# Segmenting the Swedish Sign Language Corpus: On the Possibilities of Using Visual Cues as a Basis for Syntactic Segmentation

**Carl Börstell, Johanna Mesch, Lars Wallin**

Dept. of Linguistics, Stockholm University
S–106 91 Stockholm, SWEDEN
calle@ling.su.se, johanna.mesch@ling.su.se, wallin@ling.su.se

**Abstract**

This paper deals with the possibility of conducting syntactic segmentation of the Swedish Sign Language Corpus (SSLC) on the basis of the visual cues from both manual and nonmanual signals. The SSLC currently features segmentation on the lexical level only, which is why the need for a linguistically valid segmentation on e.g. the clausal level would be very useful for corpus-based studies on the grammatical structure of Swedish Sign Language (SSL). An experiment was carried out letting seven Deaf signers of SSL each segment two short texts (one narrative and one dialogue) using ELAN, based on the visual cues they perceived as boundaries. This was later compared to the linguistic analysis done by a language expert (also a Deaf signer of SSL), who segmented the same texts into what was considered syntactic clausal units. Furthermore, these segmentation procedures were compared to the segmentation done for the Swedish translations also found in the SSLC. The results show that though the visual and syntactic segmentations overlap in many cases, especially when a number of cues coincide, the visual segmentation is not consistent enough to be used as a means of segmenting syntactic units in the SSLC.

**Keywords:** Swedish Sign Language, corpus linguistics, segmentation, nonmanuals

## 1. Introduction

### 1.1. Segmenting sign language

Previous studies have shown that nonmanual markers in sign language (e.g. eye blinks, eyebrow movement, gaze, head movement etc.) readily have syntactic functions, but they also have prosodic functions (see Pfau and Quer (2010) for an overview). For instance, Wilbur (1994) argues that eye blinks can mark units with a variety of different functions in American Sign Language (ASL), such as syntactic, prosodic, discourse, and narrative units. Nonmanual markers have been shown to work in complex patterns, and boundaries between prosodic units are often aligned with those between syntactic units, with boundaries usually characterized by a change in several of the nonmanual features (Nespor and Sandler, 1999).

Other studies have tried to investigate the possibility of using prosodic and/or intonational information as a means of reliably segmenting certain linguistic units in sign language (e.g. clauses or sentences). A small-scale study on Auslan[1] compared the alignment of so-called *Intonation Units* (cf. Chafe (1994)) with syntactic units, and found that IUs often align with a single clause, although there are also cases of multiple IUs within a single clause as well as a single IU spanning multiple clauses (see Ferrara (2012) for a summary). In another study, Fenlon et al. (2007) found that signers and non-signers alike accurately identify sentence boundaries in sign language texts, and because several visual cues can coincide with each other, some boundaries were stronger (i.e. more visual cues coinciding) than others. This was shown to be true when the subjects viewed both a familiar (British Sign Language (BSL)) and an unfamiliar one (SSL). However, a study on German Sign Language (DGS) investigated whether certain formal boundary markers accurately coincide with sentence boundaries, but found that though the markers often align with the boundaries, they do so in a non-consistent and non-exclusive fashion (Hansen and Heßmann, 2007).

For corpus purposes, the idea of having syntactically segmented sign language texts is still under investigation. The annotation guidelines for the Auslan Corpus use the label *clause-like unit* (CLU) as a tentative equivalent of a ”potential clause”, corresponding to more traditional types of clause units as well as segments containing sign language specific strategies of describing or ”showing” meaning (Johnston, 2013, 50–51).

### 1.2. The Swedish Sign Language Corpus

The Swedish Sign Language Corpus—henceforth SSLC—consists of a collection of video recordings of pairs of Deaf signers, spanning various text types, e.g. semi-spontaneous dialogues, narratives, and elicitation tasks. The SSLC consists of approximately 25 hours of video data comprising 42 different signers (male and female; ages 20–82). The recordings have accompanying ELAN annotation files, which are being published as they are produced (see Mesch et al. (2014) for the current version). The ELAN annotation files currently consist of six tiers: four for sign glosses (two tiers for each signer; one for each of a signer's hands), and two for written Swedish translations (one for each signer). Signs are annotated in individual cells with glosses corresponding to Swedish translations of each sign, with additional suffixed tags for some types of signs, such as fingerspellings or gestures (see Wallin and Mesch (2014) for the current guidelines for annotating SSL). To date, the SSLC features annotation files (with

[1] Australian Sign Language

glosses and translations) for about 19% of the total amount of recorded data (Mesch et al., 2012). However, there is no segmentation done above the lexical level, i.e. cells for individual signs, which complicates data exporting and concordance viewing by not being searchable within syntactic units (e.g. clauses or sentences).

The addition of a clausal/sentential segmentation is a natural first step toward analyzing—and annotating—e.g. semantic roles or syntactic functions. Automatic annotation of such categories would be facilitated by an existing linguistic segmentation. A first attempt at an automated induction of word classes was done using the segmentation for Swedish translations as utterances (Sjons, 2013), but a segmentation done independently of another language should prove more viable.

## 2. Methodology

### 2.1. Data

For the experiment, we selected two separate texts from the SSLC data: one narrative text, 1:35 minutes long; and one dialogue text, 2:08 minutes long. In the narrative text, only one signer in the pair is signing, with the other signer acting as a receiver of the signing, in the dialogue text, both signers are signing.

### 2.2. Experiment

The experiment consisted of two parts: in the first part, a number of Deaf signers individually segmented sign language texts based on visual cues; in the second part, a SSL expert segmented the same texts based on syntactic information. The segmentations were then layered on top of each other, together with the pre-existing glosses and translations. From this, we analyzed the data with regard to the number of boundaries marked within and across participants, as well as the amount of overlap in the alignment of the identified boundaries across participants and the syntactic segmentation. The two parts of the experiment are further described below.

#### 2.2.1. Visual segmentation

For the first part, we asked seven Deaf signers (three female) to participate in a segmentation task on the two selected texts. The participants (labeled A–G) have all completed higher education in sign language linguistics. The participants first saw the video on a computer screen, and were then asked to segment the text into units based on the visual cues they interpreted as boundaries. The participants were asked to mark any occurrence of a boundary by pressing a key on the keyboard, but they were allowed to stop the video in order to go back in the video and mark the boundary's exact location. The test was conducted directly in the ELAN window, but all other annotation tiers were hidden during the experiment session. The experiment was run twice for the dialogue data, in order for the participants to segment the text in two tiers—one for each of the two signers in the video.

#### 2.2.2. Syntactic segmentation

For the second part, we had a Deaf sign language researcher analyze the two texts and segment the texts according to available linguistic information regarding semantics and syntax, thus marking segments that semantically and/or syntactically could constitute a clause (although short turns or individual feedback signs would also be segmented as separate units in the dialogue text).

#### 2.2.3. Analyzing overlaps

After the tiers from all participants and the language expert were layered on top of each other, it was clear that there was a certain amount of overlap across participants and the syntactic segmentation. However, in order to establish which segmentations should be grouped together—especially in sequences with several close consecutive segmentations—some criteria had to be defined. We assumed a point of overlap if (1) any two segment boundaries were less than 300 ms apart, and (2) the total span of the overlap point was less than 1000 ms. The latter criterion was quite generous, but deemed necessary in order to allow for cases of longer holds/pauses where some participants marked the end of the previous sign as a boundary, some in the middle of the hold/pause, and others at the beginning of the following sign. If a participant had marked two adjacent boundaries very close to each other such that both were within the scope of a single point of overlap, one boundary was included in the overlap and the other was excluded from it.[2]

## 3. Results

### 3.1. Number of segmentations

The results show that the participants differ on the number of segmentations they made for both texts. The number of boundaries marked range from 19 to 52 (average 39.7) for the narrative text, and 44 to 109 (average 73.7) for the dialogue text. Table 1 shows the number of boundaries marked by each participant as well as the boundaries in the syntactic segmentation made by the language expert (labeled "LE").

| Participant | Narrative | Dialogue |
|---|---|---|
| A | 45 | 76 |
| B | 32 | 79 |
| C | 52 | 109 |
| D | 19 | 44 |
| E | 49 | 62 |
| F | 41 | 85 |
| G | 40 | 61 |
| LE | 51 | 97 |

Table 1: Number of marked boundaries per participant.

Looking at the actual video with the aligned segmentation tiers, it is clear that while some participants (e.g. participant C) have segmented many single nonmanuals (e.g. a single eye blink), others have focused on the more extensively marked segments (e.g. eye blink, nod, and hands hold/drop at once). Comparing the participants' segmentation to the syntactic segmentation, there is one obvious difference in the task that was given to the participants compared to that

[2] If a participant had made a double segmentation with an interval of <200 ms it was counted as a single boundary.

of the researcher: while the participants task was based on marking the presence of a boundary with the help of visual cues, the researcher segmented cells that contain a syntactic clausal unit. Thus, in the researcher's segmentation, the cells are sometimes interspaced by a pause in the signing, but the boundaries marked by the participants are usually punctual and often coincide with either the start or the end of such an interspace (i.e. marking boundaries either at the end of one unit or the start of another). An example of this is given in figure 1 below, in which Signer 1 in the dialogue text puts his hands together for a short pause after a sign.

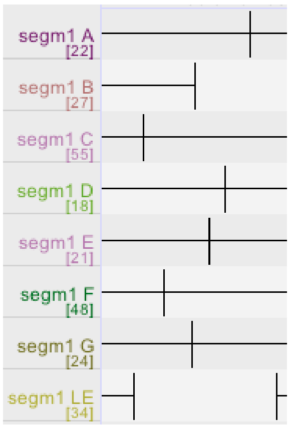


Figure 1: Differences in boundary markings.

As figure 1 illustrates, some of the participants (e.g. C and F) chose to mark the retraction of the hands as the boundary, whereas others (e.g. A and D) chose to mark the middle of the pause as the boundary. Since the language expert (LE) only annotated signing, the pause (about 1000 ms long) is thus marked by the absence of a cell. This difference in tasks could be one explanation as to why the number and placement of segmentations may vary between the language expert and the participants.

### 3.2. Amount of overlap

Turning to the points of overlap, it is visible from the data that some locations in the texts are characterized by a high number of overlap across participants and the language expert. Using the criteria described in section 2.2.3., the total number of unique boundaries was 78 for the narrative text, and 167 for the dialogue text. For some of these, only a single participant had marked it as a boundary, but the majority of boundaries are shared with at least one other participant and/or the language expert, thus constituting a point of overlap. Figures 2 and 3 below show the distribution of overlaps across participants, and whether or not these align with the syntactic boundaries made by the language expert, for each of the two texts. The X axis corresponds to the number of participants sharing a point of overlap, and the Y axis corresponds to occurrences. The different colored columns show whether or not the occurrences also overlap with a syntactic boundary.

Figure 2 shows that there is a tendency for the more agreed upon points of overlap to be marked also as a syntactic boundary by the language expert, although there are a couple of cases for which a syntactic boundary has not been marked by a single participant. However, there is a clear idiosyncrasy among (some of) the participants' segmentations, resulting in a very high number of boundaries marked by a single participant.

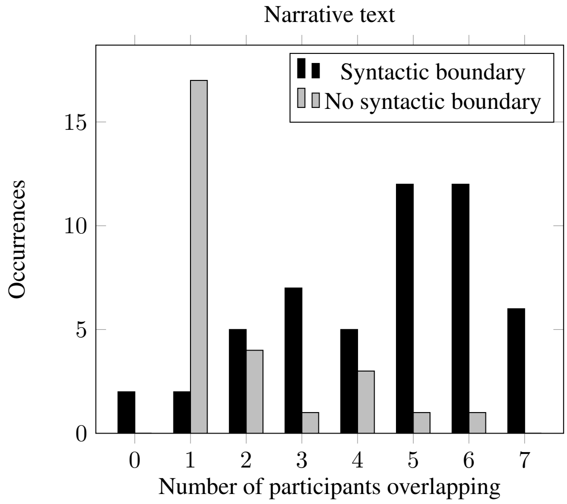


Figure 2: Overlap between visual and syntactic boundaries.

Figure 3 shows a similar pattern, again demonstrating the high amount of idiosyncrasy among the participants resulting in a high number of boundaries identified by a single participant. However, it also shows a global trend of syntactic boundaries being marked most often if many participants also marked it as a boundary, and vice versa.

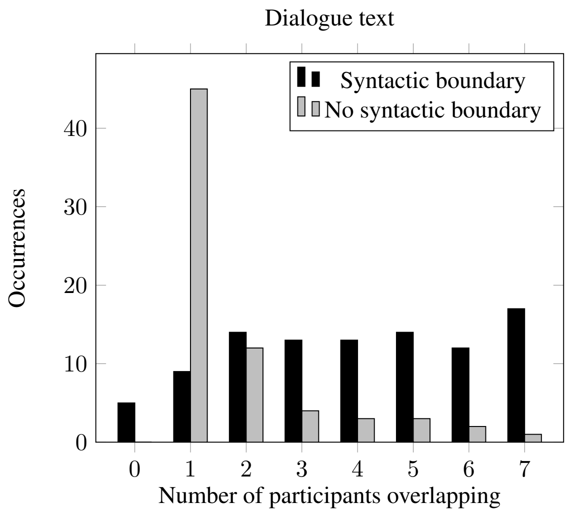


Figure 3: Overlap between visual and syntactic boundaries.

#### 3.2.1. Comparison with the translations

For the narrative text, the translator used for the SSLC—a hearing native signer (i.e. CODA[3])—made a translation without access to any of the segmentations made by either the participants or the language expert. Looking briefly at the alignment of cells in this translation, it is clear that although many of the endpoints overlap with those made in the syntactic segmentation, the number of segmentations is not the same. The translation tier generally has longer segments, often spanning over several syntactic segmentations. While the endpoints of most translation cells do align with the endpoints of some syntactic segmentations, the range of syntactic segmentations within the scope of a single translation cell ranges between 1 and 9 (average 3). Thus, the translation tier cannot be considered an accurate segmentation of the SSLC on a clausal level.

## 4. Conclusion

This minor study is a first step toward adding a linguistic segmentation to the SSLC, which could prove useful in future work of annotating e.g. word classes or syntactic functions in the corpus. The purpose of the study was two-fold: first, we wanted to see how well different signers' segmentation of sign language texts based on visual cues correspond to each other; second, we wanted to see how well a segmentation based on visual cues corresponds to a segmentation based on a more in-depth analysis of linguistic units.

Our investigation demonstrated that while signers show some agreement in their segmentation of sign language texts based on visual cues, it is not completely reliable as a means of segmenting syntactic units. For instance, the number of segmentations varies across participants doing the same task—segmenting the text based on visual cues—and these do not always align with syntactic boundaries, as suggested by Hansen and Heßmann (2007). When layering the visual segmentations, they do pattern in a way that shows that a high degree of overlap often corresponds to the presence of a syntactic boundary, but using a high number of participants for visual segmentations might prove more time-consuming than a few language experts segmenting the SSLC for syntactic units. Thus, a visual-based segmentation would be neither the most accurate nor the most practical way of adding a syntactic segmentation to the SSLC.

## 5. Acknowledgments

The Swedish Sign Language Corpus project was funded by a grant from Riksbankens Jubileumsfond (In2008-0276-1-IK).

## 6. References

Chafe, Wallace. (1994). *Discourse, Consciousness, and Time: The flow and displacement of conscious experience in speaking and writing*. University of Chicago Press, Chicago, IL.

Fenlon, Jordan, Denmark, Tanya, Campbell, Ruth, and Woll, Bencie. (2007). Seeing sentence boundaries. *Sign Language & Linguistics*, 10(2):177–200.

Ferrara, Lindsay. (2012). *The grammar of depiction*. Ph.D. dissertation, Dept. of Linguistics, Macquarie University.

Hansen, Martje and Heßmann, Jens. (2007). Matching propositional content and formal markers: Sentence boundaries in a DGS text. *Sign Language & Linguistics*, 10(2):145–175.

Johnston, Trevor. (2013). Auslan Corpus Annotation Guidelines. Centre for Language Sciences, Department of Linguistics, Macquarie University.

Mesch, Johanna, Wallin, Lars, Nilsson, Anna-Lena, and Bergman, Brita. (2012). Dataset. Swedish Sign Language Corpus project 2009–2011 (version 1). Sign Language Section, Department of Linguistics, Stockholm University.

Mesch, Johanna, Rohdell, Maya, and Wallin, Lars. (2014). Annotated files for the Swedish Sign Language Corpus. Version 2. Sign Language Section, Department of Linguistics, Stockholm University.

Nespor, Marina and Sandler, Wendy. (1999). Prosody in Israeli Sign Language. *Language and Speech*, 42(2-3):143–176, June.

Pfau, Roland and Quer, Josep. (2010). Nonmanuals: their grammatical and prosodic roles. In Brentari, Diane, editor, *Sign languages: A Cambridge language survey*, pages 381–402. Cambridge University Press, New York, NY.

Sjons, Johan. (2013). Automatic induction of word classes in Swedish Sign Language. M.A. thesis, Dept. of Linguistics, Stockholm University.

Wallin, Lars and Mesch, Johanna. (2014). Annotation guidelines for sign language texts. Version 5. [In Swedish]. Sign Language Section, Department of Linguistics, Stockholm University.

Wilbur, Ronnie B. (1994). Eyeblinks & ASL Phrase Structure. *Sign Language Studies*, 84(1):221–240.

[3] Child of Deaf Adult