1. INTRODUCTION

- The ID-glossing, i.e. lemmatization (Johnston 2008) and the creation of annotation conventions for ID-glossing Finnish Sign Language (FinSL) was started in the Corpus and SignWiki Project, at the Finnish Association of the Deaf (FAD) in Helsinki in 2014.
- Annotation conventions are mainly based on Auslan annotation conventions (Johnston 2014).
- Data: Language Policy Programme for the National Sign Languages in Finland (2010), the FinSL edition, 2½ h of signed formal monologue
- Mission: to create the basis for a multi-purpose FinSL corpus annotation and its conventions that can be enriched and developed in the future
- Lexical database: we temporarily use the administrative site of the Suvi online dictionary (see Figure 1), but we plan to replace it by the Signbank.¹
- In the ongoing first phase, the focus is on ID-glossing the fully lexical signs, while other kinds of signs are given only a general tag, apart from some exceptions, so that they can be analyzed and annotated more specifically in the future.
- Cooperation with the Sign Language Centre at the University of Jyväskylä

Figure 1. Sign OLLA ‘have’, ‘exist’ in the administrative site of the Suvi online dictionary, our temporary lexical database. The sign has two phonological variants that differ in handshape. The handshape variants are described in the database.

Figure 2. ELAN screenshot showing the tiers Id-glossi oik (ID-gloss right hand), Id-glossi vas (ID-gloss left hand), Kommentteja (comments), Alkuperäisteksti (original text), Käännös (translation) and Kommentteja_käännös (comments on translation). A tier Meaning for contextual meaning of each token is planned to be added to the basic template. The screenshot demonstrates also the annotation of two-handed signs on two tiers, the annotation of fragment buoys (fpoiju) and the signs co-occurring with them as well as the length of the annotations. In addition to the annotation tiers, the text viewer shows many different ID-glosses.

2. ID-GLOSSING PRINCIPLES FOR LEXICAL SIGNS

- Each lexeme is given a unique ID-gloss.
- The citation form of the lexeme, the lemma, and its phonological variants are described in the lexical database (see figure 1).
- Choosing the ID-gloss:
  1. A unique Finnish word is used when possible, even for (near) synonyms. The word represents the core meaning of the sign.
  2. If the same word needs to be used, the following specifications can be applied:

   a. A symbol or a word referring to a structural feature (e.g. handshape or movement) of the sign is suffixed in parentheses.
   b. A word that hints at the meaning is suffixed, separated by a dot.
   c. A number suffix is added after a dot.

HELPPO EASY
PIKKUJUTTU SMALL-THING
TUNTUA-HELPOLTA FEEL-EASY

TEHDÁ(B) WORK(B)
TEHDÁ(S) WORK(S)

JARJESTÁÁ(pompiva) ORGANIZE(bouncing)
JARJESTÁÁ(lukuva) ORGANIZE(sliding)

KUUSI SIX
KUUSI.puu SPRUCE.tree

RAHA.1 MONEY.1
RAHA.2 MONEY.2

1. Sanna Nordlund, Leena Savolainen & Päivi Mäntylä

Corpus and SignWiki Project / The Finnish Association of the Deaf
3. Start and End of a Sign

- Theoretically, we accept the view according to which a sign is a relatively long unit, and so-called transitional movements are part of the sign (Jantunen 2015). Therefore, only little space is left between the annotations, often just one frame.
- Defining exact criteria for the beginning and end of a sign has proven to be difficult, for example, the order in which the sign's structural features appear varies, and the beginning and the end of two signs often merge with each other.
- For two-handed signs, the start and end of the sign follows the dominant hand, as in BSL conventions.
- When the two hands produce different signs, the tiers remain independent, as in NGT conventions.
- For buoys and signs co-occurring with them, the tiers also remain independent (see Figure 2).

4. Other Practices Compared to the Digging Project

<table>
<thead>
<tr>
<th>TOPIC</th>
<th>THE PRACTICE IN CORPUS AND SIGNWIKI PROJECT</th>
<th>EXAMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TWO-HANDED SIGNS</td>
<td>The ID-gloss is annotated on both hand tiers</td>
<td>TEHDAB Bernardino</td>
</tr>
<tr>
<td>BUOYS</td>
<td>List buoys and fragment buoys are annotated as in BSL, except that the codes are in Finnish and in lower case. The handshape is not specified.</td>
<td>os:puju</td>
</tr>
<tr>
<td></td>
<td>On fragment buoys, also the ID-gloss of the sign of which it is a fragment is given after the colon (see Figure 2), as in Auslan (Johnston 2014).</td>
<td>os:puju</td>
</tr>
<tr>
<td>REPETITION</td>
<td>Repeated signs are annotated separately, as in BSL and NGT corpora, when articulated clearly separately. Grammatical repetition is not annotated at this stage.</td>
<td>os os os</td>
</tr>
<tr>
<td>MANUAL NEGATIVE INCORPORATION</td>
<td>A negative suffix NEG separated by a plus. Note: The word El ‘not’ is used in signs having inherent negative meaning.</td>
<td>TIE:AA+NEG</td>
</tr>
<tr>
<td>NUMBER INCORPORATION</td>
<td>The incorporated number is added in parentheses at the moment. The parenthesis might be replaced by a plus in the future to indicate the morpheme status of the number handshapes.</td>
<td>TUNTI(4)</td>
</tr>
<tr>
<td>SIGN NAMES, FINGERSPELLING, DEPICTING SIGNS AND GESTURES</td>
<td>A transparent code (grefa) in lower case letters. More specific annotation of these sign types has not been decided yet.</td>
<td>na (nimivittoma)</td>
</tr>
<tr>
<td>POINTING SIGNS</td>
<td>At the first stage, only a code, apart from some exceptions.</td>
<td>os (osoitus)</td>
</tr>
<tr>
<td></td>
<td>At the first stage, only a code, apart from some exceptions.</td>
<td>os (osoitus)</td>
</tr>
</tbody>
</table>

5. Consistency with the Practice of Glossing Spoken Languages?

The Leipzig Glossing Rules have become a worldwide standard in spoken language research. In our opinion, it would be important to discuss the consistency between the standard suggested by the Digging into Signs project and the Leipzig Glossing Rules. A consistent practice would facilitate the participation in the typological discussion. However, if differing conventions are chosen, they should be justified in the guidelines.

6. Conclusions

- Since the FinSL and BSL conventions are both based on Auslan annotation conventions, they have quite a lot in common. However, in some points we agree more with the conventions for NGT.
- According to our view, the basic principle should be that the ID-gloss tier is reserved only for the ID-glosses (lemmas), and all the other (grammatical & semantic) information is annotated on separate tiers.
- We prefer using a different word for each ID-gloss, if possible, and transparent suffixes when needed (see section 2). As well, we prefer using abbreviations for different sign types (e.g. depicting signs and fingerspelling), since they are easier for the annotator and the user of the corpus to remember than meaningless codes.

---

REFERENCES


Johnston, Trevor 2013. Formational and functional characteristics of pointing signs in a corpus of Auslan (Australian sign language) are the data sufficient to posit a grammatical class of ‘pronouns’ in Auslan? Corpus linguistics and linguistic theory, Vol. 9, No. 1, p109-159.


---