seriously endangered by the UNESCO red book on endangered languages: Gagauz (Moldovan), Crimean Tatar, Noghay (Nogai), and West-Siberian Tatar

- Caucasian: Laz (a few hundred thousand speakers), Georgian (30,000 speakers), Abkhaz (10,000 speakers), Chechen-Ingush, Avar, Lak, Lezghian (it is unclear whether this is still spoken)
- Indo-European: Bulgarian, Domari, Albanian, French (a few thousand speakers each), Polish (a few dozen speakers), German (a few dozen speakers), Ukrainian (it is unclear whether this is still spoken), and these languages designated as seriously endangered by the UNESCO red book on endangered languages: Romani (20,000–30,000 speakers) and Yiddish (a few dozen speakers)
- Neo-Aramaic (Afrasiatic): Túrùyo and Súrit (a few thousand speakers each)
- Languages spoken by recent immigrants, refugees, and asylum seekers: Afroasiatic languages: Amharic, Somali, and Tigrigna; Niger-Congo: Lingala, Swahili, and various languages spoken in Nigeria; Indo-European: Russian and Farsi; and Altaic: Kazakh and Kirghiz (hundreds of thousands of speakers)
- Foreign languages taught in secondary schools: English, French, German, and Italian

See also: Arabic; Armenian; Azerbaijani; Caucasian Languages; Endangered Languages; Greek, Modern; Kurdish; Sign Language: Interpreting; Turkic Languages; Turkish.

Bibliography


Turkic Languages

L Johanson, Universität Mainz, Mainz, Germany
© 2006 Elsevier Ltd. All rights reserved.

Development and Classification

The Turkic language family was first attested in 8th century inscriptions. Turkic-speaking groups first appeared in the Inner Eurasian steppes, from where they moved to Central Asia, Eastern Europe, the Middle East, Siberia, etc. Because of their high mobility, Turkic expanded over a huge area.

The Proto-Turkic network of varieties was dissolved by an early split of Oghur or Bulgar Turkic. Its modern representative, Chuvash, a descendant of Volga Bulgar, differs from Common Turkic by specific phonetic representations, e.g., r and ž instead of z and s in words such as sér ‘hundred’ and sül ‘year’ (Turkish yüz ‘hundred,’ yaş ‘age’). A second split is represented by Khalaj, which retains a reflex of Proto-Turkic *p- as h-, e.g., hadaq ‘foot.’ Dialect splitting has led to further differentiation of Common Turkic. There is no mutual intelligibility throughout the family today. The following division combines the current areal distribution with genealogical and typological features.

1. The Southwestern or Oghuz branch contains a western subgroup comprising Turkish, Gagauz, and Azerbaijani (Azerbaijani, Northern and Azerbaijani, Southern), a southern subgroup comprising dialects of southern Iran and Afghanistan, and an eastern subgroup comprising Turkmen and Khorasan Turkic.

2. The Northwestern or Kipchak branch has a western subgroup comprising Kumyk, Karachay-Balkar, Crimean Tatar, and Karaim, a northern subgroup comprising Tatar and Bashkir, and a southern subgroup comprising Kazakh, Karakalpak, Kipchak Uzbek, Nogai, and Kirghiz (of different origin, but strongly influenced by Kazakh).

3. The Southeastern or Uyghur-Karluk branch has a western Uzbek subgroup and an eastern Uyghur subgroup.
4. The Northeastern or Siberian branch has a southern heterogeneous subgroup comprising Sayan Turkic (Tuvan, Tofan), Abakan (Yenisei) Turkic (Khakas, Shor), Chulym Turkic, Altai Turkic (Altai, Northern and Southern), and a northern subgroup comprising Yakut (Sakha) and Dolgan.

5. Chuvash is geographically situated in the northeastern area (Volga region).

6. Khalaj is geographically situated in the southwestern area (central Iran).

Deviant languages in China are Salar, of Oghuz origin, Yellow Uyghur (Yugur, West) and Fu-yu (Manchuria), both of south Siberian origin.

One traditional classificatory criterion is the final consonant of the word for ‘nine.’ Its representation as r in Chuvash tâxxrār separates Oghur from Common Turkic (Turkish dokuz). The intervocalic consonant in the word for ‘foot’ divides most Northeastern languages, Chuvash, Khalaj, etc. from the rest, which exhibits -y- (Turkish ayak), e.g., Tuvan adaq, Khakas azax, Chuvash ara. Oghuz Turkic differs from the rest by loss of suffix-initial velars, e.g., qal-an [remain-PART] instead of qal-yan [remain-PART] ‘remaining.’ Final -G is devoiced in the Southeast (Uyghur tâ-yi̯q [mountain-DER] ‘mountainous’), preserved in southern Siberia (Tuvan da-yi̯r [mountain-DER]), and lost elsewhere (Turkish da̱-li̱ [mountain-DER]).

Most older linguistic stages are insufficiently known. Written sources, where available, provide no direct information on spoken varieties. Early Oghuz and Bulgar (East Europe, 6th–7th centuries) are unknown. There are no texts in the language of the Khazars (7th–10th centuries). Pecheneg and Kuman, predecessors of West Kipchak, are only known from loanwords, titles, and names.

**Written Varieties**

Turkic literary varieties have emerged in various cultural centers. Many older Turkic empires, however, used foreign languages for administration (Sogdian, Persian). Muslim Turks often used Persian for poetry, and Arabic for religious and scientific writing. Russian has played an important role for poetry, and Arabic for religious and scientific (Sogdian, Persian) Muslim Turks often used Persian for administration, whereas Russian has played an important role for poetry.

The Turkic-speaking world in the 20th century was increasingly obstructed transregional linguistic contacts. A dozen ‘national’ languages with a narrow radius of validity emerged. In Turkey,
Ottoman was replaced by modern Turkish. The social importance of many Turkic languages was very limited. After the recent political developments, their significance is rapidly increasing, but the varieties spoken in Iran, Afghanistan, Iraq, etc., still have poor possibilities to develop.

Various scripts and script systems have been applied to Turkic. A specific runiform script was created for Old East Turkic. Most Old Uyghur texts are written in Uyghur script, originating in the Near East and later taken over by Mongols and Manchus. It is similar to the Sogdian script, which is also used in Buddhist texts. A few Buddhist manuscripts are written in Brahmi script, Manichaean texts in Manichaean script, and Nestorian texts in Syriac script. Arabic script was used for the languages of the Islamic era (still used in China for Uyghur and Kazakh).

A unified Roman-based script was introduced for several languages in the early Soviet period, but later replaced by different Cyrillic-based scripts. A Roman-based alphabet was introduced in Turkey in 1923. Most of the newly established Turkic republics have introduced or are introducing Roman-based scripts.

Contacts

The massive displacements of Turkic-speaking groups throughout their history have led to various phenomena induced by contacts with Iranian, Slavic, Mongolic, Uralic, etc. Speakers of Turkic have copied lexical, phonetic, morphological, and syntactic elements, whereas non-Turkic (e.g., Iranian, Greek, Finno-Ugric, Samoyedic, Yeniseian, Tungusic) groups shifting to Turkic have exerted substrate influence by copying native elements into their new varieties. Languages such as Chuvash, Yakut, Salar, Yellow Uyghur, Khalaj, Karaim, and Fu-yü have long developed in isolation from their relatives, preserving old features and acquiring new ones in their environments. Long and intense interaction with Iranian in Central Asia, Iran, Afghanistan, etc., has led to profound convergence phenomena. Massive foreign influence has sometimes caused considerable typological deviations, e.g., drastic structural changes in Karaim and Gagauz under Slavic impact.

Most written languages have been strongly influenced by Persian and Arabic. In Chaghatay (Chagatai) and Ottoman, lexical borrowing contributed to a remarkable richness of the vocabularies, whereas grammar was much less affected. The overload of Persian and Arabic in Ottoman led to strong puristic efforts in the 20th century to create a so-called Pure Turkish.

Internal convergence processes have resulted in leveling of languages of the central area. Several Turkic koinés have been used as transregional codes for trade and intergroup communication, e.g., Azerbaijani in Iran and the Caucasus region.

Linguistic Features

Despite their huge area of distribution, Turkic languages share essential phonological, morphological, and syntactic features.

They have a synthetic word structure with numerous highly applicable derivational and grammatical suffixes, and a juxtaposing technique with clear-cut morpheme boundaries and predictable allomorphs. These agglutinative principles yield considerable morphological regularity and transparency. Exceptions include traces of vowel gradation in the prenominal declension, e.g., Turkish ben ‘I,’ ban-a [I-DAT] ‘to me.’ The agglutinative structure is partly deranged in languages of the northeast and southeast. Some languages, e.g., Uzbek, even display borrowed prefixes.

The syllable contains minimally a vowel with maximally one preceding and one subsequent consonant. Vowel hiatus and consonant clusters are avoided.

Most languages exhibit eight short vowel phonemes, a, å, o, u, e, i, ö, ü, classified according to the features front vs. back, unrounded vs. rounded, and high vs. low. Proto-Turkic long vowel phonemes are preserved in Turkmen, Yakut, and Khalaj. Iranian and Slavic phonetic influence has sometimes affected the front vs. back distinctions. Tatar, Bashkir, Chuvash, and Uyghur exhibit systematic vowel shifts. Chuvash, Gagauz, Karaim, etc., have developed palatalized consonants, e.g., Karaim men ‘I.’ Tuvan and Tofan exhibit a glottal element signaling strong obstruents, e.g., apt ‘horse’ vs. at ‘name.’

The most general sound harmony phenomenon is an intrasyllabic front vs. back assimilation. An intersyllabic front vs. back harmony causes neutralization of the front vs. back distinction under the influence of the preceding syllable. If applied consistently, it excludes back and front syllables in a word, e.g., Turkish ev-ler-im-e [house-PL-POSS.1.SG-DAT] ‘to my houses,’ at-lar-im-a [horse-PL-POSS.1.SG-DAT] ‘to my horses.’ Some languages only display this kind of harmony, whereas others also apply a rounded vs. unrounded harmony, neutralization of the distinction rounded vs. unrounded in high suffix vowels, e.g., Turkish el-im [hand-POSS.1.SG] ‘my hand,’ gül-üm [rose-POSS.1.SG] ‘my rose.’ Languages such as Yakut and Kirghiz apply this harmony to low-vowel suffixes as well, e.g., börö-lör [wolf-PL] ‘wolves.’ There are numerous exceptions to harmony rules in loanwords. Further allomorphs are created by various consonant assimilations.
The rules of word accent vary. A high pitch accent, interacting with a dynamic stress accent, mostly falls on the last accentable syllable of native words.

The morphological structure has remained relatively stable through the centuries. The main word classes are nominals (nouns, adjectives, pronouns, numerals) and verbals. The primary stems can be used as free forms, e.g., at ‘horse,’ att! ‘throw!’ From verbal and nominal stems, which are sharply distinguished, expanded stems are formed. Nominals take plural, possessive, case, and specific derivational suffixes. Grammatical gender is not marked. The verbal morphology comprises markers of actionality, voice, possibility, negation, aspect, mood, evidentiality, tense, person, interrogation, etc. Voice is expressed by passive, reflexive-middle, causative, and cooperative-reciprocal suffixes. The order and combinability of suffixes is basically common to all Turkic languages.

Constructions with postposed auxiliary verbs (post-verbs) express actional modifications. A few constructions have developed into aspect-tense categories, e.g., Turkish gel-iyor [come-PRES] ‘comes’ < *gel-e yori-r [come-CONV run-AOR] ‘runs coming’. Possibility markers are formed with auxiliary verbs such as bil- ‘to know’ and al- ‘to take,’ e.g., Kirghiz ber-e al-[give-CONV AUX.POTEN] ‘to be able to give.’

Turkish languages share many syntactic characteristics. With respect to relational typology, they adhere to the nominative-accusative pattern. They have a head-final constituent order, with dependents preceding their heads. The unmarked order of clause constituents is subject + object + predicate (SOV). Adjectival, genitival, and participial attributes precede the head of the nominal phrase. Postpositions are used instead of prepositions. There is no agreement in number or case between dependents and heads. The focus position is in front of the predicate core. The unmarked constituent order is often deviated from for discourse-pragmatic reasons. Contact-induced word order changes are common, e.g., in Gagauz, which has become an SVO language.

Preposed subordinate clauses are based on verbal nouns, participles, and conversbs. The use of postposed subordinative patterns with conjunctions are typical effects of Iranian and Slavic influence. Most languages possess conjunctions, even coordinative ones meaning ‘and,’ ‘or,’ and ‘but’ of Persian, Arabic, or Russian origin.

Turkish lacks definite articles. The indefinite article is formally identical with the numeral ‘one’ Genitival attributes, expressing a possessor, stand in the genitive, whereas their head, indicating a possessed entity, carries a possessive suffix, e.g., Turkish at-in baş-i [horse-GEN head-POSS.3.SG] ‘the head of the horse.’ The dominant type of nominal compounds follows the pattern noun + noun + possessive suffix, e.g., Turkish el çanta-si [hand bag-POSS.3.SG] ‘handbag.’

All Turkic varieties exhibit numerous loanwords. Arabic and Persian loans are frequent in all Islamic-Turkic languages. The Iranian influence is strong in Uyghur, Uzbek, and varieties of Iran and Afghanistan. Many languages have been subject to considerable Mongolic and Slavic influence. Loans and calques from European languages have become increasingly important. The Turkic languages spoken in China exhibit old and recent loans from Chinese.

**Bibliography**

Oghuz Turkic
Old Anatolian Turkish
Old Uyghur
Ottoman Turkish
Pecheneg
Sakha
Salar
Samoyedic
Sayan Turkic
Shor
Slavic
Sogdian
Tofan
Tungusic
Turkic languages
Tuvin
Uralic
Uyghur-Karluk
Volga Bulgar
Yakut
Yellow Uyghur
Yenisei Turkic
Yeniseian