

## IX. Handling sign language data

### 42. Data collection

1. Introduction
2. Introspection
3. Data elicitation
4. Sign language corpus projects
5. Informant selection
6. Video-recording data
7. Conclusion
8. Literature

#### Abstract

*This chapter deals with data collection within the field of sign language research and focuses on the collection of sign language data for the purpose of linguistic – mainly grammatical – description. Various data collection techniques using both introspection and different types of elicitation materials are presented and it is shown how the selection of data can actually have an impact on the research results. As the use of corpora is an important recent development within the field of (sign) linguistics, a separate section is devoted to sign language corpora. Furthermore, two practical issues that are more or less modality-specific are discussed, i.e. the problem of informant selection and the more technical aspects of video-recording the data. It is concluded that in general, publications should contain sufficient information on data collection and informants in order to help the reader evaluate research findings, discussions, and conclusions.*

#### 1. Introduction

Sign language linguistics is a broad research field including several sub-disciplines, such as (both diachronic and synchronic) phonetics/phonology, morphology, syntax, semantics and pragmatics, sociolinguistics, lexicography, typology, and psycho- and neurolinguistics. Each sub-domain in turn comprises a wide range of research topics. For example, within sociolinguistics one can discern the linguistic study of language attitudes, bi- and multilingualism, standardisation, language variation and language change, etc. Furthermore, each sub-domain and research question may require specific types of data. Phonological and lexicographical research can focus on individual lexemes, but morphosyntactic research requires a different approach, using more extensive corpora, certain language production elicitation methods, and/or introspection. For discourse related research into turn-taking, on the other hand, a researcher would need to videotape dialogues or multi-party meetings. Even within one discipline, it is necessary to first decide on the research questions and then on which methodologies can be used

to find answers. Since it is not possible to deal with all aspects of linguistic research in this chapter, we have decided to focus on data collection for the purpose of linguistic description.

In general, (sign language) linguists claim to be using either qualitative or quantitative methodologies and regard these methodologies as two totally different (often incompatible) approaches. However, we prefer to talk about a continuum of research methodologies from qualitative to quantitative approaches rather than a dichotomy. At one end of the continuum, introspection can be found as the ultimate qualitative methodology (see section 2); at the other end, experimentation as the typically quantitative one is situated. To the best of our knowledge, the latter methodology has not been used in studies of linguistic description of sign languages. In between, there are mainly methods of observation and focused description on the basis of systematic elicitation (see section 3). Next, and newer to the field of sign language research, there are corpus-based studies where a (relatively) large corpus is mined for examples of structures and co-occurrences of items that then constitute the data for analysis (see section 4). When designing a study, it is also very important to think about the selection of the informants (see section 5) and to take into account the more technical aspects of data collection (see section 6).

## 2. Introspection

### 2.1. Value and limitations

According to Larsen-Freeman and Long (1991, 15) “[p]erhaps the ultimate qualitative study is an introspective one” in which subjects (often the researchers themselves) examine their own linguistic behaviour. In linguistics (including sign language linguistics) this methodology has frequently been used for investigating grammaticality judgments by means of tapping the intuitions of the “ideal native speaker” (Chomsky 1965). Schütze (1996, 2) gives some reasons why such an approach can be useful:

- Certain rare constructions are sometimes very hard to elicit and hardly ever occur in a corpus of texts. In this case, it is easier to present a native speaker with the construction studied and ask him/her about grammaticality and/or acceptability.
- A corpus of texts or elicited data cannot give negative information, that is, those data cannot tell the researcher that a certain construction is ungrammatical and/or unacceptable.
- Through tapping a native speaker’s intuitions, performance problems in spontaneous speech, such as slips of the tongue or incomplete utterances, can be weeded out.

At the same time, Schütze (1996, 3–6) acknowledges that introspection as a methodology has also attracted a great deal of criticism:

- Since the elicitation situation is artificial, an informant’s behaviour can be entirely different from what s/he would normally do in everyday conversation.

- Linguistic elicitation as it has been done in the past few decades does not follow the procedures of psychological experimentation since the data gathering has been too informal. Sometimes researchers only use their own intuitions as data, but in Labov's terms: "Linguists cannot continue to produce theory and data at the same time" (1972, 199). Moreover "[b]eing a native speaker doesn't confer papal infallibility on one's intuitive judgments" (Raven McDavid, quoted in Paikeday 1985).
- Basically, grammaticality judgments are another type of performance. Although they supposedly tap linguistic competence, linguistic intuitions "are derived and rather artificial psycholinguistic phenomena which develop late in language acquisition [...] and are very dependent on explicit teaching and instruction" (Levelt et al. 1977, in Schütze 1996).

The last remark in particular is highly relevant for sign language linguistics, since many, if not most, native signers will not have received any explicit teaching and instruction in their own sign language when they were at school. This fact in combination with the scarcity or complete lack of codification of many sign languages and the atypical acquisition process of sign languages in many communities (which results in a wide variety of competencies in these communities) raises the question to what extent it is possible to tap the linguistic intuitions of native signers in depth (see also section 5 on informant selection).

Schütze himself proposes a modified approach in order to answer the above criticism. The central idea of his proposal is that one should investigate not only a single native speaker's linguistic intuitions, but rather those of a group of native speakers:

I argue [...] that there is much to be gained by applying the experimental methodology of social science to the gathering of grammaticality judgments, and that in the absence of such practices our data might well be suspect. Eliminating or controlling for confounding factors requires us to have some idea of what those factors might be, and such an understanding can only be gained by systematic study of the judgment process. Finally, I argue that by studying interspeaker variation rather than ignoring it (by treating only the majority dialect or one's own idiolect), one uncovers interesting facts. (Schütze 1996, 9)

Clearly, caution remains necessary. Pateman (1987, 100), for instance, argues that "it is clear and admitted that intuitions of grammaticality are liable to all kinds of interference 'on the way up' to the level at which they are given as responses to questions. In particular, they are liable to interference from social judgments of linguistic acceptability".

## 2.2. Techniques

Various techniques have been used to tap an informant's intuitions about the linguistic issue under scrutiny. Some of them will be discussed in what follows, but this is certainly not an exhaustive list.

### (i) Error recognition and correction

Error recognition is a fairly common task but has not been used all that frequently in sign language research. Here informants are presented with a number of utter-

ances and are asked to detect possible errors and to correct them if there are any. However, since many sign languages have not yet (or hardly) been codified, and since many sign language users have not been educated in their sign language, this may prove to be a difficult task for certain informants (see above). Therefore, caution is warranted here.

(ii) Grammaticality judgments

In this type of task, informants are presented with a number of utterances and are asked whether they would consider them grammatically correct and/or appropriate or not. If a negative reply is given, informants can be asked to correct the utterance as well. An example of this in sign language research is a task in which a participant who is presented with a number of classifier handshapes embedded in the same classifier construction is asked whether the particular classifier handshape is appropriate/acceptable in the context provided.

An extension of this task would be to vary certain aspects of the execution of a sign (for instance, the handshape, the location, the rate, the manner, the non-manual aspects, etc.) and to ask the informants what the consequences of the change(s) actually are (morphologically, semantically, etc.) rather than just asking them whether the modified production would still be grammatically correct and/or acceptable.

(iii) Semantic judgments

Informants can be asked what the exact meaning of a certain lexeme is and in which contexts it would typically occur or in which contexts and situations it would be considered appropriate. In sign language research, informants can also be asked whether a certain manual production would be considered a lexical, conventional sign, or whether it is rather a polycomponential construction.

(iv) Other judgment tasks

Informants can also be asked to evaluate whether certain utterances, lexemes, etc. are appropriate for a given discourse situation (for instance, with respect to politeness, style, genre, and/or register). Furthermore, they could be asked to introspect on the speech act force of a certain utterance. To our knowledge, this type of task has not been used all that frequently in sign language research. What has been done quite frequently in sign language research though, is checking back with informants by asking them to introspect on their own productions of certain elicited data and/or by asking (a group of) native signers to introspect on certain elicited data (and/or the researchers' analyses) (see also section 3.3).

### 3. Data elicitation

In the first part of this section, some examples of tasks which can be used for data elicitation will briefly be explained. A number of these have been used quite extensively by various researchers investigating different sign languages while others have been used far less frequently (cf. Hong et al. 2009). The discussion proceeds from tasks with less control exerted by the researcher to tasks with more control. The second part discusses certain decisions with respect to data collection and the impact these decisions can have on the results obtained. Finally, an integrated approach, in which various methodologies are used in sequence, is presented.

### 3.1. Elicitation techniques and materials

#### 3.1.1. Recording natural language use in its context

On the “tasks with less – or no – control” end of the continuum, one finds the recording of traditional narratives in their appropriate context. One could, for example, videotape the after-dinner speech of the Deaf club’s president at the New Year’s Eve party as an example of a quasi-literal oratorical style involving the appropriate adjustments for a large room and the formality of the occasion. A priest’s sermon would be a similar example. In studies of language acquisition, videotaping bathtime play is a good way to collect data from parent-child interaction (Slobin/Hoiting/Frishberg, personal communication). In this context, it is very important for the researcher to be aware of the “Observer’s Paradox”, as first mentioned by Labov in the late 1960s (e.g. Labov 1969, 1972). Labov argues that even if the observer is very careful not to influence the linguistic activity, the mere presence of an observer will have an impact on the participants, who are likely to produce utterances in a manner different from when the observer is not present.

#### 3.1.2. Free and guided composition

In free composition, the researcher merely provides the informant with a topic and asks him/her to talk about that topic. Again, there is little control although it is possible to use this task to elicit particular structures. An obvious example is to ask an informant about his/her past experiences, in order to get past time references in the data. An example of guided composition that has been used in sign language research, is when informants are asked to draw their own family tree and to talk about family relations in order to elicit kinship terms.

#### 3.1.3. Role play and simulation games

Role play and simulation games are tasks which can also easily be used to elicit particular grammatical structures. If one informant is told to assume the role of interviewer and another is the interviewee (a famous athlete, for instance), the elicited data are expected to contain many questions. Creative researchers can certainly invent other types of role play yielding different grammatical structures.

Simulation games are usually played on a larger scale (with more participants), but are less well-defined in that the players only get a prompt and have to improvise as the conversation progresses (for example, they have to simulate a meeting of the board of a Deaf club, a family birthday party). As such, the researcher does not have a lot of control, but this procedure can nevertheless yield focused data (to look at turn-taking or register variation, for instance).

#### 3.1.4. Communication games

Communication games have also been used in sign language research to elicit production data. An example is a game played by two people who are asked to look at

drawings which contain a number of (sometimes subtle) differences. The players cannot see each other's drawings and have to try to detect what exactly those differences are by asking questions. Other possibilities include popular guessing games like *I spy ...* or a game played between a group of people in which one participant thinks of a famous person and the others have to guess the identity of this person by asking yes/no questions (and other variants of this game).

### 3.1.5. Story retelling

In sign language research, story retelling is commonly used for data elicitation. Here we present four forms of story telling: (i) picture story retelling, (ii) film story retelling, (iii) the retelling of written stories, and (iv) the retelling of signed stories.

#### (i) Picture story retelling

In some picture story elicitation tasks, informants are presented with a picture story made up of drawings and are asked to describe the depicted events. Normally such picture stories do not contain any type of linguistic information, that is, there is no written language accompanying the pictures. The following stories have been quite widely used in sign language research:

##### *The Horse Story*

*The Horse Story* (Hickmann 2003) was originally used in spoken language acquisition research but has also been used to elicit sign language data from adult signers, especially but not exclusively with a view to crosslinguistic comparison, as well as in research on 'homesign' (see chapter 26). It is a rather short picture story made up of five drawings about a horse that wants to jump over a fence in order to be with a cow in an adjacent meadow. However, the horse hits the fence, hurts its leg, and falls down. A little bird has witnessed the scene and flies off to get a first-aid kit. This is then used by the cow to bandage up the horse's leg.

##### *The Snowman*

A longer picture story with a longer history of being used for the elicitation of sign language data is *The Snowman*, a children's book by Raymond Briggs, first published in 1978 and turned into an animated film in 1982. The story is about a boy who makes a snowman that comes to life the following night. A large part of the story deals with the boy showing the snowman appliances, toys, and other bric-a-brac in the boy's house, while they are trying to keep very quiet so as not to wake up the boy's parents. Then the boy and the snowman set out on a flight over the boy's town, over houses and large buildings, before arriving at the sea. While looking at the sea, the sun starts to rise and they have to return home. The next morning, the boy wakes up to find the snowman melted. This is the story as it appears in the book; the film has additional parts, including a snowmen's party and a meeting with Father Christmas and his reindeer.

*Frog, where are you*

Another wordless picture story often used to elicit sign (and spoken) language narratives is *Frog, Where Are You* by Mercer Mayer, published in 1969. This story is about a boy who keeps a frog captured in a jar. One night, however, the frog escapes and the boy, accompanied by his dog, goes looking for it in the woods. Before finding the frog, they go through various adventures.

## (ii) Film story retelling

Next to narratives elicited by means of drawings, there is also film story retelling: informants are shown animated cartoons or (part of) a film and are asked to re-tell what they have just seen. Usually, cartoons or films used to elicit sign language narratives contain no or little spoken or written language. Examples include clips taken from *The Simpsons*, *Wallace and Gromit*, *The Pink Panther*, and *Tweety Bird & Sylvester* cartoons as well as short episodes from *Die Sendung mit der Maus*, a German children's television series featuring a large personified mouse and a smaller personified elephant as the main protagonists. All of these animated cartoons were produced to be shown on television. There are also films that were made specifically for use in linguistic research. A well-known example is *The Pear Story*, a six-minute film developed by Wallace Chafe and his team in the mid-1970s to elicit narratives from speakers around the world (Chafe 1980). The film shows a man harvesting pears, which are stolen by a boy on a bike. The boy has some other adventures with other children, before the farmer discovers that his pears are missing. The film includes sound effects but no words. *The Pear Story* has also been used in sign language research.

## (iii) Retelling of written stories

There are some examples of signed narratives elicited by means of written stories. In the context of *Case Study 4: Sign Languages of the ECHO (European Cultural Heritage Online)* project, for example, stories from *Aesop's Fables* in written English, Swedish, and Dutch were used to elicit narratives in British Sign Language (BSL), Swedish Sign Language (SSL), and Sign Language of the Netherlands (NGT). Working with this type of translated texts can have two major drawbacks. First, the (morpho)syntax of the target language may be influenced by the source language, and second, one needs to make sure that informants have a (near-)native proficiency in both languages. At the same time, however, working with parallel corpora of translated texts can be interesting for other purposes, e.g. for translation studies.

## (iv) Retelling of signed stories

Some of the NGT-fables mentioned above were used as elicitation materials during more recent NGT-data collection sessions: signers were shown the signed fables and asked to retell them (Crasborn/Zwitserlood/Ros 2008). These signed stories can then again be used for analysis towards linguistic description.

## 3.1.6. Video clip description

In the 1970s, Supalla created elicitation materials designed to elicit polycomponential verbs of motion and location (Supalla 1982). The original materials, known as the *Verbs*

of *Motion Production Test* (VMP), include some 120 very short video clips showing objects moving in specific ways. Informants (American deaf children in the original Supalla (1982) study) are asked to watch the animated scenes and to describe the movement of the object shown in the clip. The VMP task can easily be used to study verbs of motion and location in other sign languages and/or produced by other groups of informants, although Schembri (2001, 156) notes that the task may be of less use with signers from non-Western cultures because the objects include items that may not be familiar to members of these cultures. There is a shorter version of the VMP task which consists of 80 coded items and five practice items. This version is included as one of twelve tasks in the *Test Battery for American Sign Language Morphology and Syntax* (Supalla et al., no date). Both the longer and the short version of the VMP task have been used in a number of studies on different sign languages and are still used today, for example, in the context of some of the corpus projects discussed in section 4.

A somewhat comparable set of stimuli are the *ECOM clips* from the Max Planck Institute for Psycholinguistics (Nijmegen): 74 animations showing geometrical entities that move and interact. These have also been used in sign language research, mainly to study classifier constructions. A set of stimuli consisting of 66 videotaped skits of approximately 3–10 seconds depicting real-life actors performing and undergoing certain actions was used to study indexicality of singular versus plural verbs in American Sign Language (Cormier 2002).

### 3.1.7. Picture description

Picture description may take the form of a question-and-answer session. Participants are asked to look at a picture or a series of pictures (or drawings) and then answer questions designed to elicit particular structures under study. This is a fairly common procedure in lexicographical research, but has also been used to target certain grammatical patterns. In such a question-and-answer session, there is linguistic interaction between the informant and at least one other person. In another task involving picture description, the signer describes a specific picture to an interlocutor who subsequently has to select the correct picture (i.e. the picture described by the signer) from a series of (almost identical) pictures or drawings. This elicitation procedure is often used to elicit specific forms or structures, e.g. plural forms, locative constructions, or classifier constructions. A well-known example in sign language research is the, by now classical, study on word order in Italian Sign Language, for which Volterra et al. (1984) designed elicitation materials. Since then, these materials have been used for the analysis of constituent order in declarative sentences in a number of other sign languages (Johnston et al. 2007).

In the Volterra et al. task, eighteen pairs of drawings with only one contrastive element (e.g. 'A cat is under a chair' versus 'A cat is on a chair') are used to elicit sentences describing three distinct states of affairs: six locative states of affairs (e.g. 'The tree is behind/in front of the house'), six non-reversible states of affairs (e.g. 'The boy/girl eats a piece of cake'), and six reversible states of affairs (e.g. 'The car is towing the truck/The truck is towing the car'). The person videotaped is a signer who has the drawings before him/her, and for each pair, one of the drawings is marked with an arrow. The interlocutor, another signer who is not being videotaped, has the same

drawings, but without arrows. The first signer is asked to sign one sentence describing the drawing marked with the arrow; the interlocutor is asked to indicate which of the two drawings of each pair is being described.

The main purpose of studies using this elicitation task has been to analyse whether the sign language under investigation exhibits systematic ordering of constituents in declarative utterances that contain two arguments, and if this is the case, to determine the patterns that occur.

A variant of the Volterra et al. task makes use of elicitation materials that consist of sets of pictures, e.g. four pictures, with only one deviant picture. The signer is asked to describe the picture that is different. This task may, for example, be used to elicit negative constructions, when the relevant picture differs from the others in that there is something missing.

### 3.1.8. Elicited translation

In elicited translation, the researcher provides the informant with an isolated utterance in one language (usually the surrounding spoken language, but it could also be another sign language) and asks the informant to translate the utterance into his/her own (sign) language. This procedure has been widely used in sign language research, especially in its early days, but has more recently been regarded with suspicion as it is faced with the risk of interference from the source language onto the target language. Consequently, (mostly morphosyntactic) linguistic descriptions of target sign language structures elicited by means of this method may be less valid.

A slightly less controlled form of elicited translation consists in presenting informants with verbs in a written language and asking them to produce a complete signed utterance containing the same verb. In order to further minimize possible interference from the written language, these utterances can subsequently be shown to another informant who is asked to copy the utterance. It is the final utterance which is then used for the analysis. This would be an example of elicited imitation (see next sub-section).

### 3.1.9. Elicited imitation

In elicited imitation, the researcher produces an utterance containing a certain linguistic structure and asks the informant to repeat what s/he has just produced. If the utterance is long enough, the informant will not be able to rely on working memory, but will have to rely on semantic and syntactic knowledge of the language. To our knowledge, this procedure has not been used in sign language research yet, but it could yield interesting results when executed correctly. One could imagine that this procedure might be used to study non-manuals, for instance.

### 3.1.10. Completion task

In a manner fairly similar to the previous one, informants are asked to complete an utterance started by the researcher. This type of task can be used to study plural

formation, for instance. The researcher signs something like “I have one daughter, but John has ...” (*three daughters*). As far as we know, this technique has only rarely been used in sign language research.

### 3.1.11. Structured exercises

In structured exercises, informants are asked to produce certain sentence structures in a predetermined manner. Informants can be presented with two clauses, for instance, and asked to turn them into one complex sentence (e.g. by means of embedding one of the clauses as a relative clause into the other), or can be asked to turn a positive utterance into a negative one. Again, this technique has been used in sign language research, but certainly not on a large scale.

## 3.2. Data selection and impact on results

The selection of data can, of course, have a major impact on research results. When examining the degree of similarity across the grammars of different sign languages, for instance, looking at elicited utterances produced in isolation may lead to a conclusion which is very different from the overall picture one would get when comparing narratives resulting from picture story descriptions. The latter type of data contains many instances where the signer decides to “tell by showing” (“dire en montrant”; Cuxac 2000), and it seems likely that the resulting prominence of visual imagery in the narratives – among other issues – yields more similarity across sign languages (see, for instance, Vermeerbergen (2006) for a more comprehensive account). In general, the strategy of ‘telling by showing’ is (far) less present in isolated declarative sentences, and it is in these constructions where we find more differences between different sign languages (Van Herreweghe/Vermeerbergen 2008).

The nature of the data one works with might also influence one’s opinion when it comes to deciding on how to approach the analysis of a sign language. Does one opt for a more ‘oral language compatible view’ or rather decide on a ‘sign language differential view’?

On the one hand, there is the oral language compatibility view. This presupposes that most of SL structure is in principle compatible with ordinary linguistic concepts. On the other hand, there is the SL differential view. This is based on the hypothesis that SL is so unique in structure that its description should not be primarily modelled on oral language analogies. (Karlsson 1984, 149f.)

Simplifying somewhat, it could be argued that the question whether ‘spoken language tools’, that is, theories, categories, terminology, etc. developed and used in spoken language research, are appropriate and/or sufficient for the analysis and description of sign languages will receive different answers depending on whether one analyzes the signed production of a deaf comedian or a corpus consisting of single sentences translated from a spoken language into a sign language. A similar observation can be made with regard to the relationship between the choice of data and the issue of sign lan-

languages as homogeneous systems or as basically heterogeneous systems in which meanings are conveyed using a combination of elements, including linguistic elements but also components traditionally regarded as not being linguistic in nature.

### 3.3. An integrated approach

When it comes to studying a certain aspect of the linguistic structure of a sign language, we would like to maintain that there is much to be gained from approaching the study by using a combination of the above-mentioned methodologies and techniques. An example of such an integrated line of research for the study of negatives and interrogatives in a particular sign language might include the following steps:

- Step 1: Making an inventory of the instances of negatives and interrogatives in a previously collected and transcribed corpus of monologues and dialogues in the sign language studied.
- Step 2: Eliciting more focused production data in which negatives and interrogatives can be expected, such as a role play between signers in which one informant takes the role of an interviewer (asking questions) and the other of the interviewee (giving negative replies), or by means of communication games.
- Step 3: Transcribing the data collected in step 2 and making an inventory of the negatives and interrogatives, followed by an analysis of these occurrences.
- Step 4: Checking the analysis against the intuitions of a group of (near-)native signers by means of introspection.
- Step 5: Designing a more controlled judgment study in which one group is confronted with (what the researchers think are) correct negative and interrogative constructions and another with (what the researchers think are) incorrect negatives and interrogatives.
- Step 6: Proposing a description of the characteristic properties of negatives and interrogatives in the sign language under scrutiny.

## 4. Sign language corpus projects

### 4.1. Why corpus linguistics?

Corpus linguistics is a fairly new branch of linguistic research which goes hand in hand with the possibilities offered by more and more advanced computer technology. In the past, any set of data on which a linguistic analysis was performed was called a ‘corpus’. However, with the advent of computer technology and corpus-based linguistics, use of the term ‘corpus’ has become more and more restricted to any type of collection of texts in a machine-readable form. Johnston (2009, 18) argues: “Corpus linguistics is based on the assumption that processing large amounts of annotated texts can reveal patterns of language use and structure not available to lay user intuitions or even to

expert detailed linguistic analyses of particular texts.” In corpus linguistics, “quantitative analysis goes hand in hand with qualitative analysis” (Leech 2000, 49) since

[e]mpirical linguists are interested in the actual phenomena of language, in the recordings of spoken and written texts. They apply a bottom-up procedure: from the analysis of individual citations, they infer generalizations that lead them to the formulation of abstractions. The categories they design help them understand differences: different text types, syntactic oppositions, variations of style, shades of meaning, etc. Their goal is to collect and shape the linguistic knowledge needed to make a text understandable. (Mahlberg 1996, iv)

The same obviously holds for sign language corpora. However, since they contain face-to-face interaction, they are more comparable to spoken language corpora than to written language corpora, and according to Leech (2000, 57),

[t]here are two different ways of designing a spoken corpus in order to achieve ‘representativeness’. One is to select recordings of speech to represent the various activity types, contexts, and genres into which spoken discourse can be classified. This may be called *genre-based* sampling. A second method is to sample across the population of the speech community one wishes to represent, in terms of sampling across variables such as region, gender, age, and socio-economic group, so as to represent a balanced cross-section of the population of the relevant speech community. This may be called a *demographic* sampling.

In sign language corpora, it is especially the latter type of sampling that has been done so far. Moreover, sign language corpora are similar to spoken language corpora (and not so much to written language corpora) since they are only machine-readable when transcriptions and annotations are included (for the transcription of sign language data, we refer the reader to chapter 43).

## 4.2. Sign language corpora

In sign language linguistics, corpus (at least in its more restricted sense of machine-readable corpus) linguistics is still in its infancy, although rapidly growing. Johnston (2008, 82) expresses the need for sign language corpora as follows:

Signed language corpora will vastly improve peer review of descriptions of signed languages and make possible, for the first time, a corpus-based approach to signed language analysis. Corpora are important for the testing of language hypotheses in all language research at all levels, from phonology through to discourse [...]. This is especially true of deaf signing communities which are also inevitably young minority language communities. Although introspection and observation can help develop hypotheses regarding language use and structure, because signed languages lack written forms and well developed community-wide standards, and have interrupted transmission and few native speakers, intuitions and researcher observations may fail in the absence of clear native signer consensus of phonological or grammatical typicality, markedness or acceptability. The past reliance on the intuitions of very few informants and isolated textual examples (which have remained essentially inaccessible to peer review) has been problematic in the field. Research into signed languages has grown dramatically over the past three to four decades but progress in the field has been hindered by the resulting obstacles to data sharing and processing.

One of the first (if not *the* first) large-scale sign language corpus projects is the corpus of American Sign Language (ASL) collected by Ceil Lucas, Robert Bayley, and their team (see, for instance, Lucas/Bayley/Valli 2001). In the course of 1995, they collected data in seven cities in the United States that were considered to be representative of the major areas of the country: Staunton, Virginia; Frederick, Maryland; Boston, Massachusetts; Olathe, Kansas/Kansas City, Missouri; New Orleans, Louisiana; Fremont, California; and Bellingham, Washington. All of these cities have thriving communities of ASL users and some also residential schools for deaf children and as such long-established Deaf communities. 207 African-American and white working and middle-class men and women participated in the project. They could be divided into three age groups: 15–25, 26–54, and 55 and up. All had either acquired ASL natively at home or had learned to sign in residential schools before the age of 5 or 6 (see Lucas/Bayley/Valli 2001). For each site, at least one contact person was asked to identify fluent, lifelong ASL users who had to have lived in the community for at least ten years. The contact persons, deaf themselves and living in the neighborhood, assembled groups of two to seven signers. At the sites where both white and African-American signers were interviewed, two contact persons were appointed, one for each community. All the data were collected in videotaped sessions that consisted of three parts. In the first part of each session, approximately one hour of free conversation among the members of each group was videotaped, without any of the researchers being present. In a second part, two participants were selected and interviewed in depth by the deaf researchers. The interviews included topics such as background, social network, and patterns of language use. Finally, 34 pictures were shown to the signers to elicit signs for the objects or actions represented in the pictures. It was considered to be very important not to have any hearing researcher present in any of the sessions: “It has been demonstrated that ASL signers tend to be very sensitive to the audiological and ethnic status of an interviewer [...]. This sensitivity may be manifested by rapid switching from ASL to Signed English or contact signing in the presence of a hearing person.” (Lucas/Bayley 2005, 48). Moreover, the African-American participants were interviewed by a deaf African-American research assistant, and during the group sessions with African-American participants, no white researchers were present. In total, data from 62 groups were collected at community centers, at schools for deaf children, in private homes, and at a public park. At the same time, a cataloguing system and a computer database were developed to also collect and store metadata, that is, details as to when and where each group was interviewed and personal information (name, age, educational background, occupation, pattern of language use, etc.). Furthermore, the database also contained details about phonological, lexical, morphological, and syntactic variation, and further observations about other linguistic features of ASL that are not necessarily related to variation. The analysis of this corpus has led to numerous publications about sociolinguistic variation in ASL (see chapter 33 on sociolinguistic variation).

Since this substantial ASL corpus project, for which the data were collected in 1995, sign language corpus projects have been initiated in other countries as well, including Australia, Ireland, The Netherlands, the United Kingdom, Germany, China (Hong Kong), Italy, Sweden, and France, and more are planned in other places. Some of these corpus projects also focus on sociolinguistic variation, but most have multiple goals, and the data to be obtained cannot only be used as data for linguistic description,

but also for the preservation of older sign language data for future research (i.e. the documentation of diachronic change) or as authentic materials to be used in sign language teaching. The reader can find up to date information with respect to these (and new) corpus projects at the following website: <http://www.signlanguagecorpora.org>.

### 4.3. Metadata

When collecting a corpus it is of the utmost importance to also collect and store metadata related to the linguistic data gathered. In many recent sign language corpus projects, the *IMDI metadata database* is being used, an already existing database which has been further developed in the context of the *ECHO project* at the Max Planck Institute for Psycholinguistics in Nijmegen (The Netherlands) (Crasborn/Hanke 2003; also see [www.mpi.nl/IMDI/](http://www.mpi.nl/IMDI/)). This approach is being increasingly used in smaller research projects as well. A good example is presented in Costello, Fernández, and Landa (2008, 84–85):

We video-record our informants in various situations and contexts, such as spontaneous conversations, controlled interviews and elicitation from stimulus material. Each recording session is logged in the IMDI database to ensure that all the related metadata are recorded. The metadata relate to the informant, for example:

- age, place of birth and sex
- hearing status, parents' hearing status, type of hearing aid used (if any)
- age of exposure to sign language
- place and context of sign language exposure
- primary language of communication within the family
- schooling (age, educational program, type of school)

and also to the specific context of the recording session, such as:

- type of communicative act (dialogue, storytelling, question and answer)
- degree of formality
- place and social context
- topic of the content.

Another important piece of information to include in the metadata is birth order of the informant and hearing status of siblings, if any. There are, for instance, clear differences between the youngest/oldest deaf person in a family with hearing parents and three older/younger deaf siblings and the youngest/oldest deaf person in a family with hearing parents and three older/younger hearing siblings.

## 5. Informant selection

Not all users of a specific language show the same level of language competence. This is probably true of all language communities and of all languages, but it is even more true of sign language communities. This is, of course, related to the fact that across the world, 90 to 95 percent (or more, cf. Johnston 2004) of deaf children are born to hearing parents, who are very unlikely to know the local sign language. Most often

deaf children only start acquiring a sign language when they start going to a deaf school. This may be early in life, but it may also be (very) late or even never, either because the deaf child's parents opt for a strictly oral education with no contact with a sign language or because the child does not go to school at all. Consequently, only a small minority of signers can be labelled "*mother tongue speaker*" in the strict sense of the word, and in most cases, these native signers' signing parents will not be/have been native signers themselves. When deaf parents are late learners of a sign language, for instance, when they did not learn to sign until they were in their teens, this may be reflected in their sign language skills, which may in turn have an effect on their children's sign language production.

In spoken language research, especially in the case of research on the linguistic structure of a given language, the object of study is considered to be present in its most natural state in the language production of a native speaker (but see section 2 above). When studying form and function of a specific grammatical mechanism or structure in a spoken language, it would indeed be very unusual to analyse the language production of non-native speakers and/or to ask non-native speakers to provide grammaticality judgments. The importance of native data has also been maintained for sign language research, but, as stated by Costello, Fernández, and Landa (2008, 78), "there is no single agreed-upon definition of native signer, and frequently no explanation at all is given when the term is used". The "safest option model of native signers" (Costello/Fernández/Landa 2008, 79) is the informant who is (at least) a second generation deaf-of-deaf signer. However, in small Deaf communities, such ideal informants may be very few in number. For example, Costello et al. themselves claim that they have not managed to find even seven second-generation signers in the sign language community of the Basque Country, a community estimated to include around 5,100 people. Johnston (2004, 370f.) mentions attempts to locate deaf children of deaf parents under the age of nine and claims that it was not possible to locate more than 50 across Australia. Especially in small communities where there is merely a handful of (possibly) native signers, researchers may be forced to go for the second best and decide to stipulate a number of criteria which informants who are not native signers must meet. Such criteria often include:

- early onset of sign language acquisition; often the age of three is mentioned here, but sometimes also six or seven;
- education in a school for the deaf, sometimes stipulating that this should be a residential school;
- daily use of the sign language under investigation (e.g. with a deaf signing partner and/or in a deaf working environment);
- prolonged membership of the Deaf community.

Note that it may actually be advisable to apply these criteria to native signers as well. At the same time, we would like to make two final comments:

- (1) In any community of sign language users, small or large, there are many more non-native signers than native signers. This means that native signers most often have non-native signers as their communication partners and this may affect their intuitions about language use. It may well be that a certain structure is over-used by

non-native signers so that that structure is seen as “typical” of or “normal” for the language, although it is not very prominent in the language production of native signers. One can even imagine that a structure (e.g. a certain constituent order) which results from the influence of the spoken majority language and is frequently used by non-native signers is characterized as “acceptable” by native signers even though the latter would not use this structure themselves, at least not when signing to another native language user.

- (2) If one wants to get an insight into the mechanisms of specific language practices within a certain sign language community (e.g. to train the receptive language skills of sign language interpreter students), it might be desirable in certain sign language communities not to restrict the linguistic analysis to the language use of third-generation native signers. Because non-native signers make up the vast majority of the language community, native signers are not necessarily “typical” representatives of that community.

Natural languages are known to show (sociolinguistic) variation. It seems that for sign languages, region and age are among the most important determining factors, although we feel it is safe to say that in most, if not all, sign languages the extent and nature of variation is not yet fully understood. Thus, variation is another issue that needs to be taken into account when selecting informants. Concerning regional variation in the lexicon of Flemish Sign Language (VGT), for example, research has shown that there are five variants, with the three most centrally located areas having more signs in common, compared to the two more peripheral provinces. Also, there seems to be an ongoing spontaneous standardization process with the most central regions “leading the dance” (Van Herreweghe/Vermeerbergen 2009). Therefore, in order to study a specific linguistic structure or mechanism in VGT, it is best to include data from all different regions. Whenever that is not possible, it is important to be very specific about the regional background of the informants because it may well be the case that the results of the analysis are valid for one region but not for another.

Finally, we would like to stress the necessity of taking into account the anthropological and socio-cultural dimensions of the community the informants belong to. When working with deaf informants, researchers need to be sensitive to the specific values and traditions of Deaf culture, which may at times be different from those of the surrounding culture. Furthermore, when the informants belong to a Deaf community set within a mainstream community that the researcher is not a member of, this may raise other issues that need to be taken into consideration (e.g. when selecting elicitation materials). A discussion of these and related complications, however, is beyond the scope of this chapter.

## 6. Video-recording data

### 6.1. Recording conditions

Research on sign languages shares many methodological issues with research on spoken languages but it also comprises issues of its own. The fact that data cannot be audio-recorded but need to be video-recorded is one of these sign language specific

challenges. Especially when recording data to study the structure of the language, but also when it comes to issues such as sociolinguistic research on variation, one of the major decisions a researcher needs to make is whether to opt for high quality recording or rather to try to minimize the impact of the data collection setting on the language production of the informants. It is a well-known fact that language users are influenced by the formality of the setting. Different situations may result in variations in style and register in the language production. This is equally true for speakers and signers, but in the latter group, the specific relationship between the sign language and the spoken language of the surrounding hearing community is an additional factor that needs to be taken into account. In many countries, sign languages are not yet seen as equal to spoken languages, but even if a sign language is recognized as a fully-fledged natural language, it is still a minority language used by a small group of language users surrounded by a much larger group of majority language speakers. As a result, in many Deaf communities, increased formality often results in increased influence from the spoken language (Deuchar 1984).

A problem related to this issue is the tendency to accommodate to the (hearing) interlocutor. This is often done by including a maximum of characteristics from the structure of the spoken language and/or by using structures and mechanisms that are supposedly more easily understood by people with poor(er) signing skills. For example, when a Flemish signer is engaged in the Volterra et al. elicitation task (see section 3.1.7) and needs to describe a picture of a tree in front of a house, s/he may decide to start the sentence with the two-handed lexical sign HOUSE followed by the sign TREE and a simultaneous combination of a ‘fragment buoy’ (Liddell 2003) referring to HOUSE on the non-dominant hand and a ‘classifier’ referring to the tree on the dominant hand, thereby representing the actual spatial arrangement of the referents involved by the spatial arrangement of both hands. Alternatively, s/he might describe the same picture using the three lexical signs TREE + IN-FRONT-OF + HOUSE in sequence, that is, in the sequential arrangement familiar to speakers of Dutch. In both cases, the result is a grammatically correct sentence in VGT, but whereas the first sentence involves sign language specific mechanisms, namely (manual) simultaneity and the use of space to express the spatial relationship between the two referents, the same is not true for the second sentence, where the relationship is expressed through the use of a preposition sign and word order, exactly as in the Dutch equivalent *De boom staat voor het huis* (‘the tree is in front of the house’). One way to overcome this problem in an empirical setting is by engaging native signers to act as conversational partners. However, because of the already mentioned specific relationship between a sign language and the majority spoken language, signers may still feel that they should use a more ‘spoken language compatible’ form of signing in a formal setting (also see the discussion of the ‘Observer’s Paradox’ in section 3.1).

Because of such issues, researchers may try and make the recording situation as informal and natural as possible. Ways of doing this include:

- organising the data collection in a place familiar to the signer (e.g. at home or in the local Deaf club);
- providing a deaf conversational partner: This can be someone unknown to the signer (e.g. a deaf researcher or research assistant, a deaf student), although the presence of a stranger (especially if it is a highly educated person) may in itself

have an impact on the language production of the informant. It may therefore be better to work with an interlocutor the signer knows, but at the same time, it should not be an interlocutor the signer is too closely related with (e.g. husband/wife or sibling) because this may result in a specific language use (known as ‘within-the-family-jargon’) which may not be representative of the language use in the larger linguistic community;

- avoiding the presence of hearing people whenever possible;
- only using one (small-size) camera and avoiding the use of additional recording equipment or lights;
- not using the first ten minutes of what has been videotaped; these first ten minutes can be devoted to general conversation to make sure that the signer is at ease and gradually forgets the presence of the camera.

## 6.2. Technical issues

In certain circumstances, for instance when compiling a corpus for pedagogical reasons, researchers may opt for maximal technical quality when recording sign language data. Factors that are known to increase the quality of a recording include the following:

- **Clothing:** White signers preferably wear dark, plain clothes and black signers light, plain clothes to make sure there is enough contrast between the hands and the background when signs are produced on or in front of the torso. Jewellery can be distracting. If the informant usually wears glasses, it may be necessary to ask him/her to take off the glasses in order to maximize the visibility of the non-manual activity (obviously, this is only possible when interaction with an interlocutor is not required).
- **Background:** The background can also influence the visibility of the signed utterances. Consequently, a simple, unpatterned background is a prerequisite, and frequently, a certain shade of blue or green is used. This is related to the use of the chroma key (a.k.a. bluescreen or greenscreen) technique, where two images are being mixed. The informant is recorded in front of a blue or a green background which is later replaced by another image so that the informant seems to be standing in front of the other background. If there is no intention to apply this technique, then there is no need for a blue or green background, simply “unpatterned” is good enough. However, visual distraction in the form of objects present in the signer’s vicinity should be avoided.
- **Posture:** When a signer sits down, this may result in a different dimension of the signing space as compared to the same signer standing upright (and this may be a very important factor in phonetic or phonological research, for instance).
- **Lighting:** There clearly needs to be enough foreground lighting. Light sources behind the signer should be avoided as much as possible since it results in low visibility of facial expressions. The presence of shadows should be avoided as much as possible.
- **Multiple cameras:** How many cameras are necessary, their position, and what they focus on will be determined by the specific research question(s); the analysis of non-manual activity, for example, requires the use of one camera zooming in on

the face of the informant(s) (although nowadays it is also possible to afterwards electronically zoom in on a selected area within the image).

- Position of the camera(s) in relation to the signer(s): In order to fully capture the horizontal dimension of the signed production, some researchers avoid full frontal recording and prefer a slight angle. A top view facilitates the analysis of the relationship of the hands and the body, which may be important when studying the use of space.
- Use of elicitation materials: The signer should not hold any papers or other things in his/her hands while signing and should not start to sign while (still) looking at the materials.

### 6.3. Issues of anonymity

One major disadvantage of the necessity to video-record sign language production is related to the issue of anonymity. When presenting or publishing their work, researchers may wish to illustrate their findings with sequences or stills taken from the video-recorded data. However, not all signers like the idea of their face being shown to a larger public. In the age of online publishing, this problem becomes even more serious. Obviously, making the signer unrecognisable, for instance, by blurring his/her face – a strategy commonly used to anonymise video-taped speakers – is not an option because important non-manual information expressed on the face will be lost. It may therefore be necessary to make use of a model reproducing the examples for the purpose of dissemination. This solution may be relatively easy for individual signs or for constructions to be reproduced in isolation but may be problematic in the case of longer stretches of language production. The problem of privacy protection is, of course, also highly relevant in the case of on-line publication of sign language video recordings and annotations. This issue cannot be further dealt with here, but we would like to refer to Crasborn (2008), who discusses developments in internet publishing of sign language data and related copyright and privacy issues.

The fact that sign language production needs to be video-recorded also has consequences in terms of research design. A well-known research design to study language attitudes is the “matched guise” technique developed by Lambert and colleagues (Lambert et al. 1960) to study attitudes towards English and French in Montreal, Canada. The visual nature of sign languages makes it difficult to apply this technique when studying sign language attitudes because it will soon be obvious that one and the same signer is producing two samples in two different languages or variants. Fenn (1992, in Burns/Matthews/Nolan-Conroy 2001, 189) attempted to overcome this by selecting physically similar signers, dressed in a similar fashion. However, he encountered another difficulty since many of his subjects recognized the signers presenting the language samples.

## 7. Conclusion

In this chapter, we have attempted to give a brief survey of data collection techniques using different types of elicitation materials and using corpora. We have also focused

on the importance of deciding which type of data should be used for which type of analysis. Furthermore, we have discussed the problem of informant selection and some more technical aspects of video-recording the data. Throughout the chapter, we have focused on data collection in the sense of collecting sign language data. Sign language research may also involve other types of data collection, such as questioning signers on matters related to sign language use or (sign) language attitudes. In this context, too, the sociolinguistic reality of Deaf communities may require a specific approach. Matthews (1996, in Burns/Matthews/Nolan-Conroy 2001, 188) describes how he and his team, because of a very poor response from deaf informants on postal questionnaires, decided to travel around Ireland to meet with members of the Deaf community face to face. They outlined the aims and objectives of their study (using Irish Sign Language) and presented informants with the possibility to complete the questionnaire on the spot, giving them the opportunity to provide their responses in Irish Sign Language (which were later translated into written English in the questionnaires). Thanks to this procedure, response rates were much higher.

Finally, we would also like to stress the need for including sufficient information on data collection and informants in publications in order to help the reader evaluate the research findings, discussion, and conclusions. It is quite customary to collect and provide metadata in the context of sociolinguistic research and it has become standard practice in the larger corpus projects as well, but we would like to encourage the collection of the above type of information for all linguistic studies, as we are convinced that this will vastly improve the comparability of studies dealing with different sign languages or sign language varieties.

## 8. Literature

- Briggs, Raymond  
1978 *The Snowman*. London: Random House.
- Burns, Sarah/Matthews, Patrick/Nolan-Conroy, Evelyn  
2001 Language Attitudes. In Lucas, Ceil (ed.), *The Sociolinguistics of Sign Languages*. Cambridge: Cambridge University Press, 181–216.
- Chafe, Wallace L.  
1980 *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex.
- Chomsky, Noam  
1965 *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.
- Cormier, Kearsy  
2002 *Grammaticization of Indexic Signs: How American Sign Language Expresses Numerosity*. PhD Dissertation, University of Texas at Austin.
- Costello, Brendan/Fernández, Javier/Landa, Alazne  
2008 The Non-(existent) Native Signer: Sign Language Research in a Small Deaf Population. In: Quadros, Ronice M. de (ed.), *Sign Languages: Spinning and Unraveling the Past, Present and Future. TISLR 9: Forty Five Papers and Three Posters from the 9<sup>th</sup> Theoretical Issues in Sign Language Research Conference, Florianopolis, Brazil, December 2006*. Petrópolis/RJ, Brazil: Editora Arara Azul, 77–94. [Available at: [www.editora-arara-azul.com.br/EstudosSurdos.php](http://www.editora-arara-azul.com.br/EstudosSurdos.php)]

- Crasborn, Onno  
2008 Open Access to Sign Language Corpora. Paper Presented at the 3<sup>rd</sup> Workshop on the Representation and Processing of Sign Languages (LREC), Marrakech, Morocco, May 2008 [<http://www.lrec-conf.org/proceedings/lrec2008>, 33–38].
- Crasborn, Onno/Hanke, Thomas  
2003 Additions to the IMDI Metadata Set for Sign Language Corpora. Agreements at an ECHO workshop, May 2003, Nijmegen University. [Available at: [http://www.let.kun.nl/sign-lang/echo/docs/SignMetadata\\_May2003.doc](http://www.let.kun.nl/sign-lang/echo/docs/SignMetadata_May2003.doc)]
- Crasborn, Onno/Zwitserslood, Inge/Ros, Johan  
2008 *Corpus NGT. An Open Access Digital Corpus of Movies with Annotations of Sign Language of the Netherlands*. Centre for Language Studies, Radboud University Nijmegen. [Available at: <http://www.ru.nl/corpusngt>]
- Cuxac, Christian  
2000 *La Langue des Signes Française. Les Voies de l'Iconicité* (Faits de Langues No 15–16). Paris: Ophrys.
- Deuchar, Margaret  
1984 *British Sign Language*. London: Routledge & Kegan Paul.
- Hickmann, Maya  
2003 *Children's Discourse: Person, Space and Time Across Languages*. Cambridge: Cambridge University Press.
- Hong, Sung-Eun/Hanke, Thomas/König, Susanne/Konrad, Reiner/Langer, Gabriele/Rathmann, Christian  
2009 Elicitation Materials and Their Use in Sign Language Linguistics. Poster Presented at the *Sign Language Corpora: Linguistic Issues Workshop*, London, July 2009.
- Johnston, Trevor  
2004 W(h)ither the Deaf Community? Population, Genetics, and the Future of Australian Sign Language. In: *American Annals of the Deaf* 148(5), 358–375.
- Johnston, Trevor  
2008 *Corpus Linguistics and Signed Languages: No Lemmata, No Corpus*. Paper Presented at the 3<sup>rd</sup> Workshop on the Representation and Processing of Sign Languages (LREC), Marrakech, Morocco, May 2008. [<http://www.lrec-conf.org/proceedings/lrec2008/>, 82–87]
- Johnston, Trevor  
2009 The Reluctant Oracle: Annotating a Sign Language Corpus for Answers to Questions We Can't Ask Any Other Way. Abstract of a Paper Presented at the *Sign Language Corpora: Linguistic Issues Workshop*, London, July 2009.
- Johnston, Trevor/Vermeerbergen, Myriam/Schembri, Adam/Leeson, Lorraine  
2007 “Real Data Are Messy”: Considering Cross-linguistic Analysis of Constituent Ordering in Auslan, VGT, and ISL. In: Perniss, Pamela/Pfau, Roland/Steinbach, Markus (eds.), *Visible Variation: Comparative Studies on Sign Language Structure*. Berlin: Mouton de Gruyter, 163–205.
- Karlsson, Fred  
1984 Structure and Iconicity in Sign Language. In: Loncke, Filip/Boyes-Braem, Penny/Lebrun, Yvan (eds.), *Recent Research on European Sign Languages*. Lisse: Swets and Zeitlinger, 149–155.
- Labov, William  
1969 Contraction, Deletion, and Inherent Variability of the English Copula. In: *Language* 45, 715–762.
- Labov, William  
1972 *Sociolinguistic Patterns*. Philadelphia, PA: University of Pennsylvania Press.

- Lambert, Wallace, E./Hodgson, Richard C./Gardner, Robert C./Fillenbaum, Samuel  
1960 Evaluational Reactions to Spoken Language. In: *Journal of Abnormal and Social Psychology* 60, 44–51.
- Larsen-Freeman, Diane/Long, Michael H.  
1991 *An Introduction to Second Language Acquisition Research*. London: Longman.
- Leech, Geoffrey  
2000 Same Grammar or Different Grammar? Contrasting Approaches to the Grammar of Spoken English Discourse. In: Sarangi, Srikant/Coulthard, Malcolm (eds.), *Discourse and Social Life*. Harlow: Pearson Education, 48–65.
- Liddell, Scott K.  
2003 *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press.
- Lucas, Ceil/Bayley, Robert  
2005 Variation in ASL: The Role of Grammatical Function. In: *Sign Language Studies* 6(1), 38–75.
- Lucas, Ceil/Bayley, Robert/Valli, Clayton  
2001 *Sociolinguistic Variation in American Sign Language*. Washington, DC: Gallaudet University Press.
- Mahlberg, Michaela  
1996 Editorial. In: *International Journal of Corpus Linguistics* 1(1), iii–x.
- Mayer, Mercer  
1969 *Frog, Where Are You?* New York: Dial Books for Young Readers.
- Paikeday, Thomas M.  
1985 *The Native Speaker Is Dead*. Toronto: Paikeday Publishing Inc.
- Pateman, Trevor  
1987 *Language in Mind and Language in Society: Studies in Linguistic Reproduction*. Oxford: Clarendon Press.
- Schembri, Adam  
2001 *Issues in the Analysis of Polycomponential Verbs in Australian Sign Language (Auslan)*. PhD Dissertation, University of Sydney, Australia.
- Schütze, Carson T.  
1996 *The Empirical Base of Linguistics. Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.
- Supalla, Ted  
1982 *Structure and Acquisition of Verbs of Motion and Location in American Sign Language*. PhD Dissertation, University of California, San Diego.
- Supalla, Ted/Newport, Elissa/Singleton, Jenny/Supalla, Sam/Metlay, Don/Coulter, Geoffrey  
no date *The Test Battery for American Sign Language Morphology and Syntax*. Manuscript, University of Rochester.
- Van Herreweghe, Mieke/Vermeerbergen, Myriam  
2008 Referent Tracking in Two Unrelated Sign Languages and in Home Sign Systems. Paper Presented at the Workshop “Gestures: A Comparison of Signed and Spoken Languages” at the 30<sup>th</sup> Annual Meeting of the German Linguistic Society (DGfS), Bamberg, February 2008.
- Van Herreweghe, Mieke/Vermeerbergen, Myriam  
2009 Flemish Sign Language Standardisation. In: *Current Issues in Language Planning* 10(3), 308–326.
- Vermeerbergen, Myriam  
2006 Past and Current Trends in Sign Language Research. In: *Language and Communication* 26(2), 168–192.

Volterra, Virginia/Corazza, Serena/Radutsky, Elena/Natale Francesco

- 1984 Italian Sign Language: The Order of Elements in the Declarative Sentence. In: Loncke, Filip/Boyes-Braem, Penny/Lebrun, Yvan (eds.), *Recent Research on European Sign Languages*. Lisse: Swets and Zeitlinger, 19–48.

*Mieke Van Herreweghe, Ghent (Belgium)*

*Myriam Vermeerbergen, Antwerp & Leuven (Belgium)*

## 43. Transcription

1. Introduction
2. Transcription at the level of phonology
3. Transcription at the level of morphology
4. Multimedia tools
5. Conclusion
6. Literature and web resources

### Abstract

*The international field of sign language linguistics is in need of standardized notation systems for both form and function. This chapter provides an overview of available means of notating components of manual signs, non-manual devices, and meaning. Attention is also paid to problems of representing simultaneous articulators of hands, face, and body. A final section provides an overview of several tools of multimedia analysis. Standardization, in the twenty-first century, requires attention to computer-based storage and processing of data; numerous links are provided to web-based facilities. Throughout, the chapter addresses theoretical problems of defining and relating linguistic levels of analysis in the study of sign languages.*

*“What is on a transcript will influence and constrain what generalizations emerge”.*

Elinor Ochs (1979, 45)

### 1. Introduction

Transcription serves a number of functions, such as linguistic analysis, pedagogy, providing deaf signers with a writing system, creating input to an animation program, and others. Because this chapter appears in a handbook of sign language linguistics, we limit ourselves to those notation systems that have played a role in developing and advancing our understanding of sign languages as linguistic systems. Although most notation schemes have been devised for the descriptive study of particular sign lan-