–9
 verbal negation markers 58
 verbal phrases 297–8
 verbs 290–5
 vowel notation xxi
 vowels 31, 283
 word formation 286
- Tatarstan 7
- Tekke people 5
- Telengit 403
- Teleut 403
- temporal inclusion 64
- tense 41, 45
 anterior 42
 presents 43–4
 tense-indifference 45
- terminal converbs 47
- terminality 43–4, 64
- terminative case 39
- terminativity 64
- thematic suffixes 114
- third actants 53
- Tibetan script 133–4, 138
- Tofa (Karagas) 83, 87, 403–15
 consonants 31
 people 11
- Toja 403
 people 11
- Tokharian 78
- Tokharian Brahmi script 132–3
- topic (sentence topic) 58–9
- transcendence 55–6
- transcription xviii–xix
- transformatives 42
- Transoxania 8
- Tungusic 78
- Turcological notation xviii
- Türk Kaghanate 19–21
- Turkey 3
- Turkic–Mongolic relationship 76–7
- Turkish 82, 87, 203–34
 action clauses 60–1
 action nouns 46
 actionality 214
 adjective phrases 39
 adjectives 210–11
 adverbs 48, 216
 agreement rules 225–6
 attributive modifiers 218–19
 case-marking 221, 225
 case suffixes 209–10
 causatives 56, 213, 228
 clause combining 229
 clause junctions 65–6
 comparatives 57, 210–11
 complement clauses 229–30
 compounds 219
 conditional clauses 233
 conditionals 216
 conjunctions 48, 217
 consonant assimilations 34, 206–7
 consonants 31, 204–5
 constituent order 224
 converb clauses 63, 213–14, 231–3
 converbs 216

cooperative-reciprocal 213
 coordination 227–8
 declension 208–11
 definiteness 225
 demonstrative pronouns 40, 211, 218
 derivation 209, 213–14
 determiners 218
devrik cümle 224
 dialects *see* Turkish dialects
 diathetic relations 228
 distributive numerals 41, 212
 enclitic particles 37, 216–17
 equatives 54, 57, 210
 finite thematic forms 43, 214–15
 first actants 53, 220, 228
 focus 59, 224
 future 215
 gender 209
 genitive constructions 49, 218–19, 225
 harmony *see* sound harmony
 identity apposition 217–18
 imperative 215
 imperative clauses 224
 impossibility 213
 indeclinables 216–17
 indefinite article 218
 indirective forms 45, 215
 intensive forms 40, 211
 interrogation 215–16
 interrogation markers 57, 226
 interrogative pronouns 226–7
 interrogative sentences 226
 intraterminal converbs 47, 216
 intraterminals 43, 214
izâfat 219
 labial (rounded vs. unrounded)
 harmony 33, 206
 language reform 242–6
 lexicon 234
 literary language 2
 loanwords (lexical copies) 35, 37, 204–7, 213, 234
 manner adverbials 54
 mnemonic past 215
 modal forms 215
 morphology 208–17
 morphophonological variations 205–7
 necessitative 215
 negation 213–14, 227
 negative converb markers 65
 negative copulas 41, 227
 nominal compounds 50, 219
 nominal phrases 49, 217–22
 nominals 208–12
 non-finite clauses 229–31
 non-finite forms 216
 non-modifying converb clauses 64, 231–2
 nouns 38, 208–10
 numerals 51, 212
 optative 215
 orthography 207–8
 palatal (front vs. back) harmony 33, 206
 participant reference 59, 223
 participles 46, 216
 passive 55, 213, 228
 past tense 214
 past tense copula 41, 214
 periphrases 214
 personal markers 41, 45, 212
 personal pronouns 211
 pitch/stress accent 34, 35, 207
 plural suffixes 51, 209
 position/specificity 59, 225
 possessive ('have') constructions 56, 226
 possessive pronouns 40, 211
 possessive suffixes 51, 209
 possibility markers 43, 213
 post-terminals 42, 44, 214
 postpositions 48, 222
 postpositive clauses 65, 229
 postpredicative elements 58, 225
 postverbs 214
 predications 52
 presumptives 45, 212, 214–15
 primary stems 32, 205
 pronominal *n* 209
 pronouns 40, 211–12
 prosodic phenomena 207
 prospective 215
 quantifiers 218
 reciprocal 213
 reciprocal pronouns 211
 reflexive markers 54–5, 213
 reflexive pronouns 211
 relative clauses 61–3, 219–20, 231
 repeated forms (echolaly) 209
 second actants 53
 sentences 223–33
 sound harmony 206
 sound system 203–8
 speakers 3–4

specificity 225
 stress (dynamic) accent 35, 207
 subjective predicatives 54
 subjectless clauses 223
 subordinated clauses 223–4
 subordinated finite clauses 230–1
 suffix *-ki* 208
 suffix notation xx–xxii
 suffixes 37, 206–7, 208
 superlatives 210–11
 syllable structure 205
 syntax 217–33
 terminal converbs 47, 216
 text syntax 234
 thematic suffixes 214–15
 third actants 53
 topicalisation marker 58, 224
 verbal adjectives 216
 verbal adverbs 216
 verbal morphology 41–2, 212–16
 verbal nouns 216
 verbal phrases 42, 223
 verbal stems 38
 verbals 212–16
 voice 213
 vowels 30, 203–4
 word structure 35–6
 'yumuşak ğ' 136, 204
 Turkish dialects 236–40
 consonant assimilations 238
 consonants 237–8
 morphology 239–40
 morphophonology 238–9
 palatalisation 238
 phonetics 237–8
 phonology 237–8
 syntax 240
 vowels 237
 Turkish Language Society (TDK) 243–6
 Turkmen 82, 87, 261–71
 accent 262
 action clauses 60
 action nouns 46
 adjectives 263–5
 adverbs 269
 categorical pasts 44
 comparison 263–4
 conjunctions 269
 consonants 261–2
 converbs 271
 copulas 265
 declension 264
 dialects 271, 274–5
 equatives 57
 first actants 52
 intraterminal participles 46
 intraterminals 44
 lexicon 271
 literary language 2, 87, 261
 morphology 263–70
 morphophonology 262–3
 negative copulas 41
 nominal compounds 50
 nominal phrases 270
 nouns 263–5
 numerals 265
 numerals 51
 origins 86
 orthography 261
 participles 270–1
 particles 269–70
 people 5
 phonetics/phonology 261–2
 phonotactics 262
 possessive pronouns 40
 postpositions 269
 post-terminal (perfect) participles 46
 presumptives 45
 pronouns 264–5
 relative clauses 62
 sentences 270–1
 sound harmony 263–4
 sound system 261–3
 suffix notation xx–xxi
 syntax 270–1
 topicalisation marker 58
 verbal negation markers 58, 266
 verbs 265–9
 vowels 30, 261
 word formation 263
 Turkmenia 357
 Turkmenistan 5, 8, 12, 29, 357
 Tuvan 83, 87, 88, 403–15
 actionality suffixes 42
 cardinal numerals 41
 causatives 56
 classification 83
 collective numerals 41
 conjunctions 48
 consonant assimilations 34
 consonants 31
 dative 54
 demonstrative pronouns 40
 intraterminal converbs 47
 intraterminal participles 46
 negative converb markers 65

nominal compounds 50
 non-modifying converb clauses 64
 participium nondum facti 46
 people 11
 possibility markers 43
 post-terminal converbs 47
 primary stems 32
 prospective participles 46
 reflexive markers 54
 reflexive pronouns 40
 relative clauses 61
 verbal phrases 42
 twin words (binomes) 50

Ukraine 5
 Ulus 26–7
 unquoting particles 48–9
 Uralic 78
 urheimat 16, 68
 Uyghur 83, 87, 379–96
 accent 382
 adjuncts 49
 adverbs 47
 alphabet 386–7
 case marking 389
 causatives 56
 classification 83
 collective numerals 41
 comparatives 57
 conditional clauses 64
 conjunctions 48
 consonant assimilation 385–6
 consonant dissimilation 386
 consonants 381–2
 converb forms as adverbs 48
 declension 388–9
 dialects 395–6
 equatives 57
 first actants 52
 genitive constructions 49
 harmony 383–5
 indirective forms 45
 intraterminals 43–4
 lexicon 383, 394–5
 literary language 379–80
 modal forms 44
 morphology 387–93
 morphophonology 382–6
 negative copulas 41
 nominal compounds 50
 nouns 388–9
 orthography 136, 386–7, 455
 passive 55

people 3, 9–10, 379
 phonetics/phonology 380–2
 plural marking 388
 position/specificity 59
 possession marking 388–9
 possessive constructions 56
 possessive pronouns 40
 post-terminals 44
 postpositions 48
 presumptives 45
 primary stems 32
 pronouns 390
 prosody 382
 reflexive pronouns 40
 relative clauses 61–2
 sentences 394
 sound system 380–7
 suffixes 383–5
 syllable structure 386
 syntax 393–4
 umlauting 383
 verbal negation markers 57–8
 verbal phrases 42, 150–1
 verbs 390–3
 vowel assimilation 383–5
 vowel devoicing 382
 vowel raising 385
 vowels 30, 380–1

Uyghur (southeastern) branch of Turkic 82

Uzbek 83, 87, 357–78
 action clauses 60–1
 action nouns 46
 adjective phrases 50
 adjectives 39, 361
 adverbs 47–8, 369
 aorist 368
 case suffixes 39
 comparative case 39
 comparatives 57
 complement clauses 375
 conjunctions 48
 consonant assimilations 34, 359
 consonants 31, 357–8
 converb clauses 63, 375–7
 converbs 364–5
 coordination 374
 copulas 363
 declension 360–1
 demonstrative pronouns 40, 362
 dialects 378
 enclitic particles 37, 48
 equatives 57, 361

first actants 53
 gender 372
 genitive constructions 49
 indirective forms 45
 intensive forms 39
 interrogatives 373
 intraterminal participles 46
 intraterminals 43, 44, 367–8
 lexicon 37, 378
 literary language 2, 87
 modal forms 44, 373–4
 morphology 360–71
 morphophonology 359–60
 negation 373
 negative copulas 41
 nominal compounds 50
 nominal phrases 49, 371–2
 nouns 360–1
 numerals 363
 numerators 51
 ordinal numerals 41
 orthography 357–8, 360, 454
 passive 55
 people 2–3, 8–9, 27–8, 357
 personal markers 45
 phonetics/phonology 358–9
 pitch/stress accent 35
 possessive constructions 56
 possessive pronouns 40
 possessive suffixes 51
 post-terminal (perfect) participles 46
 post-terminals 44
 postpositions 48, 369–71
 postpositive clauses 65
 primary stems 32
 pronouns 361–3
 relative clauses 61–2, 374–5
 second actants 53
 sentences 372–7
 sound system 358–60
 superlatives 57
 syntax 371–7
 terminals 44
 third actants 53
 verbal phrases 42
 verbal stems 38
 verbs 363–9
 vowel notation xix
 vowels 30, 358

Uzbekistan 2, 4, 8–9, 10, 12, 28, 29
 language contacts 84

verbal morphology 41–7
 verbal negation 58
 verbal phrases 42
 verbal stems 37–8
 notation xxii
 verbals 41–9
 verbs
 finite forms 43–5
 non-finite forms 46–9
 Proto-Turkic 75
 Viryal dialect 451–2
 voice 41–3
 Proto-Turkic 75
 Volga Bulghar 7
 inscriptions 85
 people 23
 Volga Khanates 27
 Volga Turki 87
 vowels 30–1
 diachronic developments 89–95
 diachronic suffix phonology 107–9
 notations xviii–xix
 Proto-Turkic 69–71, 89–95, 107–9
 suffix notation xx–xxi

West Kipchak languages 301–16
see also Crimean Tatar; Karachay-Balkar; Karaim; Kumyk

West Middle Turkic 86
 West Old Turkic 79
 Western Thrace 4
 word classes 38–49
 indeclinable 47–9
 word formation 37–9
 Proto-Turkic 73
 word order
 discourse-pragmatic functions 57, 118
 word structure 35–7
 Proto-Turkic 72–3
 writing systems xviii, 1, 20–1, 126–37

Xinjiang 8, 10, 379–80, 394–6
 Xinjiang Uyghur Autonomous Region (Eastern Turkestan) 9, 10
 Xunhua Autonomous County 10

Yakut (Sakha) 83, 87, 88, 417–33
 accent 418
 adjective phrases 51
 adjectives 421–2
 aorist 427

- case suffixes 39
- case syntax 428–9
- categorical pasts 44
- classification 83
- collective pronouns 422
- comparative case 39
- consonants 418
- converb clauses 63
- converbs 427
- coordination 432
- declension 421–2
- demonstrative pronouns 40, 422–3
- dialects 432–3
- diathetic relations 432
- diphthongs 417–18, 420
- equatives 57
- exclamations 418
- genitive constructions 49
- indirective forms 45
- interrogation 423
- intraterminal converbs 47
- labial (rounded vs. unrounded)
 - harmony 33, 419
- lexicon 417, 432
- morphology 421–7
- morphophonology 419–20
- negation 431
- nominal compounds 50
- nominal phrases 427–9
- non-modifying converb clauses 64
- nouns 421–2
- numerals 423–4
- orthography 420–1
- people 11–12
- personal pronouns 422
- phonetics/phonology 417–18
- possessive pronouns 422, 423
- primary stems 32
- pronouns 422–3
- reflexive pronouns 40
- relative clauses 61, 428
- second actants 53
- sentences 430–2
- sound harmony 419
- sound system 417–21
- syllable structure 31
- syntax 427–32
- terminal converbs 47
- verbal phrases 429–30
- verbs 424–7
- vowels 30, 417–18
- Yakutia 11
- Yellow Uyghur 9–10, 83, 397–9
 - classification 83
 - consonants 398–9
 - first actants 52
 - lexicon 399
 - numerals 52, 74
 - possession marking 399
 - sound system 398–9
 - vowels 398
 - Yughur people 397
- Yenisey Turkic 83, 403–15
 - see also* South Siberian Turkic
- yes/no questions 57–8
- Yughur *see* Yellow Uyghur