

Sign Language Resources in Sweden: Dictionary and Corpus

Johanna Mesch, Lars Wallin & Thomas Björkstrand

Department of Linguistics, Sign Language Section

Stockholm University

SE-10691 Stockholm, Sweden

E-mail: johanna.mesch@ling.su.se, wallin@ling.su.se, bjorkstrand@ling.su.se

Abstract

Sign language resources are necessary tools for adequately serving the needs of learners, teachers and researchers of signed languages. Among these resources, the *Swedish Sign Language Dictionary* was begun in 2008 and has been in development ever since. Today, it has approximately 8,000 sign entries. The Swedish Sign Language Corpus is also an important resource, but it is of a very different kind than the dictionary. Compiled during the years 2009–2011, the corpus consists of video recorded conversations among 42 informants aged between 20 and 82, from three separate regions in Sweden. With 14 % of the corpus having been annotated with glosses for signs, it comprises total of approximately 3,600 different signs occurring about 25,500 times (tokens) in the 42 annotated sign language discourses/video files. As these two resources sprang from different starting points, they are independent from each other; however, in the late phases of building the corpus the importance of combining work from the two became evident. This presentation will show the development of these two resources and the advantages of combining them.

Keywords: Swedish Sign Language, sign language dictionary, sign language corpus, web-based sign language resources

1. Introduction

In Sweden, a sign language resource, namely the online Swedish Sign Language Dictionary, henceforth SSLD, is frequented by language learners, schools, teachers, curious visitors, researchers, and annotators. Not only has it been on the ten top list of Stockholm University websites, but it also ranks high among more general educational resources in *Skollink*. Another forthcoming resource, The Swedish Sign Language Corpus, henceforth SSLC, will also be an important resource, but it will have different uses than the dictionary. Created during the years 2009–2011, the corpus consists of video recorded conversations of 42 informants aged between 20 and 82, from three regions in Sweden (see Mesch & Wallin, this volume).

The two resources have had different starting points and sources and have also followed different timelines. Work on the SSLD began many years ago and is still under development, while the SSLC was created according to other criteria and includes semi-spontaneous dialogues and narratives that the dictionary does not have (Mesch, Wallin, Nilsson & Björkstrand, 2010). Although they are clearly independent of each other, in the late phase of corpus creation, we discovered areas in which the SSLD and the SSLC could be improved if they were to work in cooperation.

2. Online Swedish Sign Language dictionaries

The first version of the online dictionary was created in 2001; it was called the *Digital version of the Swedish Sign Language Dictionary* and included 3,132 sign entries. This was the result of lexicology work initiated in 1988 by the Sign Language section of the Department of Linguistics at Stockholm University. This dictionary includes video files for signs taken from a printed

dictionary of the Swedish National Association of the Deaf (1997). Then, nine databases with specialized lexicons in Swedish Sign Language, with a total of 4,300 sign entries, were created during the years 2004–2011. The *Swedish Sign Language Dictionary*, the SSLD, was created in 2008 and has been in development since. Today, it has approximately 8,000 dictionary sign entries, including parts of the specialized databases mentioned above (see figure 1). The lexicographic work group documents Swedish Sign Language vocabulary and continuously updates the web-based dictionaries. All these databases are created as FileMaker Pro database files and are available online at <http://www.ling.su.se/teckensprak>

The SSLD has four versions: a) an Internet version, b) a Tecklex (simpler version), c) a mobile version, and d) a mobile app for the iPhone (and later for the Android).

Databases	Year	Number of signs	Variants	Number of sentences
Swedish Sign Language Dictionary, SSLD	2008–12 Dec 2011	7,905	1,370	2,188
Digital version of Swedish Sign Language Dictionary	2001–2008	3,132		3,132
Specialized databases				
Signs for sport	2011	1334	133	
Signs used in health care settings	2010	911		-
Signs with incorporated numerals	2008	179		(581 video clips)
Signs to talk about sex and family planning	2007	391		-
Signs for religious concepts	2006	568		-
Swedish towns and provinces	2006	515		-
Signs for linguistic concepts	2005	414		-
Signs used when playing bridge	2005	370		-
Mathematical concepts	2004	469		-

Figure 1. Lexicographic SSL databases

Each sign entry consists of four tabs: sign, variant, example (sentence), and photo and sign notation. In the tab for 'sign', the sign demonstration is shown with a description of the shape and keyword(s) (see figure 2).

From there you can also click through to the other signs of the same forms but with other meanings (cf. homonyms) or to other signs with the same meaning (cf. synonyms). In the tab for 'variant', signs are shown as one or more variants, for example, a sign can be performed by one or two hands, or it may be performed as a compound sign. It also contains the form description of the variant. In the tab 'example', there are one or more usage examples provided as sentences in which the sign is used for the translation into Swedish (see figure 3). In the tab for 'photo & notation', there is a sequence of one or more still photos of sign demonstration (for example, for printing) and notation.



Figure 2. Sign entry in the SSLD

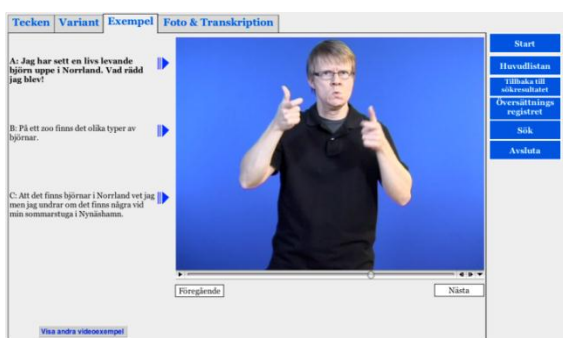


Figure 3. Example usage in sentences in the SSLD



Figure 4. Searching fields in the SSLD

As shown in Figure 4, there is a variety of search paths one can take. These paths are a) Swedish word; b) subject, which is a list of signs associated with a specific domain, such as technology, family and clothing; c) Swedish words in translated sentences, which can be selected to show the dictionary user how specific signs are used in sentences; d) numeral, which consists of a list of numeral sign or sign with incorporated number; e) ID-number, for those cases in which the user knows the number of the sign entry; f) manual alphabet, which displays the Swedish manual alphabet as a list; g) place of articulation; i) handshape for one or two hands; j) unusual/old signs

that exist but are not in common use today (often signs used by older generations, but not younger ones); k) regions, in which signs are divided by region of location and former location of deaf schools; l) sign language phrases such as 'good morning', 'how are you?' or 'cannot afford'; and m) fingerspelling, which consists of fingerspelling signs, fingerspelling affixes and fingerspelling parts in sign compounds.

Each sign entry consists of a unique identification number, and each new entry is assigned a unique number. This makes it possible for a sign to be referred to via phone, e-mail or written communication.

When searching on handshapes or place of articulation, there are several approaches that can be used. If the search is for a sign through handshape, users can answer a prompt that asks for the number of hands and, in the case of two hand signs, whether both hands are active, or just one. After users make their selection, another box is presented, and the user can still choose to see the sign list (figure 5), or search further, for instance, for a specific handshape. If a search for place of articulation is conducted, then a box with a list of the different places of articulation appears. After it is chosen another box appears and asks for the number of hands followed by a new box where the user can choose to see the sign list or go on to look at handshape.

Avsluta	Start	Sök	Sökresultat: 81	Utskrift	Mitt Freläsvet	Säkers 1, 2, 3...	Säkers 4, 5, 6...	INFO
			05489 listig	Kvädlinget, uppåtriktat och framsvängt, kontakt bredvid erat, förs sedan framåt samtidigt som det alla framåt				
			05490 listig	Kvädlinget, vänstertriktat och inåtriktat, kontakt bredvid kinden, förs nedåt samtidigt som det alla nedåt				
			05492 komedi	Kvädlinget, vänstertriktat och inåtriktat, förs nedåt med bibeållan kontakt med kinden, uppåtriktat				
			05494 raka sig	Kvädlinget, uppåtriktat och framsvängt, förs nedåt med bibeållan kontakt med kinden, uppåtriktat				
			05495 rakbyvel	Kvädlinget, uppåtriktat och framsvängt, förs nedåt med bibeållan kontakt med kinden, uppåtriktat				
			05496 går inte	Kvädlinget, uppåtriktat och vänstertriktat, förs åt vänster med medial kontakt med munnen samtidigt som det alla åt vänster				
			05497 före inte	Kvädlinget, uppåtriktat och vänstertriktat, förs åt vänster med medial kontakt med munnen samtidigt som det alla åt vänster				
			05498 omöjlig	Kvädlinget, uppåtriktat och vänstertriktat, förs åt vänster med medial kontakt med munnen samtidigt som det alla åt vänster				

Figure 5. Sign list in the SSLD

We have looked at which of the paths are used by most SSLD users, in January 2012. Of the top five, Swedish Words is the most common search path, followed by Subject, Hands, Swedish Words in Translation Sentences, and Place of Articulation (see table 1).

Searching field	hits
Swedish word, e.g. mindre	32536
Subject, e.g. sport	1570
Handshape, e.g. hooked finger hand	506
Swedish word in translated sentences, e.g. mindre	416
Place of articulation, e.g. nose	368

Table 1. The most common search paths for SSLD users

3. Upcoming: Swedish Sign Language Corpus

The Swedish Sign Language Corpus, or SSLC, (2009–2011) is a project being carried out by the Department of Linguistics, Sign Language Section, Stockholm University, and is funded by *Riksbankens Jubileumsfond*. The aim of the project is to publish an accessible collection of sign language discourses, or a 'corpus' with (Swedish) glosses and a translation into Swedish. It is intended to provide an accurate impression of what Swedish Sign Language sentences look like and contribute new signs and variants of signs for the SSLD on the web. Making Swedish Sign Language discourses accessible like this means the corpus can also be used to develop teaching materials for Swedish Sign Language, and it will offer the opportunity to show or analyze at the level of a specific sign, groups of sentences, or an entire chunk of discourse when teaching sign language. Corpus-based studies will be of major importance to future research in sign linguistics, making it possible to analyze Swedish Sign Language grammar, as well as to facilitate research in such areas as the sociolinguistics of sign language and translation studies.

All of the corpus material has been edited, but only 14% of the 25 hours of material we have has been annotated with (Swedish) glosses and a translation into Swedish to date; approximately 2 hours and 30 minutes have so far been annotated with the annotation software ELAN (Crasborn, et al, 2008) (see figure 6). The annotation work is very time-consuming. During the project, the transcription conventions for the corpus work were supplemented with instructions for annotating polysynthetic signs (Wallin, Mesch & Nilsson, 2011), and they will be updated again in the future.

We are currently investigating which web portal will be best suited for the corpus material. The aim is to publish the first fully annotated video files, with (Swedish) glosses and a translation into Swedish during 2012 (Mesch, Wallin, Nilsson, Bäckström, Johansmide & Bergman, 2012).

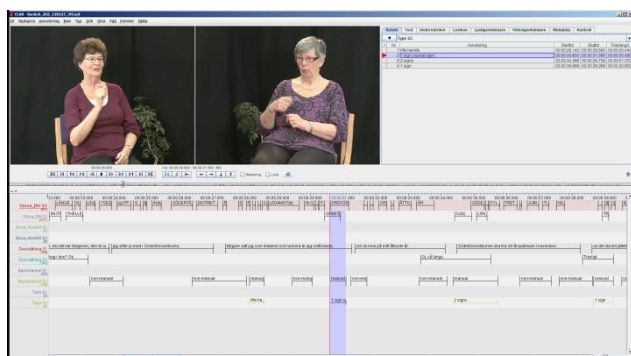


Figure 6. An example illustrating conversation materials in the SSLC with ELAN

4. Two resources for different uses

As shown in this paper, there are two different sign language resources, the SSLD and the SSLC. Some opportunities for searching these resources are presented here.

The SSLD has many search paths: Swedish word, subject, Swedish words in translated sentences, numeral signs, ID-number, manual alphabet, place of articulation, handshape, old signs, regional signs, and phrases, fingerspelling. The SSLD shows signs, variants, constructed sentences and still photos with sign notation.

The SSLD has approximately 8,000 sign entries. When a search is conducted for a sign that is made with one hand, the user learns that there are 3,732 one-handed signs (47.21%) found in the SSLD. The user can then sort his search results by selecting a handshape from the pictures; for example, a hooked finger hand is one of the shapes that is indexed. This selection reduces the sign list to 81 signs produced with that handshape (see, for example, figure 5).

The SSLC is also an important resource for searching for signs, but it is a very different resource than the SSLD.

Although the raw material has been compiled, work on the SSLC is ongoing. With 14% of the corpus having been annotated with glosses for signs, it comprises approximately 3,600 signs (including compounds) occurring a total of approximately 25,500 tokens in 42 annotated sign language discourses/video files (dated 1 March 2012). One of the aims of the SSLC is to give users an accurate representation of Swedish Sign Language sentences. The SSLC has, for example, some other possibilities for investigating such things as the frequency of signs in the sign language discourses and obtaining a concordance view (figure 7) that are not possible for the SSLD.



Figure 7. A view of the concordance view in the SSLC (with ELAN)

5. Combined resources

As described earlier, a user can search for words with Swedish translations in the SSLD. However, one cannot search for a sign in such sentences (cf. Ordbog over Dansk Tegnsprog). Thanks to work done on the SSLC, we are going to create a new path in the SSLD. The SSLD learns from the SSLC's annotation system with glosses. All example sentences will be annotated with glosses, and by using these glosses, users can then look for signs in the examples that come in different sign entries. In this way, users can get an idea of how a sign is used in different contexts in the SSLD.

Although the SSLD has nearly 8,000 signs, it is missing many signs that can be found in the SSLC, particularly signs with genuine mouth movements (henceforth genuine signs). They are not always as easy to find as signs with Swedish-influenced mouth movements. Looking at a Swedish word, one can associate with a sign but not with genuine signs; rather, it is in the context where these signs appear that is important. This problem may be overcome by using the SSLC when transcribing signs from the corpus material. A corpus annotator has a sign but not a good gloss on this sign. When he is looking for a sign in the SSLD, he finds that this sign is not documented. Lexicographers of the SSLD discover what is missing and then can use material from the SSLC to fill in these gaps, and thus complement their work.

Another task that is currently underway with the SSLD is enhancing the example sentences used in the dictionary. Although many of these are fictitious, they are still acceptable; however, a greater issue is that some nouns can become quite overused in the examples. For example, too many sentences include the signs 'children' and 'friends'. With the help of the SSLC, lexicographers can create more variation in the example sentences and thus make them less repetitious.

Another problem with such sentences is that they are in monologue form, in that there is an actor who demonstrates signs. This creates a problem for the actor in how to demonstrate signs that are typically used in dialogues. We will modify and implement such sentences with two actors.

Additionally, a new resource, a web site pertaining to grammatical information, will be added. There are also plans to add useful links to the SSLD, and for this, technical solutions will need to be studied.

The next task for the project will be to create study material of sign language structure, using both of these resources. The material should include in sign description of the sign structure (how one or two hands perform a movement in a particular position), sign formation processes, parts of speech, sentence formation, spatial organization and variation in language use. The material is supposed to consist of short texts with hyperlinks to the

video examples. It will be available on the web portal of the Department of Linguistics probably in the beginning of 2013, with links to the SSLD and the discourses of SSLC.

6. Conclusion

Work on both resources, the SSLD and the SSLC, is ongoing. The two language resources are independent from each other, but during a late phase in building the corpus, the importance of combining and this work and advising each other became apparent. And since both projects reside in the same building, there are many opportunities for collaboration.

The two (as well as a third) resources of Swedish Sign Language are of recognized value to not only both educational and sign language linguistic research, but to users of the resources, e.g. parents and second language learners. Thus, these are a part of the development of a focus on teaching methods in sign language. Pilot material is intended for educators in sign language, teacher students and sign language instructors.

7. Acknowledgements

This work was supported in part by Riksbankens jubileumsfond (In2008-0276-1-IK) and by financial support from 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.
- Mesch, J., Wallin, L., Nilsson, A-L. & Björkstrand, T. (2010). Corpus and Dictionary – separate or together? Poster presented at Sign Linguistics Corpora Network Workshop 4: Exploitation, December 3, 2010, Berlin.
- 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>
- Skollink. <http://www.skollink.se>
- Svenskt teckenspråkslexikon (2008 -). [Swedish Sign Language Dictionary] On-line publication containing also the specialized databases <http://www.ling.su.se/teckensprakslexikon>
- Svenskt teckenspråkslexikon. (1997). Sveriges dövas riksförbund.
- Wallin, L., Mesch, J. & Nilsson, A-L. (2011). Transkriptionskonventioner för teckenspråkstexter. Version 3. Stockholms universitet: Institutionen för lingvistik.