From meaning to signs and back: Lexicography and the Swedish Sign Language Corpus

Johanna Mesch, Lars Wallin
Department of Linguistics, Sign Language Section
Stockholm University
SE-10691 Stockholm, Sweden
E-mail: johanna.mesch@ling.su.se, wallin@ling.su.se

Abstract

In this paper, we will present the advantages of having a reference dictionary, and how having a corpus makes dictionary making easier and more effective. It also gives a new perspective on sign entries in the dictionary, for example, if a sign uses one or two hands, or which meaning "genuine signs" have, and it helps find a model for categorization of polysynthetic signs that is not found in the dictionary. Categorizing glosses in the corpus work has compelled us to revisit the dictionary to add signs from the corpus that are not already in the dictionary and to improve sign entries already in the dictionary based on insights that have been gained while working on the corpus.

Keywords: Swedish Sign Language, lexicography, sign language corpus, sign entries

1. Introduction

For many years, a primary challenge for lexicography work in sign languages has been to describe signs and the meaning of the signs. The choice of signs to include in a sign language dictionary is often based on the intuition of native signers and their experiences as sign language researchers, lexicographers and sign language teachers for beginners. One of the troublesome tasks in Swedish Sign Language dictionary-making, as for many other dictionaries, is to create a dictionary entry for a sign that has more than one meaning (cf. Ordbog over Dansk Tegnsprog and Suvi - Suomalaisen viittomakielen verkkosanakirja). Lemmatization of signs in signed language corpora is one of the ways corpus work can inform dictionary-making. But there is no clear consensus on the definition of lemmatization in Swedish Sign Language, despite lexicographers having many years of experience with this problem.

Sign language corpora constitute a revolution in dictionary-making, giving lexicographers a new opportunity to transfer some meanings (contextual meaning) of a sign from corpus to dictionary, and to demonstrate use in context. Here we discuss the advantages of having a reference dictionary (see Johnston, 2008; Ormel & et al., 2010; Schembri & Crasborn, 2010) and how having access to a corpus should make dictionary-making easier and more effective. This article also provides a new perspective on sign entries in the dictionary, from different points of view.

2. Swedish Sign Language Corpus

The Swedish Sign Language Corpus Project, henceforth SSLC, was created over a three-year span, from 2009–2011. During this period, we recorded 42 informants aged 20 to 82 from several different regions in Sweden. These recordings consist of free conversations and storytelling, as well as retellings of, for example, "The Snow Man". During the project, we developed transcription conventions for annotation work with sign language discourses (Wallin et al., 2011), thanks to a great deal of earlier experience with transcription in sign language discourse, but there is much to be learned from others, and we need to develop the transcription work with ELAN further (for more about ELAN, see Crasborn et al., 2008).

The aim of the SSLC is to publish an accessible collection of sign language discourses on the web in autumn 2012 (Mesch, Wallin, Nilsson, Bäckström, Johansmide & Bergman, 2012), and it is intended that the corpus will provide an accurate impression of what Swedish Sign Language sentences look like, as well as to contribute new signs and variants for the online version of the SSLD. The SSLC was especially constructed for the purpose of documenting sign language discourses and to facilitate dictionary-making rather than for sociolinguistic analysis, as has been the case for the sign language corpora used for British Sign Language (BSL), Sign Language of the Netherlands (NGT) and, to some extent, the Australian Sign Language (Auslan). Additionally, the German Sign Language Corpus (e.g. Hanke, 2002) is close to being a dictionary itself. At any rate, it is our hope that SSLC-based studies will be of great importance to future research in sign linguistics, not only by making it possible to analyze Swedish Sign Language grammar, but also to be useful for a variety of educational purposes.

3. Lexicography

A lexicographic work of Swedish Sign Language was initiated in 1988 in the Sign Language section of the Department of Linguistics at Stockholm University. In 2001, the first dictionary resulting from the project went online, and it was entitled the *Digital version of Swedish Sign Language Dictionary.* A follow-up dictionary called the *Swedish Sign Language Dictionary* was created in 2008 and has been in development since (see Mesch, Wallin & Björkstrand, in this volume). Today, it has

approximately 8,000 sign entries. There are also nine lexical databases with specialized lexicons in Swedish Sign Language, totaling 4,300 sign entries, of which some are also found in the Swedish Sign Language Dictionary, henceforth the SSLD. In the SSLD, a sign is described in terms of its handshape, orientation, location, movement, Swedish translations, and sign transcription. The lexicographic work group associated with this project documents the vocabulary of Swedish Sign Language and continuously updates the web-based dictionaries.

New entries are created not only as a result of observations of the various lexicographers involved with the project, but also user searches for words not currently in the dictionaries. There are also some specialized databases pertaining, for instance, to healthcare. The work of the lexicographers is often based on their intuitions concerning language, but also on their studies in sign language linguistics. With respect to the method for creating sentences, the (extern) actors are given still photos only, in which a person is presenting a sign, without any written Swedish words being introduced. This creates associations with many signs of different meanings. Then the sign actors have to devise their own usage examples in the form of sentences. This elicitation may be good, but in the future it will be combined with the possibility for lexicographers to look up related sentences in the corpus and change them to a suitable form for use in the dictionary.

4. Annotation work with glosses

In the annotation work for the corpus, a suitable gloss for a sign/utterance is needed that is intended to be equivalent to the SSLD sign entry. There it is possible to find a sign through by several through several possible fields (see Mesch, Wallin & Björkstrand, this volume). Corpus annotators can search handshape, movement, place of articulation, or Swedish keyword, and receive a specific gloss for this sign. Annotators have often checked some signs in the SSLD thanks to possibilities in searching for example handshape. A gloss ARBETA 'to work' is intended for a sign ARBETA/ARBETE 'to work/work', see fig. 1. The gloss would be the same in the SSLC and the SSLD, but it is crucial because of their different points of view.

Thus, the SSLC team has created the controlled vocabulary in ELAN for the SSLC with only the signs/utterances that appear in dialogues and elicited narratives in the corpus. The controlled vocabulary does not contain all the sign entries from the SSLD. Additionally, annotators use the SSLD only to check which gloss is written for a sign in the SSLD However, it is sometimes difficult for an annotator to guess a gloss in annotation work through only the controlled vocabulary because the controlled vocabulary show only the glosses of the Swedish word, not the meaning.

In earlier work (Mesch & Wallin, 2008), we discussed sign variants, for example, the sign SPRINGA 'run', and how to annotate these instead of using annotation/ transcription conventions with SPRINGA-1, SPRINGA-2 and SPRINGA-3. Such variants highlight the need for a reference dictionary so that annotators can select the gloss that is used in the SSLC. In the SSLC, we have chosen to use the convention GLOSS (handshape for sign transcription for SSL) like this:

- SPRINGA(G) *'RUN-clenched hand'*
- SPRINGA(Lböjd) *'RUN-hook finger hand'*
- SPRINGA(Ω) *'RUN-double hook hand'*

Annotators would find suitable glosses to use for signs in the SSLD, as a prototype is shown in figure 1. See also, for example, the Danish Sign Language Dictionary (Kristoffersen & Troelsgård, 2010).

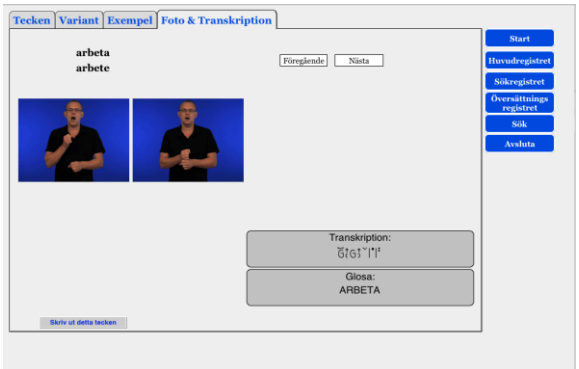

Figure 1: Reference dictionary with suitable gloss for annotators

5. Different point of views for glossing

Here we provide some examples of different points of view for glossing between the SSLD entry and SSLC gloss. Some problems are shown in figures 2-4. One sign entry can have many keywords in Swedish in the SSLD (see figure 2). The sign in SSLC is written in only one gloss, SKRIVA-PÅ (*sign*). This means that many sign entries would be created in the SSLD, if it is to follow SSLC annotation conventions.

Figure 2: Dictionary entry: skriva på (*sign for a football team*), underteckna (*sign a letter*), justera (*verify protocol of meeting*), signera (*sign your initials*).

As shown, there are some challenges to finding a suitable gloss in the SSLC annotation. For example, for the sign EREKTION '*erection*' in figure 3, the corpus annotator

would write a more detailed gloss with the purpose of indicating a correct meaning, whereas in the SSLD a more general meaning would be chosen. That is, it does not pertain only to the stiffening of a man's penis, but also to a woman's clitoris. In the SSLC, this would be glossed as PENIS-STYV (*'penis-stiff'*). This means that this particular sign entry in the SSLD needs to be changed.

Figure 3: Dictionary entry: erektion (*erection*)

A gesture-like sign is one debatable example in annotation work; for example, 'inte mitt problem' (*not my problem*) is found in the SSLD and also figure 4. In the corpus annotation for the SSLC, the gloss '(g-)PF' (*palm fronted*) is used. In the corpus material, a similar sign with a different meaning "200 members and *not more*" is found (see figure 5). (g-)PF in the SSLC has a different meaning than in the SSLD 'not my problem'. Further analysis will be needed to resolve inconsistencies like this.

Another example in the SSLD is the sign entry BJÖRN with two keywords in Swedish: 'nalle' (*teddy bear*) and 'bear' (*bear*). But in the SSLC, this is transcribed as NALLE (*teddy bear*) or BJÖRN (*bear*) following the mouth movement rather than the signs. Such examples show the importance of having a reference dictionary.

The annotation work for SSLC has created many questions for the SSLC and lexicographic work and creates a great opportunity to develop the dictionary to a greater degree. Usage examples for signs will not have to always be constructed by lexicographers but can also be taken from the corpus material.

Figure 4: Dictionary entry: inte mitt problem (*not my problem*).

Figure 5: A sentence from the SSLC '…ni fick två hundra medlemmar till, om ni fick det…' (*'...you had two hundred members more, if you had that...*).

6. Back to the dictionary

Categorizing glosses in the corpus work has compelled us to return to our dictionary and refine sign entries. Signs in the corpus that are not already in the dictionary need to be added. Therefore, suitable glosses are not taken from the SSLD to the SSLC, only keywords. After years of annotation, we see the need to have a part of the gloss-ID in the SSLD for future uses.

Annotating corpus material also gives some ideas of what signs to include in the SSLD, such as semi-lexicalized signs, (p-)signs with a high frequency (figure 6).

Figure 6: Gloss in SSLC: (p-)VARELSE(L)-KOMMA *'(p-)being(index finger)-COME'*

Some new dictionary entries will be affected by the corpus annotation work, for example, two dictionary entries, NALLE (*teddy bear)* and BJÖRN (*bear),* would be separated and added in the SSLD. Also, signs with different parts of speech, e.g. subject and verb, are already separated in the SSLD, but they need to be verified when we get more grammatical information from the corpus.

7. Conclusion

The annotation work for Swedish Sign Language Corpus has given its annotators and the lexicographers from the Swedish Sign Language Dictionary an opportunity to discuss the material and learn from each other. Glossing corpus material provides clues as to which signs to include as we continuously update the online SSLD. The lexicographic work will not strictly follow some regulations, but it will be open to opportunities for further development and to serve the needs of users better.

Other avenues for future collaboration will include creating shortcuts between the SSLD entries and places in corpus recordings where a particular sign is used. Although corpus material cannot cover all signs and their uses in different registers and situations, if used in the right way, it can be an important part of the learning process by illustrating some of the ways in which signs can be used through examples of actual usage.

We have found in our work that we must often shift from the SSLD to the SSLC and back to the SSLD when annotating our corpus materials. And, as we do, we find difficulties arising one after the other. Thus, we see the importance of having a reference dictionary; however, it is important to note that if they are associated too strongly, both individual parts lose autonomy and the strength of their appeal to different "cultures", or points of view. Anyway, it is good for them to make a contribution to each other, for example, to see if a sign has one or two hands for a main sign entry and its variant, and to find a model for categorization of polysynthetic signs that are not found in the dictionary. This kind of discussion between the two projects helps both the SSLD and the SSLC become more useful resources for students, teachers and researchers.

9. Acknowledgements

This work was supported in part by Riksbankens jubileumsfond (In2008-0276-1-IK) and by financial support from the Swedish Ministry of Education and Research.

8. References

Crasborn, O. & Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In: Proceedings of LREC 2008, Sixth International Conference on Language Resources and Evaluation, 39–43.

Johnston, T. (2008). Corpus linguistics and signed languages: no lemmata, no corpus. In LREC (Language Resources and Evaluation Conference) 2008, Proceedings of the 3rd workshop on The Representation and Processing of Signed Languages: Construction and Exploitation of Sign Language Corpora, Marrakech, Morocco. pp. 82-87.

Hanke, T. (2002). iLex – A tool for sign language lexicography and corpus analysis. In: M. González Rodriguez, Manuel and C. Paz Suarez Araujo (eds.): Proceedings of the third International Conference on Language Resources and Evaluation. Las Palmas de Gran Canaria, Spain; Vol. III. Paris: ELRA. 923-926.

Mesch, J. & Wallin, L. (2008). Use of sign language materials in teaching. In LREC (Language Resources and Evaluation Conference) 2008, Proceedings of the 3rd workshop on The Representation and Processing of Signed Languages: Construction and Exploitation of Sign Language Corpora, Marrakech, Morocco. pp. 82-87.

Mesch, J., Wallin, L., Nilsson, A.-L., Bäckström, J., Johansmide, L. & Bergman, B. (2012). Korpus för det svenska teckenspråket (Swedish Sign Language Corpus). Dataset to be published. Sign Language Section, Department of Linguistics, Stockholm University.

Ordbog over Dansk Tegnsprog. On-line publication: http://www.tegnsprog.dk

Ormel, E., Crasborn, O., van der Kooij, E., van Dijken, L., Nauta, Y.E., Forster, J. & Stein, D. (2010). Glossing a multi-purpose sign language corpus. In LREC (Language Resources and Evaluation Conference) 2010, Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. Valetta, Malta. pp. 186-191.

Schembri, A. & Crasborn, O. (2010). Issues in creating annotation standards for sign language description. I: Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. Valletta, Malta. pp. 212-216.

Suvi – Suomalaisen viittomakielen verkkosanakirja [The on-line dictionary of FinSL]. Finnish Association of the Deaf, 2003. On-line publication: http://suvi.viittomat.net

Svenskt teckenspråkslexikon (2008 ---online). [Swedish Sign Language Dictionary] On-line publication , containing also the specialized databases: http://www.ling.su.se/teckenspraklexikon

Troelsgård, T. & Hedegaard Kristoffersen, J. (2008). An electronic dictionary of Danish Sign Language 9th. Theoretical Issues in Sign Language Research Conference, Florianopolis, Brazil, December 2006. (2008) R. M. de Quadros (ed.). Editora Arara Azul. Petrópolis/RJ. Brazil. http://www.editora-arara-azul.com.br/EstudosSurdos.php

Wallin, L., Mesch, J. & Nilsson, A-L. (2011). Transkriptionskonventioner för teckenspråkstexter. Version 3. Stockholms universitet: Institutionen för lingvistik.