1. INTRODUCTION

- Multi-camera HD (1920x1080) 25-50 fps video material on FinSL and PipeSL
- Collected at the Sign Language Centre, University of Jyväskylä (Puupponen & al. 2014)
- Informants age 20 to 80 years from different parts of Finland, performing several tasks.
- Annotation conventions are being developed in several Finnish projects: here we discuss the conventions used in ProGram (2013-2018) and CF/NSL (2014-2015) projects (see Tables 1-3).

2. SEGMENTING SIGN STREAM INTO SIGNS

- Our annotation process is divided into cycles which, in turn, are organized into phases.
- During each phase, we check - and, if needed, correct - the outcome of the previous phase.
- In the first phase, we delimit signs from the stream into the tier Sign_sequence.

2.1 What is our definition of the sign?

- We define signs as distributionally free combinations of meaning and form.
- Consequently, we accept that signs may be composed of several concatenated morphemes (e.g. compounds, various plural forms involving spellings etc.).
- As functionalists, we prioritize meaning over form (ex. our definition of compounds).

2.2 Where in the sign stream the sign begins and where it ends?

- In contrast to the mainstream view, we consider signs to be relatively long units that melt together in continuous signing (see Figure 1).
- Moreover, we consider signs to be conceptually different units from annotation cells.
- Consequently, our annotation cells are long durationally, and the beginnings and endings of our annotation cells do not (necessarily) correspond to the (exact) beginnings and endings of signs (see Figure 2).

3. IDENTIFYING THE CONTEXTUAL MEANINGS OF SIGNS

- In the second annotation phase, we identify the contextually grounded meaning of each sign and write it into the tier Raagl_gloss (see Figure 3).
- We use two types of glosses for this purpose: Meaning glosses and Description glosses.
- Meaning glosses specify those signs that have a lexical or otherwise easily identifiable meaning. Meaning glosses are written in capital letters.
- Description glosses specify various depicting signs and gestures. Description glosses are written in small letters.
- The fact that we prioritize meaning over form (see Section 2.1) and take into account the context in which the sign is used means that, on the level of Meaning glosses, we distinguish, for example, between WORK (a noun) and TO-WORK (a verb), regardless of their fairly similar forms.

4. SYSTEMATIZED GLOSSING

- The goal of this phase is to systematize Meaning glosses and Description glosses (see Section 3) so that one sign is always identified with the same gloss (cf. ID glossing as proposed by Johnston 2008).
- We process Meaning glosses on the tier Meaning_gloss and Description glosses on the tier Description_gloss (see Figure 4).
- After the systematization is finished, the two tiers are merged into the tier S-gloss.
- The systematization of Meaning glosses requires a corpus lexicon.
- Our current lexicon is Excel-based but we plan to adopt Signbank for the purpose.

Figure 1. Traditionally delimited short version (left) and the revised long version (right) of the FinSL sign BLACK. For a detailed discussion about the two views, see Jantunen (2015). Images adapted from PSK.

Figure 2. ELAN tiers illustrating the difference between the short (upper tier in red) and the long sign (bottom tier in blue). In average, the long sign is 3.8 frames longer than the short sign (with 25 fps video) (Jantunen 2015).

Figure 3. ELAN screenshot showing, signer-specifically, the tiers Raagl_gloss (Raagl_gloss), Transition (Käännös), and Remarks (Huomioita) used in the second annotation phase. The Grid Viewer (upper right hand corner) illustrates various Meaning and Description glosses.

Figure 4. ELAN screenshot showing the tiers Raagl_gloss (Raagl_gloss), Meaning_gloss (Mechan_gloss), Description_gloss (Ratikanl_gloss), and S-gloss (SigW/s) for both signers. The Grid Viewer (upper right hand corner) illustrates various S-glosses. Note that the annotation of signs in the screenshot follows the specific conventions of the ProGram project and includes additional information about the word-class prefixed to the gloss (n-nominal, v-verbal, u-unspecified).

5. CONVENTIONS FOR WRITING S-GLOSSES

<table>
<thead>
<tr>
<th>Type of the sign</th>
<th>Example</th>
<th>Explanation and notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple lexical sign</td>
<td>POKA, SYÖDA</td>
<td>Words in capital letters</td>
</tr>
<tr>
<td>Long sign</td>
<td>j.UMI</td>
<td>BKnKO</td>
</tr>
<tr>
<td>flooding</td>
<td>On:Oo</td>
<td>Gloss-ON (IP in English). Meaning is specified after a colon.</td>
</tr>
<tr>
<td>“Negative” signs</td>
<td>EHALUA</td>
<td>Word s (in English) added to the beginning of the gloss.</td>
</tr>
<tr>
<td>Fingers alphabet</td>
<td>m=</td>
<td>Prefixes s (in English) is used also for the signugging.</td>
</tr>
<tr>
<td>Numbers</td>
<td>w, VIV</td>
<td>Prefixes s (in the same prefix also in English)</td>
</tr>
<tr>
<td>Name sign</td>
<td>w=VIILE</td>
<td>Prefixes r (in English)</td>
</tr>
<tr>
<td>Gender</td>
<td>=KAMMEN-YLOE</td>
<td>Prefixes s (r in English). descriptions are written in small letters.</td>
</tr>
<tr>
<td>Description</td>
<td>/ep</td>
<td>Prefixes r (in English). descriptions are written in small letters.</td>
</tr>
</tbody>
</table>

Table 2. Specific combinations of meaning and form. Only synonymy requires overt marking. This is done by adding structural information about some of the basic parameters of the sign (handshape, location, movement) into the parenthesis after the main gloss.

Table 3. Our system of writing glosses. The Description gloss syntax is a reduced version of the Meaning gloss syntax

6. CONCLUSION

- Our annotation conventions have so far been developed and tested primarily in ProGram project. CF/NSL project is likely to simplify some conventions and introduce new ones.
- A very salient characteristic of the annotation work in ProGram has been that the sign has been glossed from the perspective of the dominant hand only. In CF/NSL project, the activity of both hands will be described.
- In the future, we will consider the possibility to add yet a fourth phase into our annotation cycle: ID-glossing in the sense currently developed in the Corpus and SigW/s Project by Helsinki team. In finding the common ground, the results of the Digging into Signs project (Crasborn & al. 2012) will be of great assistance.

REFERENCES


Figure 5. The systematization of Meaning glosses requires a corpus lexicon. After the systematization is finished, the two tiers are merged into the tier S-gloss.

Figure 6. ELAN screenshot showing the tiers Raagl_gloss (Raagl_gloss), Meaning_gloss (Mechan_gloss), Description_gloss (Ratikanl_gloss), and S-gloss (SigW/s) for both signers. The Grid Viewer (upper right hand corner) illustrates various S-glosses. Note that the annotation of signs in the screenshot follows the specific conventions of the ProGram project and includes additional information about the word-class prefixed to the gloss (n-nominal, v-verbal, u-unspecified).