# 4 Collecting and Analyzing Sign Language Data: Video Requirements and Use of Annotation Software

*Pamela Perniss*



**Chapter Overview**

This chapter provides researchers with a resource guide to making and managing video recordings of sign language data and to undertaking linguistic annotation with available software. Instead of focusing on technical details, the chapter focuses on how to obtain video data that are good and usable in terms of quality of the recording and appropriate in terms of answering specific research questions. With respect to annotation, the chapter provides guidelines on devising a coding scheme for specific research questions and on setting up a workflow in order to implement the coding scheme in the annotation software for successful and effective use. The chapter is aimed at researchers with a background in linguistics interested in working with sign language video data.

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

## Introduction

Any researcher interested in answering questions about the linguistic structure and usage patterns of a sign language must obtain and analyze primary language sources, that is, examples of actual sign language use and production. As sign language is produced visually, using the hands, face, and body as articulators, data collection necessarily relies on video recording. Subsequent analysis of the collected material requires transcription and coding of the video stream, a process that has been greatly facilitated in recent years by the development of sophisticated annotation software. This chapter provides the aspiring sign language linguist with a resource guide to making video recordings of sign language and to undertaking linguistic annotation of the data with available tools.

The next section focuses on how to obtain good and usable video data. "Good and usable" is meant here both in terms of the quality of the recording and in terms of getting the kind of data that are appropriate to answering specific research questions about the sign language under consideration. The section does not focus heavily on specific technical aspects of video recording like the specific kind of equipment to use, since the speed of technological development will soon render any details set out here obsolete.

The third section deals with how to develop and implement a workflow for the annotation and coding of video data. Software that allows annotations to be directly linked to and time-aligned with the video stream has been a great boon to working with sign language data. Primary examples are ELAN[1] and Anvil.[2] In addition, software that allows the creation of a lexical database in concert with time-aligned annotation, for example iLEX[3] and LEXUS,[4] is valuable to sign language linguists interested in lexicography and corpus analysis. The exposition in this section turns predominantly around the use of ELAN, the program most familiar to the author. Again, this section does not concentrate on technical details, nor does it go into the details of transcription methods. It offers the researcher guidelines on how to make decisions about what kind of annotation is needed for answering particular research questions and on how to best set up a workflow in order to carry out successful annotation. For ELAN, this will involve the elaboration of a number of key features, as well as some step-by-step instructions for specific things that personal experience has shown to be critical; these instructions will ultimately make annotating and coding sign language data more efficient and more enjoyable. Other annotation software is discussed at the end of this section, where differences between ELAN and other programs are highlighted.

The final section concludes the chapter by delineating its contribution from what has already been published on the topic of sign and spoken language video data collection and analysis. In addition, the conclusion stresses the overall importance of using video data to annotate and research language, in both the signed and spoken modalities, in order to understand the complex multimodal phenomenon of human communication and interaction.

## Collecting Sign Language Data

Obtaining good video data of the sign language you wish to study is important for a number of reasons. These pertain, on the one hand, to the technical quality of the video and, on the other hand, to the content of the video and the type of data collected. In order to be able to work with video data as a record of language use, the conditions for filming need to be appropriate and the quality of the image needs to be good. Subsequent to filming, it must be ensured that the video is in a format that is suitable for further management and analysis. However, before actually obtaining the video data, there are important decisions to be made regarding the type of data to be collected. This is in the first instance a matter of deciding what kind of data – say, spontaneous narrative or stimulus-based elicitation – are best suited to answering a specific research question. We deal with this issue first, before moving on to the more technical aspects of data collection.

### *Type of data to collect*

Two main types of language data are distinguished here: naturalistic data and elicited data. Both types are valuable and important in their own right, and both have advantages and disadvantages. Either way, it is important to make careful decisions about the what, the where, and the how of data collection. Data collection is rarely something that can be redone under the same circumstances. For example, if you are dependent on traveling to a specific – and possibly distant – field site, you may not have the time and resources to return to that site again and to interact with the same language users. Moreover, because time will have passed and circumstances changed, it will be impossible – and often scientifically unsound – to collect the same data from the same people for a second time. Particularly with language elicitation, repetition of a task may affect the data in ways that cannot be controlled for. This would not apply, however, to some types of longitudinal studies where the methodology may rely on obtaining data by using the same materials at different points in time. Finally, collecting sign language data of any kind requires good contact with the deaf[5] community and should always be done by and in close collaboration with deaf researchers and members of the deaf community.

### *Naturalistic data*

Naturalistic language data consist in language produced spontaneously and naturally in typical interactional settings and environments. They reflect actual language use that is not influenced by the desires and demands – or even by the presence – of the researcher.

When would it be useful and advantageous to collect naturalistic data? Any research interest focused on understanding interaction and the dynamics of natural conversation must rely on naturalistic data. Examples are studying how turn-taking is managed in dyadic or multi-party conversations (see Coates and Sutton-Spence, 2001), or studying how interlocutors handle sequences of conversational repair – that is, the repair of communication problems that result from errors or misunderstandings (see Enfield et al., 2013). A researcher may also be interested in understanding how language is used in dyadic or multi-party interactions in a particular setting, for example over dinner, at a social event, or in meetings. The conversations occurring in these typical settings can only be captured through naturalistic data collection.

Other areas of research for which the collection of naturalistic data may be encouraged are sign language acquisition (see Baker, van den Bogaerde, and Woll, 2008) and the nature of child-caregiver interactions (see Smith and Sutton-Spence, 2008). Naturalistic data would be recommended, for example, in a longitudinal study of a deaf child acquiring sign language from deaf parents, as they would capture the natural interactions between the child and his/her parents and would track the development of increasingly complex linguistic productions by the child.

Finally, a growing number of sign language researchers is involved in the creation of sign language corpora. As sign languages have no written form, the existence of such corpora is particularly important for providing a store and record of sign language data that are accessible to sign language researchers and to the deaf community. In terms of research, corpora are hugely important for answering questions about the linguistic structure of a language and for understanding patterns of sociolinguistic variation or phenomena of language contact and language change (see Schembri, 2008). Obtaining naturalistic data is crucial to the creation of a sign language corpus. The corpus should include language use that is representative of a particular sign language as it is used by a wide range of signers – of different ages, from different regions, and of different genders.

A comprehensive corpus should also strive to include language use on a range of different topics and spanning different genres (e.g., personal narratives, interviews, descriptions). To achieve this, a researcher may want to introduce a measure of control into the type of data that are obtained. That is, the researcher may provide materials to elicit a certain type of language or may provide certain topics or formats for discussion. We are now getting into the realm of elicited data, the subject of the next section.

## *Elicited data*

A major advantage of collecting elicited data is the control it offers over the type of language captured on video. If you are interested in analyzing a particular domain of language – say, possessive, reciprocal, or locative constructions – there is no guarantee that collecting naturalistic data will provide enough occurrences of the structures and constructions that you are interested in. Having enough occurrences is important for achieving a thorough understanding of the linguistic structure or domain at hand. Moreover, it is necessary for any substantive quantitative analysis.

Stimulus materials designed to elicit a certain kind of language can take different forms. For sign language elicitation, they should be in non-linguistic formats: pictures (e.g., a spatial scene), short video vignettes of single events (e.g., a motion or reciprocal action event), or selected episodes of a cartoon or movie (e.g., the Canary Row cartoon). The use of linguistic formats, like written words or sentences of the surrounding spoken language, runs the risk of structural influence from the spoken language on the sign language data. Elicited picture descriptions, event descriptions, and extended narratives can provide a wealth of valuable data on sign language structure in the targeted domains, for both adult and child language production. The use of a picture description task for eliciting complex verb constructions from children across different age groups in order to study the development of such constructions (Morgan, Herman, and Woll, 2002) is a good example of how an elicitation task may be fruitfully used in sign language acquisition studies.

In general, elicited descriptions and narratives should be produced for a (deaf) addressee, not for the camera or the researcher. This enhances the communicative, discursive nature of the situation and gives more assurance that the elicited material is representative of language use as it would occur in a naturalistic setting. Addressees may also be asked to perform a task, which may further improve the chances of getting the type of language targeted by the elicitation. For example, if the addressee must pick the correct picture out of a set, the signer may be more likely to give a full description of a spatial relationship. Similarly, if the addressee is asked to re-narrate a cartoon event, the signer may be more likely to provide details about the event and to make an effort to produce a coherent, cohesive story. Tasks involving stimulus materials may also be performed collaboratively between two participants. For example, participants can engage in a picture-matching task or in a picture-comparison task; such tasks have been used to elicit possessive and existential constructions across sign languages (Perniss and Zeshan, 2008).

Having comparable data from multiple signers – that is, data elicited using the same materials – is necessary for any serious quantitative and statistical analysis and for assessing whether the data obtained are representative of actual language usage. It addition, it facilitates cross-linguistic comparability, allowing controlled investigation of variation between sign languages in the expression of particular domains. Ideally, an empirical investigation of sign language structure, or the collection of a sign language corpus, should combine both types of data. In any case, as a researcher, you need to make sure that you get the type of data you need for the questions you want to answer. Care must also be taken that the method of data collection does not introduce artifacts into the data – that is, elements that are present as a result of the process of data collection itself. This means creating the best possible situations and settings for data collection. Making decisions about how and where to set up your recording equipment in order to create these optimal conditions is the topic of the next section.

## *Technical aspects of data collection*

Technology is developing fast these days. Hence there is minimal focus here on specific technical details relating to equipment. What is stressed instead in this section is how to collect high-quality archivable data suited to your research purposes. This topic includes selecting the optimal settings and conditions for the filming itself, and ensuring that your data files are properly managed for further use and processing once you have finished filming.

## Equipment

Increasingly, the video cameras available on the market are high-definition camcorders that record directly onto an internal hard drive (HDD), SD memory card, or DVD.[6] These cameras have essentially replaced consumer devices that record onto tapes – that is, DV or HDV cameras; however, the use of tapes remains satisfactory, especially if you already possess such a device.

If and when you purchase equipment – camera(s) and tripod(s) – for collecting sign language data, inform yourself about what is out there and what best meets your demands. You can directly compare prices and specifications of different brands and models at any store that sells digital equipment, or you can do it online. If you work at a university or in a research institute, you may find it useful to consult with the IT staff. Once you have equipment at your disposal, especially if it is new, take time to familiarize yourself with using it and setting it up. Make sure you have enough storage space available for the amount of data you intend to collect. This means having enough space on the drive, disc, or memory card (or having a sufficient number of DV or HDV tapes). If possible, keep your camera plugged in to a power outlet while filming. If this is not possible, make sure you have enough battery power for the intended session of filming. Battery life is optimal with continuous recording, in which episodes of starting, stopping, and zooming are kept to a minimum.

When data have been collected, they must be transferred from the storage device to your computer. If you are recording to tape, the transfer of the digital video to a computer involves real-time capture from the camera and requires a program with which such capture is possible (e.g., Adobe Premiere, WinDV, iMovie, Final Cut Express, Final Cut Pro). The captured video will generally be in a format (MPEG, MOV, or AVI) that can be used in available annotation programs like ELAN or Anvil. Transfer from a drive or memory card is faster and easier by comparison: it requires an SD card slot if the data are stored on a card, or a connector cable (USB or Firewire) if they are stored on a camera-internal drive. However, the compressed format of these data (MTS, M2TS) requires conversion into one of the above formats (e.g., with TMPGEnc, ffmpeg, ffmpegX) before other editing and use with annotation programs is possible. For DVD discs, the underlying MPEG compression format can be used in available annotation programs, but the files need to be ripped from the disc to the computer using special software applications (which may be provided by the manufacturer). The use of DVD-RW discs allows the flexibility of re-recording on the same disc.

## Settings, conditions, and techniques for filming

Filming may seem an easy thing – turn the camera on, aim, press record – but it is in fact important to be aware of – and follow – guidelines for the optimal settings, conditions, and techniques for filming in order to obtain good sign language data.

The first thing to bear in mind is that the camera should not be treated as a roving extension of your own gaze. Mount the camera on a tripod and keep the camera steady and static. Do not pan and do not zoom. Doing so creates shaky and blurry images that are difficult or impossible to analyze. It also increases the chances of missing valuable interaction – for naturalistic data in particular – and attracts

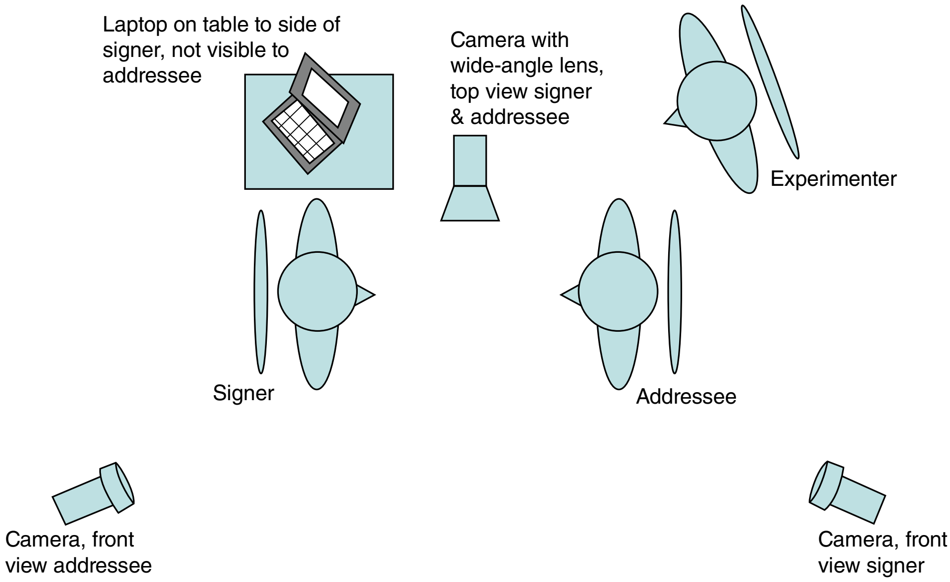


*Figure 4.1* Example of camera and participant set up for the collection of elicited data.

unwanted attention to the presence of the camera. For both naturalistic and elicited data, however, getting a steady and static shot is more than simply a matter of setting the camera down.

If you are eliciting data in a controlled setting, you should prepare in advance a protocol that maps out where the video camera(s) will be positioned, where the participants will be seated, and where stimulus materials will be presented (see Figure 4.1). If you have a studio available, which you are using for multiple recording sessions, you may want to mark out the camera and participant positions on the floor, so that equipment and materials may always be placed in exactly the same locations.

Participants should be in good view, with enough space around the body, so that the signer's hands will not leave the frame during the stream of signing. A good rule of thumb is to frame the bottom of the shot at the knees to mid-thigh and to leave a head's worth of space above the head. Signers should be filmed at a slight side angle, not fully from the front, and slightly from above, with the camera angled slightly downward. This will lend better visual depth and distinguishability to the video. If you are using one video camera, frame both the signer and the addressee in the shot as sitting across from each other, turned halfway to the camera, halfway to each other. If you have more than one camera, you may want to focus one on the signer, another on the addressee. Alternatively, if you are interested in the use of non-manual features (e.g., information conveyed on the face), you may want one camera zoomed into the signer's face. If you are interested in the specific locations and movements of signs, getting a top view of the signer(s) is of extreme value. This can be done by placing the tripod and camera onto a table or by mounting the camera on a microphone stand with boom. If you use more than one camera, the use of a cue (e.g., clapping your hands together) that is captured by all cameras is highly recommended to signal the beginning of filming. The cue should be performed deliberately, not sloppily, so that it can be utilized both visually and auditorily. This greatly aids in the synchronization of video clips, which is necessary for annotation that uses more than one video at the same time. For example, for visual synchronization, all clips would start from the frame at which the hands make contact.

If you are collecting naturalistic data, it is more challenging to frame a shot that will capture targeted interactions on camera. The situation being filmed may be dynamic, with people turning and moving around, even wandering into and out of the captured scene. This means that some of the language production and interaction you are interested in may be obscured from view or may occur off camera. In addition, participants may refer to something or someone located off camera. For this reason, it is important to be familiar with the surroundings: both the constant surroundings – the physical layout of the environment – and the specific surroundings of the filmed interaction – the people and entities involved, those nearby, or anyone relevant in some way. For each filming session, the positions of the participants and of the camera with respect to the surroundings should be carefully recorded. This will make it possible to later reconstruct references to things in the environment that are not visible in the camera shot. This is particularly important in small communities, for example, in research on village sign languages, where signers may regularly refer to people by pointing to their dwelling place within the village.

In any data collection scenario, avoid filming against the light – against a window, if you are indoors, or against the sun, if you are outdoors. Filming against the light will result in a silhouetted signer, and details of the hands and face will be veiled in shadow. Always take care that the lighting is good. The face should be well lit and visible, but not overexposed. Finally, do not shy away from the camera's manual settings. In particular, auto-focus and auto-exposure may not be desirable, depending on the degree and nature of the movement and on the lighting conditions during a filming session.

## *Archiving and accessibility of video data*

Another important aspect of dealing with video data is keeping files identifiable, manageable, and accessible over time. It is best to transfer the video to your computer as soon as possible after filming and to name the files immediately upon transfer. (If you are using tapes, you should label and lock the tapes immediately, such that the material cannot be recorded over accidentally.) It is generally useful to further edit the raw data by cutting them into smaller segments that correspond, for example, to the length of individual tasks or of stimuli, or by marking identifiable breaks in longer interactions or events. Shorter segments will alleviate potential problems related to memory and capacity load, making it easier to work with the files during transfer, viewing, and annotation.

It is also important to create metadata (that is, data about the data) for all video recordings as soon as possible. The metadata are crucial for archiving and for the structured accessibility of data; they should include information about the time, date, and place of recording, the content of the recording, and the participants (e.g., age, language background – paying heed to the ethics of data protection).[7] The metadata should be linked to the video they refer to via the label or file name assigned to the video. It is important to use consistent, systematic, and descriptive names that will be recognizable and informative to you later. Make sure that you have at least one back-up of all your data. It is necessary to keep in mind that the data format will need to be updated over time, in order to ensure that the data remain accessible as technological advances make older formats obsolete. Good labeling and metadata practices are crucial in this regard.

Annotating and coding the data is also a part of keeping them accessible and useful for scientific or community purposes (see Thieberger and Berez, 2012). Preserving language data – the recordings themselves, as well as the annotations and descriptions associated with them – is especially important for sign language, which lack records and descriptions of language use and structure (there is no written form of sign languages, and widespread documentation has only recently become possible, through advancements in video technology). The use of software to annotate sign language data is the subject of the next section.

## Using Software to Annotate Sign Language Data

Data annotation refers to the process of labeling elements in the language data that you want to describe, analyze, quantify, and so on. These elements can be at any level – phonological, morphological, syntactic, semantic, pragmatic, prosodic, social, interactive, temporal, articulatory – and the annotations identifying and labeling them can be descriptive or analytical in nature. Modern software (e.g., ELAN, Anvil) allows you to time-align and link annotations directly to the corresponding segments of video. Using such software is highly recommended for the annotation of video data, particularly of sign language data. Non-time-aligned annotation, for instance using a spreadsheet or a table, is also possible, of course. However, the advantages of doing time-aligned annotation far outweigh the perceived difficulties of becoming a proficient user of the software. The reliability and functionality of programs is steadily increasing, as developers work in association with researchers to develop the features that will be most useful and functional.[8]

Before you start annotating data, however, you need to have a clear idea about what you want to annotate. The first subsection here focuses on devising an annotation or coding scheme suited to your research question. When you have done this, you can move on to implementing the scheme within the program you are using. As stated in the introduction, the exposition offered here focuses on sign language annotation using ELAN. The second subsection elaborates a workflow in ELAN that will help you work efficiently and effectively with sign language data. The third subsection discusses other annotation software, specifically Anvil and iLex, highlighting similarities with and differences from ELAN.

### *Devising a coding scheme: Deciding what to annotate*

Decisions as to what elements of the data to analyze, and according to what criteria, need to be made in advance of the actual annotation. In effect, this means devising a coding scheme. The annotations you use to transcribe and code your data should allow you to extract all the information you are interested in when you get to the stage of analysis and quantification.

There is a vast multitude of possible research questions. You should approach your data – and indeed data collection itself – with an established idea of what you want to know and what your research aims are. For example, you may be interested in the properties of classifier predicates used to express location and motion; in co-articulation effects in the expression of a single sign in different environments; in the scope and temporal alignment of non-manual markers with the stream of manual signs (see Crasborn, van der Kooij, Waters, Woll, and Mesch, 2008); in how possession is expressed in a particular sign language (see Zeshan and Perniss, 2008); or in the semantic and syntactic behavior of a sign that functions as a perfect marker (see Meir, 1999). Some discussion of how you would approach your data to devise a coding scheme to answer such questions follows.[9]

For example, what would you need to extract from your data if you were interested in the use of classifier predicates? You would want to know which (classifier) handshapes are used for which types of entities. You would want to know where the hands are placed in space in order to encode referent location and how the hands move in space in order to encode referent motion. You may want to know to what extent the use of space is topographic or iconic with respect to the scene or event being described. Your coding scheme should thus include ways to identify and label different handshapes (and the entities to which they refer) and different locations and areas in sign space. You may also want to know about the entire construction in which classifier predicates occur: for example, whether the predicates are preceded by nominal signs, whether they occur in final position in an utterance, or whether and when they occur in simultaneous constructions (see Vermeerbergen, Leeson, and Crasborn, 2007 on simultaneity in sign languages).

Similarly, if you are interested in studying co-articulation effects, you will need to identify all the occurrences of a particular sign or group of signs in your data and describe how articulation is affected in each case by the environment – that is, by the preceding and the following signs. Your coding scheme in this case would concentrate on phonetic and phonological aspects of the data. You would need to label the handshape, the place of articulation, and the movement characteristics of each sign occurrence and compare it to the phonologically specified "target" form of the sign. Similar descriptions of the preceding and the following signs would allow you to analyze the effect of phonological environment on sign articulation and the phonetic variability of individual signs.

Finally, if you are interested in the scope and temporal alignment of non-manual markers with concurrent manual signing, you will need to label the occurrence of all non-manual markers of the kind you have decided to investigate (e.g., brow raise), marking the precise onset and offset (e.g., the moment at which the brows begin to move up and the moment at which they return to neutral position). These onsets and offsets would need to be temporally evaluated with respect to the lexicon and syntax of the manual signing stream. This kind of coding relies crucially on temporal alignment and on being able to compare the timing of different articulatory channels that simultaneously contribute information to the language signal. Programs like ELAN or Anvil, which allow time-aligned annotation, are indispensable for this kind of coding.

### *Implementing the coding scheme: Learning how to annotate*

After you have decided what elements of the data you want to code and according to what criteria, it is time to implement the coding scheme in the annotation program. This section aims at providing a workflow that will allow researchers interested in annotating sign language data to work effectively with ELAN. The section does not go into detail about the technicalities of the interface or into optimizing the use of functions, for instance the use of keyboard shortcuts. These things can best be learned by practice, by trial and error, and by consulting the user guides and manuals that are available for download.[10]

The defining feature of ELAN in terms of the annotation framework it provides is the simultaneous display of multiple tiers on a timeline that is linked to one or more language data video files (see Figure 4.2). All tiers are user-defined and reflect your specific annotation needs.[11]

Tiers are associated with various attributes – linguistic type, stereotype, controlled vocabulary, and parent – which define the constraints that apply to each tier and its annotations. For example, it is possible to create dependencies between tiers (that is, by assigning parent tiers and by defining dependency stereotypes). In addition, tiers can be assigned a fixed set of annotation values (that is, by creating controlled vocabularies associated with linguistic types). Understanding how tier attributes are related to each other and defining their values in an order that reflects their interrelationships is important to making ELAN work for you. Table 4.1 lists the tier attributes specified for each tier shown in the ELAN screen shot in Figure 4.2. Below, the tier

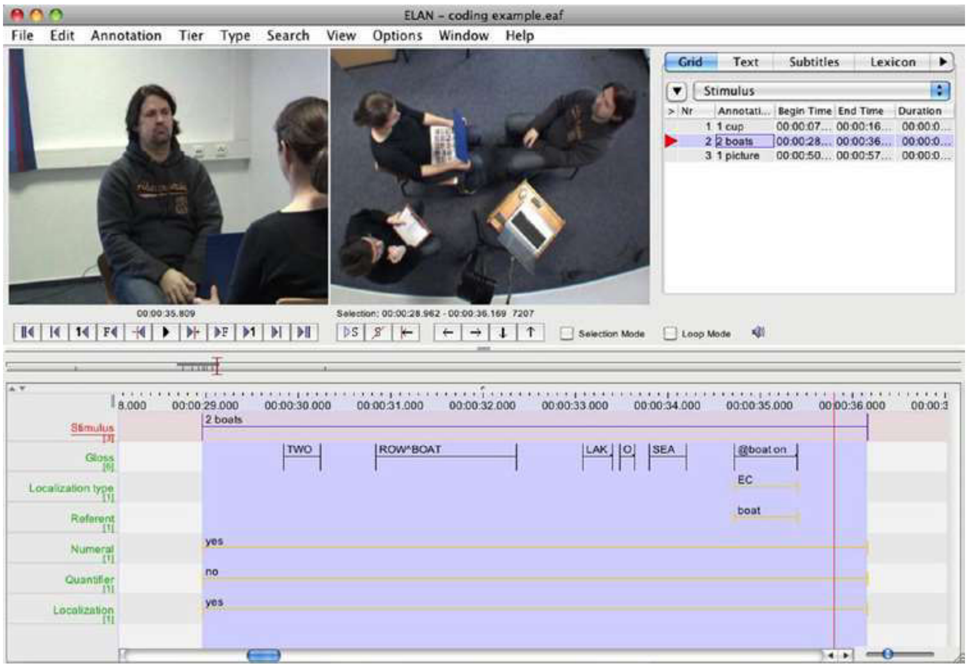


*Figure 4.2* Screen shot of ELAN, showing multiple, simultaneously coded annotation tiers.

*Table 4.1* Example of table listing tier names and tier attributes.

| *Tier name* | *Parent tier* | *Linguistic Type* | *Stereotype* | *Controlled Vocabulary* |
|---|---|---|---|---|
| Stimulus | none | Stimulus | none | none |
| Gloss | Stimulus | Sign gloss | Include In | none |
| Localization type | Gloss | Loc type | Symb. Assoc. | Loc type |
| Referent | Gloss | Referent | Symb. Assoc. | none |
| Numeral | Stimulus | Numeral | Symb. Assoc. | Yes/no |
| Quantifier | Stimulus | Quantifier | Symb. Assoc. | Yes/no |
| Localization | Stimulus | Localization | Symb. Assoc. | Yes/no |

attributes are defined and explained with reference to the example in Table 4.1, and this is followed by a description of how to create tiers and specify their attributes in ELAN. It is advisable to make a table of tier names and attributes (as the one in Table 4.1) before actually creating tiers in ELAN itself.

The tier name should identify the information being annotated. (A tier's linguistic type, which is specified together with the tier's name, can be similarly descriptive; we return to the linguistic type further below.) In our example the stimulus tier identifies the stimulus item (a picture of two boats) that has elicited the stretch of signing marked out by the annotation (i.e., the utterance comprising the description of the stimulus picture). We can see in Table 4.1 that the stimulus tier is the parent of a number of other tiers: the gloss, numeral, quantifier, and classifier tiers. The annotations on these dependent tiers are constrained by (or they refer to) the annotations on the parent tier in some way. For example, we can see in Figure 4.2 that the annotations on the numeral, quantifier, and classifier tiers are exactly the same length as the annotation on their parent stimulus tier. This type of dependency relationship – specified by the stereotype attribute – is called "symbolic association." Annotations on tiers with a stereotype of symbolic association are automatically the same length as annotations on their parent tier. Establishing this as an attribute of a tier can improve the time-efficiency of coding. For instance, on the numeral tier in our example, we want to know whether or not a numeral sign appears in the stimulus description. Annotations on the numeral tier (*yes* in this case) thus refer to the full length of the annotations on the parent stimulus tier, and the stereotype is specified accordingly as symbolic association.

Another type of dependency, or referring, relationship is reflected in the annotations on the gloss tier. On this tier the length of individual annotations is freely determined on the basis of the length of individual signs in the video data; but, overall, all glosses associated with a particular stimulus description are fully contained within the length of the annotation on the parent stimulus tier. The dependency stereotype for this type of relationship is called "included in." In Figure 4.2, for example, it would not be possible to create annotations on the gloss tier to the left or right of the *two boats* annotation on the parent stimulus tier (i.e., to the left or right of the selection marked in blue).

Defining a controlled vocabulary for use with a tier can greatly relieve manual annotation work. If you know that your annotations on a given tier will be from a

*Table 4.2* Example of table listing names of controlled vocabularies and their entry values.

| *Controlled Vocabulary* | *Entry Values* |
|---|---|
| Loc type | EC, HC, SASS, noun |
| Numeral | yes, no |
| Quantifier | yes, no |
| Localization | yes, no |

fixed set (e.g., you will annotate only *yes* or *no*, as on the numeral tier), it is possible to predefine this set as a controlled vocabulary. Selecting annotations from a fixed set (that appears as a drop-down menu) is more time-efficient than typing each annotation in by hand, and is recommended when possible. In our example, four different tiers have a controlled vocabulary associated with them: the localization type, numeral, quantifier, and classifier tiers. It is useful to create a separate table listing the controlled vocabularies you would want to use, with all their entry values specified (see Table 4.2). Names for controlled vocabularies can be descriptive, like the tier name. For example, the "loc type" vocabulary associated with the localization type tier has four specified entry values that identify different types of signs that may be localized in space in the spatial description data from which our example is taken, namely EC (entity classifier), HC (handling classifier), SASS (size and shape specifier), and noun signs.

How do you get these attributes set up and defined in ELAN? For setting up tiers and working most successfully and effectively with them, it is important that steps are taken in an order that reflects the interrelationships between tiers, types, and their attributes. In ELAN, creating a tier requires defining three things: the tier name; the tier's linguistic type; and the tier's parent tier (i.e., dependency relationship), if applicable. Creating a linguistic type, in turn, also requires specifying three things: the type name; the stereotype (i.e., the type of dependency relationship); and the use of a controlled vocabulary, if applicable. Thus it is best to work backwards along these interrelationships in getting your tier structure set up.

With your tables of names and values at hand, first create all the controlled vocabularies that you want to work with. To do this, go to Edit in the menu bar and select Edit Controlled Vocabulary. In the window that appears, you can specify the names and entry values of each controlled vocabulary you wish to use. Once this is done, all linguistic types should be defined (Type > Add New Linguistic Type). Here the available stereotypes[12] and the controlled vocabularies you have created appear for selection from drop-down menus. The type name serves primarily to link a given tier to its type attributes (i.e., to the stereotype and controlled vocabulary values). This link can be made transparent by giving a tier and its associated linguistic type the same (or a very similar) name (see Table 4.1). Finally, the tiers themselves can be created (Tier > Add New Tier). The parent tier and the linguistic type are selected from the drop-down menus, which are populated on the basis of, and as a result of, your creation of tiers and types. The tier name, finally, is what you see in the ELAN annotation window itself. As the tiers are created, they appear on the left side of the timeline, vertically arranged.

It may be, of course, that you want to make changes to your tiers and tier attributes once you have started annotating. As you go along, it is possible to create new controlled vocabularies and to edit existing ones, and to associate an existing tier with a new or a modified controlled vocabulary. It is also possible to associate a tier with a different linguistic type after the tier has been created. It is difficult, however, to change the dependency relationships between tiers, in other words to specify the parent tier and the stereotype of a given tier after it has been created. When you have created all of your tiers – in effect, when you have implemented your coding scheme in ELAN – it is useful to save the file in order to use it as a coding template. Individual annotation files can be created and saved by linking the appropriate video files to the program file in each case (Edit > Linked Files). ELAN can display up to four video files simultaneously. The use of multiple videos allows you to display different views of the same signers at the same time, for example the front view and the top view displayed in the ELAN screen shot in Figure 4.2.

When you start annotating, your research question and the type of information you want to code will help you to fine-tune the use of program functions, tools, and modes that will be most suited and useful. You will become practiced at using the keyboard shortcuts (View > Shortcuts, for a list) and you will learn to navigate the active tiers and annotations; to copy, paste, or duplicate annotations between and along tiers; to modify the time alignments of existing annotations; and to modify the content of existing annotations as needed.

The annotation of sign language data requires careful attention to the video stream. You are likely to create many of your annotations in the default annotation mode (Options > Annotation Mode) by moving the red crosshair frame by frame along the timeline to find the beginning and end points of an individual sign. However, the more automated, coarse-grained method of creating annotations that the segmentation and transcription modes offer (Options > Segmentation Mode; Transcription Mode) may prove useful for an initial pass through the data. The segmentation mode allows the fast placement of annotations if you hit the Enter (or Return) key (↵) as the video plays at the set speed. For example, if your aim is to analyze the use of a particular sign as a perfect marker in a sign language (see de Vos, 2012), you may want to use the segmentation mode for an initial identification of all occurrences of this sign within a video file. The annotations created in the segmentation mode are empty. The transcription mode can prove useful for filling them with the intended label (e.g., the gloss identifying the perfect marker). This text-oriented mode allows you to efficiently enter the annotation text into each of the marked segments. Taking these grossly marked segments as a point of departure, you can return to the annotation mode and modify the annotations so as to fit the precise duration of the sign in each case and to create new annotations according to your aims.

When you have completed the annotation of your data, the next step is to use these annotations to analyze and quantify your data and to extract the relevant information about the sign language, according to your specific research questions. ELAN allows various functions within the program itself, including acquiring simple statistics (View > Annotation Statistics) and more sophisticated search functions within and between tiers, which can be implemented by using both temporal and structural constraints (Search > Find; Search Functions). In addition, ELAN allows the export of tiers and annotations as a tab-delimited text file, which can be opened and processed as an Excel spreadsheet. This is possible for single (Export As > Tab-Delimited Text) or multiple ELAN files (Export Multiple Files As > Tab-Delimited Text).

## *Other annotation software*

This section offers an overview of other software that similarly allows time-aligned annotation of sign language data and highlights options for sign language lexicography and lemmatization (for a more comprehensive overview and discussion of such software, see Rohlfing et al., 2006).

Anvil is another popular annotation tool that provides the simultaneous display of multiple annotation tiers and time-locked video (for a detailed description of the program, see Kipp, 2014).[13] Anvil offers many of the same features as ELAN, albeit with different names. Anvil uses tracks instead of tiers, elements instead of annotations, and dependency relationships between tiers/tracks are reflected in the track type instead of the stereotype (for example the singleton track type corresponds to ELAN's symbolic association stereotype, creating automatic temporal overlap). Anvil also offers the option of creating and using controlled vocabularies, which it calls "value sets."

A significant difference between Anvil and ELAN lies in the file structure. Anvil keeps annotation files separate from the coding scheme files – that is, the specification files where the track structure is specified. All annotations point to, or are associated with, a particular specification file. This means that any changes made to your coding scheme along the way (like the addition of tracks, or the change of track attributes) will be automatically made within the annotation files. This is an advantage when dealing with large amounts of annotated data files, as existing coding schemes may develop and change over time. In ELAN, where this information is not stored separately, making changes to a coding scheme template file does not automatically implement these changes in the annotated files that have already been created using this template.[14]

A further feature of Anvil not offered by ELAN is that it supports motion capture viewing. Anvil can read the most commonly used motion capture data formats (BVH and ASF/AMC), allowing the viewing of the 3D image of the body and the time-aligned display of curve analyses (e.g., velocity and acceleration). Finally, Anvil is notable in that it allows users to code spatial information directly on the video frame. This is done by means of an attribute that stores screen locations associated with individual video frames. The Anvil GUI allows these screen locations to be defined by clicking directly on the video. ELAN files may be imported into Anvil, which offers the possibility of using these features of Anvil with data annotations (previously) created in ELAN. Finally, one drawback of Anvil compared to ELAN is its limited video format compatibility. Thus it may be difficult to get video to play in Anvil. In contrast, ELAN supports a wide range of video formats and codecs.

iLex is a tool that offers an annotation environment that is directly integrated with a lexical database. This is valuable for sign language lexicography and corpus analysis, particularly for the process of lemmatization (where different forms of a sign, such as inflected forms, must be associated to a single entry). In iLex, glosses in the annotation environment refer to token occurrences of signs and are linked automatically to type entries in the lexical database. This achieves consistent type/token matching that does not depend on the labeling consistency of the glosses themselves (Hanke and Storz, 2008). In addition, having an integrated lexical database means that the database is created simultaneously with the transcription and glossing of the sign language data.

This feature of lexical database incorporation is not offered by ELAN or Anvil. However, ELAN can be used in conjunction with the lexicon tool LEXUS, such that sign glosses (in ELAN) can be associated to lexical entries. Achieving consistent type/token matching with this combined functionality relies on consistent glossing across occurrences of different forms of a sign.

## Conclusion

This chapter has aimed to provide researchers interested in working with sign language data with guidelines for recording video data and for annotating them with the help of dedicated software. It has focused on the decision-making and workflow required to translate research questions into obtaining suitable data, and on devising and implementing appropriate coding schemes. The technical details of using video equipment and annotation software have not been stressed.

This chapter fills a gap in the available literature on research methods. Research on sign language has been included in previous volumes that target research methods within a specific theoretical framework, for instance cognitive linguistics (see Wilcox and Morford, 2007), or within a specific area of research, such as sign language acquisition (see Baker et al., 2008). In addition, research method handbooks have included chapters dedicated to studying, coding, and recording co-speech gesture (see Mittelberg, 2007; Sweetser, 2007 within a cognitive linguistics framework; Seyfeddinipur, 2012 on linguistic fieldwork methods; and Müller et al., 2013 for a wide range of methods chapters) and chapters including general guidelines for using video to document language usage (e.g. Margetts and Margetts, 2012). A chapter dedicated to obtaining and annotating sign language data within a general volume on research methods for sign language research is an important addition to the literature. Moreover, by focusing on the effective development and implementation of a coding scheme, it is a valuable supplement to the technical and comprehensive detail of available user manuals and guides for ELAN, including user guides intended for use of ELAN with sign language data (Bickford, 2005; Crasborn and Sloetjes, 2008).

Finally, the use of video data and time-aligned, video-based annotation should not be reserved for language expressed in the visual modality. All of language and communication is a multimodal process. In order to understand it, we need to understand how it is constructed in usage – which includes patterns of eye gaze, gesture, speech, and sign. In general, the study of language should make use of video recording and visual annotation if it aims to understand the complex phenomenon of human communication and interaction.

### Notes

1 ELAN is the Eudico Linguistic Annotation tool developed at the Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands. The software is available for free download at http://tla.mpi.nl/tools/tla-tools/elan

2 Anvil is a video annotation tool developed by Michael Kipp from the University of Applied Sciences in Augsburg, Germany. The software is available for free download at www.anvil-software.org/download/index.html

3 iLEX is the integrated lexicon tool developed at the Institute of German Sign Language and Communication of the Deaf (Institut für Deutsche Gebärdensprache und Kommunikation Gehörloser, IDGS) at Hamburg University, Germany. See www.sign-lang.uni-hamburg.de/ilex

4 LEXUS is a web-based lexicon tool developed by the Language Archive at the Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands. The software is available for free download at http://tla.mpi.nl/tools/tla-tools/lexus

5 In this chapter I use the lowercase form deaf throughout.

6 Devices allowing depth and motion sensing, e.g. Kinect devices and 3D cameras, are also on the market. These can be useful, for example, for purposes of automatic sign recognition and sign synthesis (and also for more sophisticated types of instrumented capture).

7 See, for example, www.mpi.nl/IMDI/ for information on the ISLE MetaData Initiative (IMDI) as a standard for structured metadata description.

8 The steady software improvements and general technological advancements have rendered largely irrelevant older concerns about platform and video format compatibility, as well as about processing capacity issues.

9 Independently of what your research question is, data annotation should include a transcription of the signing, which identifies and glosses all the signs in the utterances of interest. The use of consistent glosses to uniquely identify signs is crucial to creating a searchable and functional sign corpus and database (see Johnston, 2001 on ID-glosses for consistent type/token matching and Hanke and Storz, 2008 on integrated type/token matching in iLex).

10 The ELAN user guide and user manual are available for download from http://tla.mpi.nl/tools/tla-tools/elan; in addition, there is a forum for asking questions (http://tla.mpi.nl/forums/software/elan).

11 The use of a data category register (DCR) like ISOcat may aid in finding and defining tier and coding categories. ISOcat is the DCR utilized by CLARIN (Common Language Resources and Technology Infrastructure), which provides an integrated research infrastructure for language resources (see www.isocat.org and www.clarin.eu).

12 There are four predefined stereotypes. "Symbolic association" and "included in" have been mentioned; the other two are "symbolic subdivision" and "time subdivision." With these, the length of a parent tier annotation can be subdivided into multiple annotations on a dependent tier; but there can be no gaps between these individual annotations, and they are automatically assigned equal length.

13 In addition, there are good video tutorials available on the Anvil web site, as well as email support and access to a user forum (visit www.anvil-software.org).

14 To deal with this, ELAN offers options for merging and importing tiers between files, a slightly more cumbersome solution than that offered by Anvil. However, this is weighed against the ease, in ELAN, of creating tiers and implementing the coding scheme within the program: this is comparatively difficult in Anvil, where specification files are .xml files that must be created in an editor. Because creating these files from scratch may be daunting to users, the Anvil directory (to be downloaded and installed) includes a subdirectory called "spec," which contains examples of coding schemes that can be used and amended.

## Keywords

annotation software; data annotation; data coding; data collection; data management; elicited data; naturalistic data; sign language; techniques for filming; video recording

## See Also

Chapter 1; Chapter 5; Chapter 6; Chapter 7; Chapter 8; Chapter 10

## Suggested Readings

Thieberger, N. (Ed.). (2012). *The Oxford handbook of linguistic fieldwork*. Oxford: Oxford University Press.

## References


Baker, A., van den Bogaerde, B., and Woll, B. (2008). Methods and procedures in sign language acquisition studies. In A. Baker and B. Woll (Eds.), *Sign Language Acquisition*. Amsterdam: John Benjamins, pp. 1–50.

Bickford, A. (2005). Using ELAN: A getting-started guide for use with sign languages. No longer available. http://www.und.edu/dept/linguistics/textbooks/UsingElan.pdf

Coates, J., and Sutton-Spence, R. (2001). Turn-taking patterns in Deaf conversation. *Journal of Sociolinguistics* 5(4), 507–529.

Crasborn, O., and Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood, and E. D. Thoutenhoofd (Eds.), *Construction and exploitation of sign language corpora: Third workshop on the representation and processing of sign languages*. Paris: ELDA, pp. 39–43.

Crasborn, O., van der Kooij, E., Waters, D., Woll, B., and Mesch, J. (2008). Frequency distribution and spreading behavior of different types of mouth actions in three sign languages. *Sign Language & Linguistics* 11(1), 45–67.

Enfield, N. J., Dingemanse, M., Baranova, J., Blythe, J., Brown, P., Dirksmeyer, T., Drew, P., Floyd, S., Gipper, S., Gisladottir, R. S., Hoymann, G., Kendrick, K., Levinson, S. C., Magyari, L., Manrique, E., Rossi, G., San Roque, L., and Torreira, F. (2013). Huh? What? A first survey in 20 languages. In M. Hayashi, G. Raymond, and J. Sidnell (Eds.), *Conversational repair and human understanding*. New York: Cambridge University Press, pp. 343–380.

Hanke, T., and Storz, J. (2008). ilex: A database tool for integrating sign language corpus linguistics and sign language lexicography. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood, and E. D. Thoutenhoofd (Eds.), *Construction and exploitation of sign language corpora: Third workshop on the representation and processing of sign languages*. Paris: ELDA, pp. 64–67.

Johnston, T. (2001). The lexical database of Auslan (Australian Sign Language). *Sign Language & Linguistics* 4(1/2), 145–169.

Kipp, M. (2014). ANVIL: The video annotation research tool. In J. Durand, U. Gut, and G. Kristofferson (Eds.), *The Oxford Handbook of Corpus Phonology*. Oxford: Oxford University Press, pp. 420–436.

Margetts, A., and Margetts, A. (2012). Audio and video recording techniques for linguistic research. In N. Thieberger (Ed.), *The Oxford Handbook of Linguistic Fieldwork*. Oxford: Oxford University Press, pp. 13–53.

Meir, I. (1999). A perfect marker in Israeli Sign Language. *Sign Language & Linguistics* 2(1), 41–60.

Mittelberg, I. (2007). Methodology for multimodality: One way of working with speech and gesture data. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, and M. J. Spivey (Eds.), *Methods in Cognitive Linguistics*. Amsterdam: John Benjamins, pp. 225–248.

Morgan, G., Herman, R., and Woll, B. (2002). The development of complex verb constructions in British Sign Language. *Journal of Child Language* 29, 655–675.

Müller, C., Cienki, A., Fricke, E., Ladewig, S., McNeill, D., and Teßendorf, S. (Eds.) (2013). *Body—Language—Communication: An international handbook on multimodality in human interaction* (vol. 1). Berlin: De Gruyter.

Perniss, P., and Zeshan, U. (2008). Possessive and existential constructions: Introduction and overview. In U. Zeshan and P. Perniss (Eds.), *Possessive and existential constructions in sign languages* (Sign Language Typology series 2). Nijmegen, Netherlands: Ishara Press, pp. 1–31.

Rohlfing, K., Loehr, D., Duncan, S., Brown, A., Franklin, A., Kimbara, I., Milde, J.-T., Parrill, F., Rose, T., Schmidt, T., Sloetjes, H., Thies, A., and Wellinghof, S. (2006). Comparison of multimodal annotation tools: Workshop report. *Gesprächsforschung* 7, 99–123. (Online journal on verbal interaction, also at www.gespraechsforschung-ozs.de)

Schembri, A. (2008). British Sign Language Corpus project: Open access archives and the observer's paradox. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood, and E. D. Thoutenhoofd (Eds.), *Construction and exploitation of sign language corpora: Third workshop on the representation and processing of sign languages*. Paris: ELDA, pp. 165–169.

Seyfeddinipur, M. (2012). Reasons for documenting gestures and suggestions for how to go about it. In N. Thieberger (Ed.), *The Oxford handbook of linguistic fieldwork*. Oxford: Oxford University Press, pp. 147–165.

Smith, S., and Sutton-Spence, R. (2008). Adult–child interaction in BSL nursery: Getting their attention. In A. Baker and B. Woll (Eds.), *Sign language acquisition*. Amsterdam: John Benjamins, pp. 119–140.

Sweetser, E. (2007). Looking at space to study mental spaces: Co-speech gesture as a crucial data source in cognitive linguistics. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, and M. J. Spivey (Eds.), *Methods in cognitive linguistics*. Amsterdam: John Benjamins, pp. 201–224.

Thieberger, N., and Berez, A. L. (2012). Linguistic data management. In N. Thieberger (Ed.), *The Oxford handbook of linguistic fieldwork*. Oxford: Oxford University Press, pp. 90–118.

de Vos, C. (2012). The Kata Kolok perfective in child signing: Coordination of manual and non-manual components. In U. Zeshan and C. de Vos (Eds.), *Sign languages in village communitites: Anthropological and linguistic insights*. Berlin: Mouton de Gruyter, pp. 127–152.

Wilcox, S., and Morford, J. (2007). Empirical methods in signed language research. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, and M. J. Spivey (Eds.), *Methods in cognitive linguistics*. Amsterdam: John Benjamins, pp. 171–200.

Vermeerbergen, M., Leeson, L., and Crasborn, O. (Eds.). (2007). *Simultaneity in signed language: Form and function.* Amsterdam: John Benjamins.

Zeshan, U., and Perniss, P. (Eds.). (2008). *Possessive and existential constructions in sign languages* (Sign Language Typology series 2). Nijmegen, Netherlands: Ishara Press.