5 Transcription and Notation Methods

Onno A. Crasborn

Chapter Overview

This chapter discusses various methods of notating sign forms, focusing on the manual aspect of signing, for which the largest number of different systems has been proposed. Rather than entering into great detail on the advantages and disadvantages of the various orthographic notation methods and phonetic transcription systems that have been proposed throughout the past forty years, the chapter discusses how notation and transcription can aid present-day researchers that make use of multimodal annotation tools to directly time-link text to videos. Sign language corpora that are currently being created include between 20 and 300 hours of video, which are impossible to fully notate or transcribe for most research groups, let alone individual researchers. For this

Research Methods in Sign Language Studies: A Practical Guide, First Edition.
Edited by Eleni Orfanidou, Bencie Woll, and Gary Morgan.

reason, the chapter discusses various ways in which a basic annotation of a corpus in terms of glosses and sentence-level translations can be enriched with representations of aspects of the form of signs using notation methods that have been developed for other purposes.

Introduction

As Trevor Johnston emphasizes in recounting his experiences with the creation of a corpus for Auslan (the sign language used in Australia), a corpus is not a corpus unless it is machine-readable (Johnston, 2010). Despite rapid progress in computer vision technology in the past decade, we cannot directly search in video recordings of sign language. Some type of transcription is therefore required to allow for searching for aspects of the form of the language. It is this new context of the creation of annotated video resources that will be the focus of this chapter on notation and transcription. While modern technologies cannot replace manual transcription or annotation at this point in time, we are on the brink of letting computer vision techniques assist us in annotation. This chapter aims to describe present methodologies in (an)notating sign language data, while keeping an eye out on immanent changes in the way linguists work in the creation of sign resources.

Throughout the chapter I will use the term "notation" to refer to any type of text rendition, reserving "transcription" for phonetic transcription. "Annotation" will be used as a technical term to refer to a notation that is directly linked either to an audio or video source (as in the ELAN annotation tool) or to a text string. Thus a gloss annotation in ELAN is an annotation that refers to a certain time window in the video recording, while further properties of that gloss, such as meaning in context or precise phonetic form, can be added as child annotations pertaining to that gloss annotation itself. Phonetic transcription can thus be one type of annotation of a video source. Similarly, in text corpora, part-of-speech tags can be added to each word in the text, forming annotations that are not directly time-locked to audio or video. Where in the literature on text corpora "tag" and "annotation" are sometimes treated as equivalent, both being an attribute or a comment added to a piece of text, in the annotation of audio and video resources the initial layer of labeling an event in the source media file is also called annotation. "Orthographic notation" is used to refer to a notation system for everyday use, whether by language users or by linguists. (For further discussion of these different terms, see Johnston, 2010; Thompson, 2010; Baker, Hardie, and McEnerny, 2006.)

In the past decade there have been a considerable number of overviews of different notation systems for sign languages, from Miller (1994, 2001) to van der Hulst and Channon (2010) and to Frishberg, Hoiting, and Slobin (2012). Other publications have zoomed in on the transcription of child language data of sign languages (Morgan, 2005; Baker, van den Bogaerde, and Woll, 2005; Takkinen, 2005; Chen Pichler, Hochgesang, Lillo-Martin, and Quadros, 2010). Rather than trying to bundle all the details from these various overviews, this chapter aims to synthesize the issues involved in a broad outline, distinguishing orthographic and phonetic

notation from other types of annotations and indicating in which situations each of these might be of use. It then tries to determine how phonetic transcription and other types of notation may be used in the present context of the availability of large sign language corpora to researchers and the general public, and it includes practical suggestions and references to studies that have followed certain methodologies. The chapter concludes by evaluating whether there is in fact a need for phonetic transcription in sign language corpora and by discussing the expected impact of computer vision technology on our research methodologies in the near future. Taken together, all the parts of this chapter aim to provide the reader with a solid understanding of different kinds and levels of sign notation and with suggestions for the use of sign notation in relation to developing sign language corpus methodologies, linking the latter both to more traditional approaches and to expected future technologies.

Sign Orthographies and Glossing

Sign languages can be characterized as unwritten languages, in the sense that there are no Deaf[1] communities that have a written tradition. Sign orthographies for daily use – not by researchers, but by ordinary language users – have been designed, however, and there are enthusiastic proponents for each of them. While these orthographies may be used by small subcommunities or in certain educational settings, it is fair to say that the role of the orthography of the surrounding spoken language(s) is more prominent in the lives of literate deaf people.

SignWriting is probably the best known sign orthography, and was designed in the 1970s on the basis of Sutton Movement Notation, originally developed for notating various types of movements, including dance (for a historical overview of the system, see Sutton, 1999; Bianchini and Borgia, 2012; for a discussion of various similar movement notation systems, Farnell, 1996). SignWriting aims to represent not only manual behavior but also non-manual information. Like the various phonetic transcription systems discussed in the next section, SignWriting is based on the phonological analyses of ASL by Stokoe and others in the 1960s and 1970s: the phonological parameters handshape, orientation, location, and movement are all represented by individual symbols (graphemes). In addition, several features of the face (states of the eyebrows and mouth, gaze direction) have their own representation. Rather than putting symbols in a string, like in most phonetic transcriptions, SignWriting puts graphs together to form an abstract frontal view of a signer. SignWriting is not restricted to the word level: it allows for the writing of sentences and texts by ordering symbols vertically.

Other similar efforts have been made in the past; so is the typographically much more elegant SignFont (Newkirk, 1989; for a discussion, see Frishberg et al., 2012). More recently, a new orthography has been proposed for American Sign Language (ASL) under the name of "si5s" (Clark, 2012). This system appears to be especially suited for handwriting, requiring very few lines to represent signs.

Perhaps because of the limited role that orthographic notation plays in the language communities themselves, researchers wanting to notate sign language have

rarely used orthographic notation in collecting their research data or in presenting examples in publications (Miller, 2001; Frishberg et al., 2012). Rather, manual lexical items signs in sign language have been represented in text by means of glosses in a spoken language, be that the surrounding language used by signers of the sign language under investigation or the language of the international publication (often English). Thus, where the representation of examples of spoken Dutch in an English language publication would take the form of Dutch glosses using Dutch orthographic conventions with aligned English translations, the text representation of examples from Sign Language of the Netherlands would take the form of glosses in English. As the Latin script used for Dutch is based on phonetics (like many orthographies – but unlike the ideographic or semagraphic character orthography of Chinese, for instance), some information about the form of the Dutch example is also offered with a gloss. Sign language glosses, by contrast, do not contain information about the form of the sign and are necessarily based on a short translation to a spoken language. Use of glosses can therefore be deceptive, as it suggests that the semantics of the spoken language word or phrase chosen for the gloss overlaps with the semantics of the sign (Frishberg, 1983; Johnston, 2008; Frishberg et al., 2012). Such a full overlap might occur in some cases, but certainly not in all. Most signs therefore have multiple translation variants in a spoken language, but only one variant can be chosen as the gloss.

Johnston (2008) has introduced the term "ID-gloss" for a gloss that uniquely identifies a sign. This term emphasizes that a spoken language gloss for a sign should be interpreted as no more than a reference to a lexicon entry: it is not a full translation that captures all of its semantics, nor necessarily the best translation for a sign in a given context. ID-glosses could therefore just as well be numbers that refer to numbered items in a lexicon. Such a lexicon could be a traditional user's dictionary of a sign language, whether printed or electronic; but, in the context of sign language corpora, machine-readable lexicons are more likely to be scientific databases addressing the needs of the users of corpora. Spoken language words are chosen for various reasons and purposes, including to facilitate the quick reading of strings of glosses. Re-synthesis of lexical items on the basis of phonological or phonetic information in a lexicon by an avatar would help researchers to stay clear of spoken language words in glossing (Hanke, 2002). However, as with glosses for spoken languages, such a procedure would give easy access only to the form, but not to the meaning. The meaning of a sign in context has to be annotated separately, or the prototypical meaning from a lexicon needs to be made available together with the notation of the lexical item.

If one uses glosses in a publication to represent a series of signs, one typically chooses the translation of a sign that fits best, as a gloss, with that context. However, if the glosses in the example are intended as ID-glosses that refer to a lexicon or to a data collection that has been glossed in a systematic way by reference to that lexicon, the conventions of the lexicon must be followed.

The selection of glosses that function as ID-glosses in a lexicon will typically involve choosing the most frequent or the most neutral translation. This may not always be easy, as often multiple signs would all be translated by the same word if one chooses the most frequent meaning. For instance, regional sign variants referring to the same concept would ideally all be translated by the same word. In such a case, letter or number suffixes can be used to differentiate between signs, while one would

still use the same word (Johnston, 2008; Crasborn and de Meijer, 2012). The compound created by the word and the suffix then functions as the ID-gloss.

To represent the semantics of a sign in a lexicon, researchers typically fall back on a spoken language as well, simply because it can be easily written and read by researchers and/or other users like language learners. For this purpose, English would be the best choice for sign language data collections that are accessible to an international audience of researchers. The gloss annotations in the initial open access release of the NGT corpus (Sign Language of the Netherlands; see Crasborn, Zwitserlood, and Ros, 2008) are in Dutch, and for most researchers they will therefore function as a mere number, providing little to no useful semantic information. In a local non-English context, other written languages can fulfill the same function: the Dutch glosses of the Corpus NGT may be useful to Dutch people, for example. Sign lexicons have thus always been bilingual to a certain extent, at least until now, but there is no principled reason why there could not be a fully monolingual sign lexicon. Ideally lexicons, and thus ID-glosses, would be multilingual. Currently ELAN is being modified such that it will allow for the display of multilingual annotations and metadata categories. Future releases of the Corpus NGT annotations will exploit that technology so as to become accessible to a broader international audience.

Creating an orthographic notation or using glosses can also be useful when presenting examples to readers in documents – from student assignments to scientific journals. Such aids are imperative when creating sign resources for one's own research or for publication and later use by others. Word-level annotations (together with sentence-level translations) typically form the backbone of corpus annotation; other annotations such as intonational non-manual activities are related to them and can be interpreted in relation to them.

Phonetic Transcription

Various systems have been proposed to describe the phonological and/or phonetic form of sign languages. The Stokoe Notation (Stokoe, 1960) was the first to create a graphic componential representation of signs; and it inspired many others. The ASL dictionary of Stokoe, Casterline, and Croneberg (1965) used this notation method to represent the forms of signs without presenting drawings or photographs to illustrate the signs themselves. There have also been systems that are not based on Stokoe's parameter analysis of lexical items. Miller (2001) and van der Hulst and Channon (2010) discuss these notation methods in relation to parameter-based systems, and Miller (2001, p. 16) gives a good overview of their relations. Systems differ in the extent to which they are intended to be language-specific or universal and in the level of granularity or detail that they allow, including more versus less detail of the form of signs. It will thus depend both on the language one needs to transcribe and on the level of detail whether or not one can use a given system.

A series of recent publications by Johnson and Liddell (2010, 2011a, 2011b, 2012) propose a very fine-grained transcription system of the manual component of signing, one that allows for the transcription of the position of each of the fingers in quite some detail and also explicitly includes the transcription of the transitional movements

between signs – something that has not been addressed by any other notation system. Most systems make the explicit claim that such transitional movements are not linguistic units but rather phonetic epiphenomena (see Ormel and Crasborn, 2014 for a discussion). Yet, if these transitions are of interest for a specific phonetic study, the Johnson and Liddell system will prove useful. As Takkinen (2005) notes, working with an early version of the system, full transcriptions of every handshape alone can be very time-consuming and are very complex. This makes them good for searching – just as a database coding system like the ones proposed by van der Hulst and Channon (2010) or by Crasborn, van der Hulst, and van der Kooij (2001) – but impractical for reading and quickly reproducing a sign. While perhaps impractical for the transcription of longer strings, a very detailed system like Johnson and Liddell's could serve the purpose of providing standards for the coding of subaspects of manual sign forms for specific studies, where for instance only the position of the thumb is relevant. Two examples of such methodologies will be discussed later in this section.

There is a specific problem with the Stokoe Notation and its successors that, in order to represent handshapes, use letters also found in the hand alphabets; this makes them hard to use consistently for other sign languages. A letter such as "A" does not refer to the same handshape across sign languages. Iconic graphic symbols as in the HamNoSys transcription system (see Prillwitz and Zienert, 1990; and visit http://www.sign-lang.uni-hamburg.de/hamnosys) do not have that problem. Another problem that is shared among many transcription systems is that they have been developed in the context of lexicography, allowing for the sorting and searching of sign forms in dictionaries. When used for connected discourse, most transcription systems fail to accurately represent things like precise spatial locations and directions, holds of the non-dominant hand, and mouthings that may have their own time alignment.

For facial expression, some people have overcome the shortcomings of sign-specific systems by using the very detailed Facial Action Coding System (FACS), designed by psychologists to transcribe facial expressions showing emotions (Ekman and Friesen, 1969). Like any narrow phonetic transcription system, FACS has a steep learning curve and is not easily applied to long stretches of discourse simply because of the time needed for annotation. Ekman himself estimated that annotating the facial activities of signers may take as much as 2,400 times real time (Cohn and Ekman, 2005). Several researchers have productively used FACS in the analysis of intonational facial features in sign language (Baker-Shenk, 1983; Dachkovsky and Sandler, 2009; de Vos, van der Kooij, and Crasborn, 2009).

Until now, HamNoSys has been the only candidate for the sign language variant of the International Phonetic Alphabet (IPA) for the transcription of speech, especially if we focus on manual lexical items.

HamNoSys has been carefully designed for multilingual use over many years; like the IPA, it can be used for both broad (phonemic) and narrow (detailed) transcription, and it has a very elegant font for use on multiple computer platforms. Yet it is not so widely used as to be actually considered to have the status of a standard like the IPA. Many different factors may be involved in explaining why this is the case. First of all, there are some practical issues. HamNoSys takes quite some time to learn, given the large number of symbols needed for even a broad transcription of lexical forms. This problem is strengthened by the combination of a large number of frequently used symbols with the use of an ordinary keyboard, which requires the

user either to memorize a lot of keyboard shortcuts or to use special input software. Also, the transcription of sign forms sometimes leads to very long strings of symbols (see the example sentence transcribed in various ways in Frishberg et al., 2012). The limited documentation available in English for HamNoSys and its tools may play a role as well in the restricted use of this system in the literature until now. In terms of linguistic considerations, it may be the case that use of ordinary citation forms without any type of (spatial, non-manual) modification is rare in any study of sign grammar, from morphosyntax to discourse. For detailed aspects of a broad range of non-manual behavior and simultaneous constructions, HamNoSys may not (yet) be well equipped.

However, it could also be the case that, in actual practice, despite the convincing arguments in favor of phonetic transcription put forward by Miller and others, the use of glossing and/or illustration by images is sufficient for many purposes. This holds especially for the presentation of sentences as illustrations in scientific publications. It is true, as Miller (2001) notes, that a transcription helps us to focus on certain properties of the signal and disregard the wide variety of other things going on; but one may wonder how big an issue this is if we can provide immediate access to the video source in combination with specific annotations on various tiers, in order to focus our attention both on categories of events and on the timing relations between events. In present-day electronic journal publications, links to video examples stored in archives can be added, and this provides direct access to the actual utterance with its annotations (Broeder, van Uytvanck, and Senft, 2012). Such "enriched publications" presenting source data can thus include examples; moreover, they allow readers to double-check that transcriptions or other annotations have been correctly produced.

Indeed the requirements that Miller (2001) listed for a universal transcription system appear to be largely fulfilled through the use of multi-tier annotation documents time-aligned to the source video, which include any type of systematic verbal description of what can be observed. Miller's requirements included the need to record properties such as multilinearity of signing (asynchronies and overlaps of different articulators, for instance the two hands or the hands and non-manual behavior), non-manual articulations such as mouthings, manual rhythm such as holdings of the hands in space, and spatial location and direction. Each of these can be annotated at separate levels in multimedia annotation tools like ELAN (http://tla.mpi.nl/tools/tla-tools/elan), iLex (http://www.sign-lang.uni-hamburg.de/ilex), SignStream (http://www.bu.edu/asllrp/signstream), or ANVIL (http://www.anvil-software.org), which makes them easily searchable. In the context of corpus construction (Crasborn et al., 2007; Johnston, 2010), full phonetic transcriptions are simply not realistic: even glossing in annotation software can take as much as 200 times real time to do consistently – assuming there is already a full lexicon with ID-glosses available for reference. If the lexicon has to be created or enriched with new entries as annotation takes place, the time needed for glossing will be even longer. Phonetic transcription of speech may also take as much as 150 times real time (Schultze-Berndt, 2006). Ideally, glosses refer to a lexicon that stores the phonological or phonetic properties of the citation form, whether in a transcription system like HamNoSys or in other alphanumeric codes, as in databases like SignPhon (Crasborn et al., 2001) or SignTyp (van der Hulst and Channon, 2010), so that some information on the form is available through the mediation of a gloss notation.

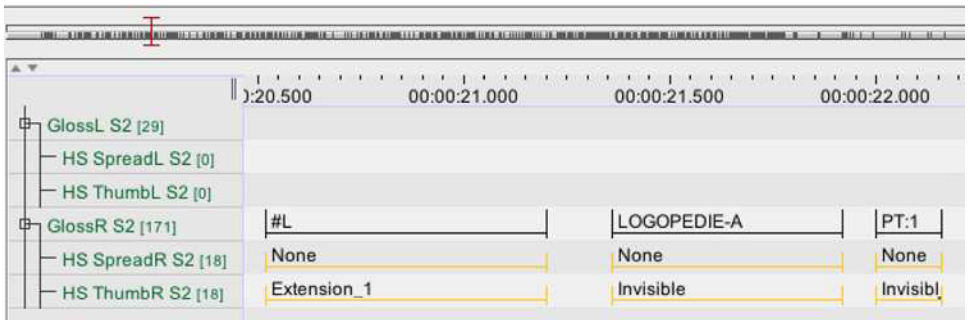

Figure 5.1 Phonetic coding appended to glosses in ELAN, for transcription of thumb position and finger spreading. "Extension_1" refers to full extension of the thumb. "Invisible" is used to indicate that the information cannot be seen on the available video because of the camera angle that was used.

Rather than creating full phonetic transcriptions for corpora, one can single out phonetic properties that are relevant for a particular study and annotate only those on specific tiers. This will speed up the process of data collection considerably in comparison to creating rich phonetic transcriptions of all data first and then extracting information from the transcriptions. Where possible, one can still follow the conventions for a transcription system like HamNoSys. As an example, take the methodology that was followed by Ormel et al. (2014). For a study of the co-articulation of handshape, they selected signs with flat handshapes from a published sign language corpus, adding a tier for the specific phonetic property of interest (precise position of the thumb) to the existing ELAN annotation documents of the corpus. In this way, rather than creating a full phonetic transcription, they transcribed only the phonetic property of interest; and this allowed for the collection of a large number of data points. Figure 5.1 illustrates how this phonetic information is appended to the gloss annotation by means of a dependent ("child") tier to the gloss tier in ELAN.

A slightly different yet related phonetic methodology, specific to the use of corpora, was followed by Russell, Wilkinson, and Janzen (2011), who investigated hand location in a corpus of ASL. This study aimed to find evidence for the phonetic undershoot of hand location. In a large corpus of conversational signing, specific signs were annotated with glosses. Within the time domain of that gloss, the frame was annotated where the active articulator was closest to the passive articulator (the place of articulation). This frame was then exported to an image file, and the location of the hand in terms of pixel coordinates was manually determined and further analyzed by using separate software. One could well imagine that, in the future, such functionality will be built into annotation tools, and that the resulting coordinate values will be stored as annotations, for further processing. Using this workflow, one is able to obtain quantitative information on the precise location of the hand that would be very hard to transcribe reliably with any transcription system.

As the two studies above show, there are ways of collecting corpus data that do not rely on full phonetic transcriptions of signs. Specific elements of signs are selected depending on the goal of the study, and specific coding schemes or measurement guidelines are developed that in some cases build on distinctions made in traditional phonetic transcription systems, yet in other cases (as the two studies illustrated)

create data collection strategies that are more efficient for a specific study. The downside of such specific strategies is that they may not lend themselves easily to reuse for other phonetic or phonological studies.

Other Types of Annotation

As some discussions of notation systems have pointed at before (Miller, 2001; Morgan, 2005; Takkinen, 2005; Frishberg et al., 2012), there are certainly properties of signed languages that are not easy to capture with current notation systems, for instance the precise spatial localization of signs and simultaneous constructions of the two hands, but also more gestural and more variable productions – like manual pantomimic gestures or emotional facial expressions. However, there may also be properties of the form of signs that are only relevant for specific studies and that one would not necessarily include explicitly in a transcription system. The minute details of the location of the hand in the study of Russell et al. (2011) are a clear case in point. Another example concerns the exact points of dominance reversal in connected signing: where exactly do signers switch from a right-dominant production to a left-dominant production? This may be implicit in some notation systems (like SignWriting or HamNoSys), where the left and the right hand can be notated as such, but it is not something that can be expressed with the Stokoe Notation, for instance. For an ongoing study of dominance reversal in NGT, we have therefore created a separate set of tiers and categories to mark various properties of hand selection (Crasborn and Sáfár, forthcoming). Examples include cases where a production moves from left-dominant to right-dominant (and vice versa) and where a turn starts with the left or the right hand as dominant. These are properties of the form of signing, annotated on a separate tier in our ELAN annotation documents. In the same vein, one could spell out various aspects of the phonetics of signing that are specific to a given study without having need for a full phonetic transcription of that data set.

Slobin et al. (2001) and Hoiting and Slobin (2002) present the Berkeley Transcription System (BTS), which does not target phonetic form but rather morphological structure. While morphological properties are sometimes included in gloss notations, these glosses are rarely fully componential, so as to distinguish the various parts of polycomponential signs like classifier constructions, for instance. The BTS aims to do just that. In part, labels for morphological categories can also be copied from glossing conventions for spoken languages, as when using ".pl" after a noun gloss to indicate a plural morpheme (see, for example, the Leipzig Glossing Conventions, on which see Comrie, Haspelmath, and Bickel, 2008). For many other properties of signing, including ones where a morphemic analysis has not yet been made and more gesture-like components are involved, conventions have yet to be developed. If one wants to make a semantic or morphological analysis of signing, investigating for instance how plurality is expressed in nominal and verbal signs, then the extra time needed for BTS transcription as compared to ID-glossing may well pay off. How much time such transcriptions will actually take to create is not made explicit in the literature.

Both the glossing of whole signs and morpheme-by-morpheme annotation provide a basic layer of annotations in a sign language corpus that will serve as an entry point for researchers with a large variety of questions. In addition to this sign-level access to a corpus, sentence-level translations are often created that likewise serve as a basic layer of annotations that will facilitate the use of any corpus by any researcher. For researchers unfamiliar with the language, translations are of course needed to understand the signing. However, also for researches that are fluent or native users of the language, translations may serve various functions. Text views of sequences of sentences allow one to scan through the content of a conversation much more rapidly than would be achievable by skimming through the video source. If translations are made of rigidly identified syntactic domains, annotations containing a translation further provide the exact sentence boundaries. These can then be used for investigating the position of glossed signs within sentences, for example. Together with gloss annotations, translations therefore provide a first segmentation of a larger text, allowing researchers to locate the basic units (words and sentences); and, in the case of multiple discourse participants, sentence annotations may also inform the researcher about who is signing and about how turns in a conversation overlap (or not).

Standardization of Annotations

A standard transcription system is seen as desirable by many researchers. While direct access to the source video may reduce the need for standard phonetic transcriptions, the need for conventions on corpus annotations becomes larger – and for the very same reason: we need to be able to "read" each other's data (Schembri and Crasborn, 2010). While "reading" will less and less refer to humans reading phonetic transcriptions and more and more to computer-readability of larger data collections, some properties of and requirements for standards are shared. From a technical point of view, the chosen font (in the case of a notation system) or document (in the case of video annotations) needs to be readable to others. Ideally, the data should be available and accessible regardless of the computer platform one is using.

Second, the annotation conventions that are chosen should be properly documented and published (see Chen Pichler et al., 2010 and Johnston, 2014 for examples).

Creating annotation standards and publishing them is not sufficient for the creation of resources that can then be used by other researchers, however. We need to be sure that the standards have actually been interpreted correctly in every instance. In other words, what is the validity of the transcriptions that are generated? Inter- and intra-transcriber reliability can be calculated, but actual access to the source videos should be possible as well. There are fewer and fewer technological barriers to publishing videos, and researchers can use open source tools like ELAN to inspect others' annotations together with the source media. Aside from checking the validity of what is annotated, researchers can also check in the source media whether certain events are not left un-annotated. Decisions on what counts as an "event" to be transcribed are not self-evident, and, no matter how good and explicit

the transcription guidelines or how standardized the categories, people will always differ somewhat in the interpretation of what is seen in the video.

With the publication of the annotation guidelines of the first published sign corpora (Johnston, 2014) and with the exchange of methodological knowledge at workshops (Crasborn, 2010), it would appear that the annotation of sign corpora is heading toward informal standardization on a number of issues, even if there is disagreement on whether and how to phonetically transcribe whole strings of signing. The current drive toward standardization in the spoken language resources world may contribute tools for this process, such as the ISOcat data category registry (Crasborn and Windhouwer, 2012). In this registry, linguistic categories of any kind can be defined, allowing annotators to be very explicit about what they mean for instance by "noun" or "handshape" by including references to these definitions in the controlled vocabularies that are related to tiers in ELAN. This explicitness and consistent referencing for specific annotation values through unique and persistent identifiers may promote standardization by avoiding ambiguity and by sharing categories and their definitions. As translations can be created in any (written) language, the same data categories can be easily reused for any sign language corpus, also if English is not the primary language of annotation.

An Outlook: Open Data and Computer-Assisted Annotation

While Miller (2001, p. 11), in his discussion of the "need for a common sign notation," acknowledges that at the time of writing (that is, around the year 2000) the use of video in the computer, and even on the Internet, was becoming more and more commonplace, he still assumed that we would always need a sign notation. Notation is useful because it provides focus, as opposed to raw video. One could debate whether this is still true, now that access to specific segments of the video on the basis of searchable annotations is possible (e.g. using ANNEX – for which visit http://tla.mpi.nl/tools/tla-tools/annex). As Miller also discusses, creating a phonetic transcription inevitably includes making theory-based decisions on what counts as an important and distinct category, both in designing a notation system and in actually transcribing video recordings. Even in the citation of examples in journal articles and e-books, perhaps we should no longer strive to use any type of phonetic transcription, but rather present the original text in its source format, which was available to the researcher: the video recording. This would allow readers to see not only an instantiation of the argument of the authors, but also the example that is brought forward in its context; thus it would also allow for an analysis of the form that is broader than just the elements highlighted by the authors. This is where the current movement toward "open data" could substantially contribute to our linguistic understanding of sign languages. The need for a common phonetic transcription method would become smaller.

Finally, developments in computer vision are proceeding rapidly, in line with the explosion of photo and video materials online, not to speak of the increase in CCTV recordings. Before long, the manual transcription of sign forms will be assisted by a

first parse of computer vision algorithms for specific features, for instance in detecting marked movements of the eye brows or head (see Piater, Hoyoux, and Du, 2010 for a recent example of the state of the art). As in automatic speech recognition, such methods will not automatically lead to full and fully correct phonetic transcriptions of whole sentences, but at least they will be able to assist in the annotation of specific features. This will increase the amount of data that are available to linguistic studies and will have an impact on how we design our research. The need for a shared phonetic transcription system may well be much smaller than we thought until only a decade ago. Although we may not see a true sign language version of the IPA, we will also forego on the disadvantages of the IPA, which transcribes some events while neglecting others – from sentence prosody to facial expression.

Conclusion

In summary, nowadays there are more economic ways of transcribing data for research purposes than relying on full phonetic transcriptions. The size of the available data sets in larger sign language corpora calls for something more efficient than the transcription of connected discourse. Again, this is often made possible by the availability of ID-glosses as a basic entry point to the data collection. The corresponding entry in a machine-readable lexicon associated to the corpus contains phonological information on the citation form. Phonetic properties such as the precise location of the hand in space or the state of the thumb can be added as properties of a gloss annotation to a sign, depending on the methodology for answering specific research questions. The selection of relevant items for studies that call for phonetic annotations can be most efficiently made once sign-level annotations have been made before.

Acknowledgments

Writing of this chapter was supported by grants from the Netherlands Organisation for Scientific Research (NWO 276-70-012) and from the European Research Council (ERC 210373).

Note

1 In this chapter the capitalized form *Deaf* is used only when it refers to social entities, not to individuals.

Keywords

corpus annotation; corpus linguistics; ELAN; glossing; notation; orthography; phonetic transcription; transcription; writing systems

See Also

Chapter 5; Chapter 9; Chapter 10; Chapter 11

Suggested Readings

Fischer, R. (1994). The notation of sign languages: Bebian's Mimographie. In H. Bos and T. Schermer (Eds.), *Sign Language Research 1994*. Hamburg: Signum Press, pp. 285–302.

Johnston, T. (2010). Adding value to, and extracting of value from, a signed language corpus through secondary processing: Implications for annotation schemas and corpus creation. In P. Dreuw, E. Efthimiou, T. Hanke, T. Johnston, G. Ruiz Martinéz, and A. Schembri (Eds.), *Proceedings of the fourth workshop on the representation and processing of sign languages: Corpora and sign language technologies*. Paris: ELRA, pp. 137–142.

Miller, C. (2006). Sign language: Transcription, notation, and writing. In K. Brown (Ed.), *Encyclopedia of language and linguistics*, vol. 8. Oxford: Elsevier, pp. 353–354.

Millet, A., and Estève, I. (2010). Transcribing and annotating multimodality: How deaf childrens' productions call into the question the analytical tools. *Gesture* 10(2/3), 297–320.

Ochs, E. (1979). Transcription as theory. In E. Ochs and B. B. Schieffelin (Eds.), *Developmental pragmatics*. New York: Academic Press, pp. 43–72.

Thoutenhoofd, E. (1992). Trans-scribing and reading: What constitutes a writing system? *Signpost* 4(2), 39–51.

References

Baker, A., van den Bogaerde, B., and Woll, B. (2005). Methods and procedures in sign language acquisition studies. *Sign Language & Linguistics* 8, 7–59.

Baker, P., Hardie, A., and McEnerny, T. (2006). *A glossary of corpus linguistics*. Edinburgh: Edinburg University Press.

Baker-Shenk, C. (1983). A microanalysis of the nonmanual components of questions in American Sign Language. PhD Dissertation, University of California, Berkeley.

Bianchini, C. S., and Borgia, F. (2012). Writing sign languages: Analysis of the evolution of the SignWriting system from 1995 to 2010, and proposals for future developments. Paper presented at the International Jubilee Congress "Science and education for the future," Varna, Bulgaria.

Broeder, D., van Uytvanck, D., and Senft, G. (2012). Citing on-line language resources. In *Proceedings of language resources and evaluation 2012*. Paris: European Language Resources Association, pp. 1391–1394.

Chen Pichler, D., Hochgesang, J. A., Lillo-Martin, D., and Müller de Quadros, R. (2010). Conventions for sign and speech transcription of child bimodal bilingual corpora in ELAN. *Language, Interaction and Acquisition* 1(1), 11–40.

Clark, A. (2012). *How to write American Sign Language*. Burnsville, MN: ASLwrite.

Cohn, J. F., and Ekman, P. (2005). Measuring facial action. In J. A. Harrigan, R. Rosenthal, and K. Scherer (Eds.), *The new handbook of methods in nonverbal behavior research*. New York: Oxford University Press, pp. 9–64.

Comrie, B., Haspelmath, M., and Bickel, B. (2008). *The Leipzig glossing rules: Conventions for interlinear morpheme-by-morpheme glosses*. Leipzig: Max Planck Institute for Evolutionary Anthropology, Department of Linguistics.

Crasborn, O. (2010). *The sign linguistics corpora network: Towards standards for signed language resources*. In *Proceedings of the seventh conference on international language*

resources and evaluation (LREC 2010). Paris: European Language Resources Association, pp. 457–460.

Crasborn, O., and & de Meijer, A. (2012). From corpus to lexicon: The creation of ID-glosses for the Corpus NGT. In O. Crasborn, E. Efthimiou, E. Fotinea, T. Hanke, J. Hedegaard Kristoffersen, and J. Mesch (Eds.), *Interactions between corpus and lexicon: Proceedings of the fifth workshop on the representation and processing of sign languages*. Paris: ELRA, pp. 13–17.

Crasborn, O., and Sáfár, A. (forthcoming). Annotation of hand dominance in the Corpus NGT. In M. Steinbach, R. Pfau, and A. Herrmann (Eds.), *Complex sentences and beyond*. Berlin: Mouton De Gruyter.

Crasborn, O., and Windhouwer, M. (2012). ISOcat data categories for signed language resources. In E. Efthimiou, G. Kouroupetroglou, and S.-E. Fotinea (Eds.), *Gesture and sign language in human–computer interaction and embodied communication*. Berlin: Springer, pp. 118–128.

Crasborn, O., van der Hulst, H., and van der Kooij, E. (2001). SignPhon: A phonological database for sign languages. *Sign Language & Linguistics* 4(1/2), 215–228.

Crasborn, O., Zwitserlood, I., and Ros, J. (2008). The corpus NGT: An open access digital corpus of movies with annotations of Sign Language of the Netherlands. Centre for Language Studies, Radboud University Nijmegen. Accessed September 6, 2014. http://www.ru.nl/corpusngtuk

Crasborn, O., Mesch, J., Waters, D., Nonhebel, A., van der Kooij, E., Woll, B., and Bergman, B. (2007). Sharing sign language data online: Experiences from the ECHO project. *International Journal of Corpus Linguistics* 12(4), 535–562.

Dachkovsky, S., and Sandler, W. (2009). Visual intonation in the prosody of a sign language. *Language and Speech* 52(2/3), 287–314.

Ekman, P., and Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* 1(1), 49–98.

Farnell, B. (1996). Movement notation systems. In P. Daniels and W. Bright (Eds.), *The world's writing systems*. New York: Oxford University Press, pp. 855–879.

Frishberg, N. (1983). Writing systems and problems for sign language notation. *Journal for the Anthropological Study of Human Movement* 2(4), 169–195.

Frishberg, N., Hoiting, N., and Slobin, D. I. (2012). Transcription. In R. Pfau, M. Steinbach, and B. Woll (Eds.), *Sign language: An international handbook*. Berlin: Mouton De Gruyter, pp. 1045–1075.

Hanke, T. (2002). HamNoSys in sign language generation context. In R. Schulmeister and H. Reinitzer (Eds.), *Progress in sign language research*. Hamburg: Signum Press, pp. 249–266.

Hoiting, N., and Slobin, D. I. (2002). Transcription as a tool for understanding: The Berkeley Transcription System for sign language research (BTS). In G. Morgan and B. Woll (Eds.), *Directions in sign language acquisition*. Amsterdam: John Benjamins, pp. 55–75.

van der Hulst, H., and Channon, R. (2010). Notation systems. In D. Brentari (Ed.), *Sign Languages*. Cambridge: Cambridge University Press, pp. 151–172.

Johnson, R. E., and Liddell, S. K. (2010). Toward a phonetic representation of signs: Sequentiality and contrast. *Sign Language Studies* 11(2), 241–274.

Johnson, R. E., and Liddell, S. K. (2011a). Toward a phonetic representation of hand configuration: The fingers. *Sign Language Studies* 12(1), 5–45.

Johnson, R. E., and Liddell, S. K. (2011b). A segmental framework for representing signs phonetically. *Sign Language Studies* 11(3), 408–463.

Johnson, R. E., and Liddell, S. K. (2012). Toward a phonetic representation of hand configuration: The thumb. *Sign Language Studies* 12(2), 316–333.

Johnston, T. (2008). Corpus linguistics and signed languages: No lemmata, no corpus. In O. Crasborn, E. Efthimiou, T. Hanke, E. Thoutenhoofd, D., and I. Zwitserlood (Eds.),

Fifth workshop on the representation and processing of signed languages: Construction and exploitation of sign language corpora. Paris: ELRA, pp. 82–87.

Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics* 15(1), 104–129.

Johnston, T. (2014). Auslan corpus annotation guidelines. Manuscript, Macquarie University, Sydney. Accessed October 29, 2014. http://new.auslan.org.au/about/annotations/

Miller, C. (1994). A note on notation. *Signpost* 7(3), 191–202.

Miller, C. (2001). Some reflections on the need for a common sign notation. *Sign Language & Linguistics* 4(1/2), 11–28.

Morgan, G. (2005). Transcription of child sign language: A focus on narrative. *Sign Language & Linguistics* 8, 119–130.

Newkirk, D. (1989). *SignFont: Handbook*. Bellevue, WA: Edmark Corporation.

Ormel, E., and Crasborn, O. (2014). Kinematic properties of lexical and transitional movements in Sign Language of the Netherlands. Manuscript under review.

Ormel, E., Crasborn, O., and de Meijer, A. (2014). Coarticulation of handshape in Sign Language of the Netherlands: A corpus study. Manuscript under review.

Piater, J., Hoyoux, T., and Du, W. (2010). Video analysis for continuous sign language recognition. Paper presented at the fourth workshop on the representation and processing of sign languages "Corpora and sign language technologies," 22–23 May, Valletta, Malta.

Prillwitz, S., and Zienert, H. (1990). Hamburg notation system for sign language: Development of a sign writing with computer applications. In S. Prillwitz and T. Vollhaber (Eds.), *Current trends in European sign language research*, vol. 9. Hamburg: Signum Press, pp. 355–379.

Russell, K., Wilkinson, E., and Janzen, T. (2011). ASL sign lowering as undershoot: A corpus study. *Laboratory Phonology* 2(2), 403–422.

Schembri, A., and Crasborn, O. (2010). Issues in creating annotation standards for sign language description. In In P. Dreuw, E. Efthimiou, T. Hanke, T. Johnston, G. Ruiz Martinéz, and A. Schembri (Eds.), *Proceedings of the fourth workshop on the representation and processing of sign languages: Corpora and sign language technologies*. Paris: ELRA, pp. 212–216.

Schultze-Berndt, E. (2006). Linguistic annotation. In J. Gippert, N. Himmelmann, and U. Mosel (Eds.), *Essentials of language documentation*. Berlin: Mouton de Gruyter, pp. 213–251.

Slobin, D. I., Hoiting, N., Anthony, M., Biederman, Y., Kuntze, L., Lindert, R., and Weinberg, A. (2001). The Berkeley Transcription System (BTS) for sign language. *Sign Language & Linguistics* 4(1/2), 63–96.

Stokoe, W. C. (1960). *Sign language structure: An outline of the visual communication systems of the American Deaf*. Buffalo, NY: Department of Anthropology and Linguistics, University of Buffalo.

Stokoe, W. C., Casterline, D. C., and Croneberg, C. G. (1965). *A dictionary of American Sign Language on linguistic principles*. Washington, DC: Gallaudet College Press.

Sutton, V. (1999). SignWriting: On the occasion of its 25th anniversary, November 1999. *Sign Language & Linguistics* 2(2), 271–282.

Takkinen, R. (2005). Some observations on the use of HamNoSys (Hamburg Notation System for Sign Languages) in the context of the phonetic transcription of children's signing. *Sign Language & Linguistics* 8, 99–118.

Thompson, P. (2010). Building a specialised audio-visual corpus. In A. O'Keeffe and M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. Abingdon: Routledge, pp. 93–103.

de Vos, C., van der Kooij, E., and Crasborn, O. (2009). Mixed signals: Combining linguistic and affective functions of eyebrows in questions in sign language of the Netherlands. *Language and Speech* 52(2/3), 315–339.