

Compiling the Slovene Sign Language Corpus

Špela Vintar, Boštjan Jerko, Marjetka Kulovec

University of Ljubljana, Dept. of Translation Studies

Aškerčeva 2, SI - 1000 Ljubljana

E-mail: {spela.vintar, bostjan.jerko, marjetka.kulovec}@ff.uni-lj.si

Abstract

We report on the project of compiling the first corpus of the Slovene Sign Language. The paper describes the procedures of data collection, the decisions regarding informant selection and plans for transcription and annotation. We outline the particularities of the Slovene situation, especially the high variability of the language, issues concerning language competence and the attitudes of the deaf community towards such data collection. At the time of writing, the data collection stage is nearly finished with over 70 recorded persons, and transcriptions with iLex are underway. The aim of the project is to use the corpus for explorations into the grammatical properties of SSL.

Keywords: Slovene Sign Language, corpus compilation, sign language transcription

1. Introduction

We are presenting SIGNOR (<http://lojze.lugos.si/signor>), a project to collect and annotate samples of the Slovene sign language (SSL) from natural signers. This is the first such endeavor for SSL, because so far all projects dealing with SSL were aimed at recording individual signs, compiling normative dictionaries and describing isolated grammatical aspects of SSL. The SIGNOR project aims to compile a representative corpus of SSL using natural signers, and then transcribe and annotate the video data to get the information needed for describing the lexicon and grammar of the language.

The project consortium includes the Faculty of Arts as the leading partner and the Slovenian Academy of Sciences and Arts as the second partner. The deaf community is represented informally through two of the researchers, of which one is a natural signer and the other a CODA and a certified SSL interpreter, but also through the support of the three key institutions for the deaf in Slovenia: Deaf and Hard of Hearing Clubs Association of Slovenia, School for the Deaf Ljubljana and the Association of Slovene Sign Language Interpreters.

The association of deaf clubs (ZDGNS)¹ is the most influential institution for the deaf and hard of hearing in Slovenia and comprises various activities related to SSL and SSL education, including the compilation of an online SSL dictionary.

2. Slovene Sign Language - background

The Slovene deaf community is estimated at between 700 and 1600 members. The exact number of the deaf is difficult to find for various reasons; some people may refuse to use the nationally provided voucher system for interpreting and thus remain "invisible", others may have become deaf at a later stage in their lives and are therefore not included in official statistics, and yet others may prefer not to be associated with the deaf community at all.

Just like elsewhere in the world, being deaf does not equal sign language user and vice versa; many deaf

people have learned to communicate primarily through lip reading and speaking, and many sign language users are hearing children of deaf parents or simply hearing users of SSL, for whatever reason.

The systematic development of SSL can be traced back to the 1970s, when Slovenia was still part of Yugoslavia and SSL was sporadically taught at seminars and courses, mostly organised by the Association (ZDGNS). Systematic activities related to interpreting have started in the mid-1980s and have resulted in the first interpreters' examination taking place in 1986 and yielding 16 new interpreters.

The public awareness of sign language as the language of the deaf started to develop after 1980, when the first TV show for the deaf was broadcast by TV Koper. Still, the general attitude towards SSL in educational institutions remained sceptical, with very heavy bias towards "inclusion"; the practice of integrating deaf children into regular schools via lip reading and speaking Slovene.

The situation changed considerably after the Act on the Use of Sign Language was adopted in 2002. The act acknowledged the fact that SSL is one of the indigenous languages in Slovenia and institutionalised the right of the deaf to use SSL in all public and private situations, and their right to use interpreters in all public situations, whereby a certain amount of interpreting services is funded by the government through a system of vouchers. The act further installs the Council of Slovene Sign Language, which is composed of members of different institutions and which should primarily monitor and enhance the development of SSL and the training of SSL interpreters.

While this system may have provided deaf people access to many public services previously unavailable to them, it remains largely insufficient in providing equal opportunities in education. Sign language is taught only at two schools in Slovenia, and interpreting services required by deaf students largely exceed the hours financed by the government. As a consequence, there are currently only about 20 deaf students in Slovenia, and specialized vocabularies are severely underdeveloped in SSL.

¹ <http://www.zveza-gns.si/>

3. Data collection

The project intends to record between 80 and 100 informants, whereby we shall ensure the representativity of the sample by including all 13 deaf clubs in Slovenia and by selecting informants on the basis of a survey of the entire deaf community. Recordings are already underway throughout Slovenia; at the time of writing we have recorded 72 informants.

In order to be able to balance the corpus and explore the sources of variability we ask each informant to provide some personal information, whereby we follow strict data protection procedures.

The personal information we are collecting from the signers include:

- When and how did deafness develop
- Age and gender
- Primary hand
- Education level
- Place and region of birth
- Place and region of education

2.1 Recording sessions

The recording sessions are composed of three parts. The first part is the informant's free signing about their life and family. This part serves as an ice-breaker and is often in the form of a dialogue between the interviewing student and the informant, the goal is to help the informant relax and get used to the camera. The second part of the session is recorded after the informant has watched an elicitation video on a general topic (e.g. politics, body, travel etc.). The last part of the recording is aimed at collecting more specialized vocabulary and can be either free narration if the signer has a favorite subject (such as a specific hobby or sport), otherwise another elicitation video is used. The videos used contain little speech and show situations from various general and specialized topics. Spoken or written Slovene is avoided in elicitation videos because such input might influence the signer in their language use.

The recording sessions are performed by deaf or hard of hearing students or CODAs. Experience gained so far shows that much better responses are obtained if the interviewer is deaf or hard of hearing. It seems that it is much easier for the informants to relax and sign spontaneously if the interviewer is an equal partner in the conversation.

2.2 Field observations

The organisation of recording sessions is performed with the help of local deaf clubs. At the beginning of the project we organised a presentation event on the premises of the association of deaf clubs, where the goals of the project and the plans for data collection were presented to local presidents. The responses of the deaf community were cautious. It became clear from the questions and comments that some fears were related to the fact that this was the first time an academic institution launched a project on the topic of SSL, and that the deaf community felt this as an unwelcome intervention or an attempt to "prescribe" or "forbid" certain SSL usage. Having made clear that the aim of the

project was primarily to describe the language as it is currently used, there was again some disappointment due to the fact that certain SSL users effectively wish for a certain level of standardisation to occur, for purely practical reasons.

After the plenary presentation of the project we e-mailed each local deaf club a presentation leaflet and asked them to help us by providing contacts to their local members. The organisation of each individual recording session, communication with the informants and the actual interviewing and recording, were performed exclusively by deaf students. Some sessions took place on the premises of the local deaf club, while others took place at the informants' place of residence. The recordings of high school pupils at the School for the Deaf Ljubljana were performed at the school premises, whereby a signed permission was obtained from each informant's parents.

Despite some initial mistrust co-operation with the local deaf clubs, the school for the deaf and the association ZDGNS was and continues to be excellent. We particularly wish to thank all informants who participated so far, because they did so on a purely altruistic basis and received no compensation of any kind.

So far we have collected data from 72 informants from different Slovenian regions. We may have to discard some material either due to some informants' inability to relax and sign naturally, or due to insufficient SSL competence of some signers.

Especially the latter issue seems difficult to delineate, because in our population SSL has been acquired in different ways and at various ages. Like in many other societies worldwide, sign language has not been systematically encouraged or taught in Slovenia until relatively recently. The older deaf generation received no schooling in SSL whatsoever and were either linguistically neglected or vigorously taught to speak and lipread. In the younger generation there are large differences with regard to the place of schooling, because there is still only a single school in Slovenia where SSL is systematically taught: the School for the Deaf Ljubljana.² Of course the extent of deafness also plays a role. For pragmatic reasons we adopted the position that for the purposes of our corpus a competent SSL user is anyone who frequently uses SSL to communicate with other SSL users. Such signers are considered to be adequate informants to our project and we make no further distinctions on the basis of linguistic competence.

4. Corpus annotation and processing

For the annotation and transcription of video data we have selected iLex (Hanke and Stolz 2008), a powerful and versatile tool providing for multi-tiered annotation. Since there is no real precedence for SSL annotation in Slovenia, numerous questions arise even before the first video has been processed.

The first annotation stages include segmentation

² SSL is taught as an independent school subject only at the secondary school level; it may be partly included into the course Communication skills at the primary school level.

(tokenization) and glossing (lemmatization). Apart from these we also intend to annotate the oral pronunciation (mouthing), and we plan to add the HamNoSys transcription at least for a part of the corpus (Schmaling and Hanke 2001). A translation into Slovene will also be provided, and for selected segments of the corpus we shall perform tests of inter-annotator agreement.

Similar to other sign languages, Slovene sign language uses modifications of signs to express syntactic and semantic relations between items in a sentence. Thus, the base form of the sign TEACH can be modified in various ways to signify *teacher* (male or female), *I teach* (1st person, active voice), *I am being taught* (1st person, passive voice) and so on. Transcribing such tokens involves deciding whether *teach* and *teacher* are to be considered forms of a single lexeme or two separate lexemes, whether the sign for *female teacher* should be tokenized as one sign or two and other similar dilemmas. SSL has a relatively poor specialized vocabulary, therefore many specific topics need to be signed "creatively", using general signs in new contexts and with new meanings. The mouthing accompanying these creative uses is necessary to infer the intended meaning, but in transcribing such signs we again need to decide whether this creative use constitutes a new lexeme or not.

Another difficult issue is the interplay between SSL and the so-called "signed Slovene", a direct transposition of spoken Slovene into signs. While most sign language users agree that signed Slovene is an artificial construct that is never used in spontaneous conversations among deaf people, it clearly influences the development of SSL in many ways simply because it is commonly seen on national television. Thus, certain signs, such as those for copula verbs and conjunctions like "and", may be used more frequently in types of discourse more influenced by signed Slovene. The impact of Slovene on SSL syntax has yet to be empirically proven, but it is believed to be considerable.

We know that many of these issues have been described - and some successfully resolved - by other researchers. We are aware of a large body of previous research in many sign languages of the world, and we plan to lean primarily on those bordering on Slovenia; Austrian (Krammer et al. 2001, Dotter 2011), Italian (Prinetto et al. 2011), Croatian (Tarczay 2010); as well as those with exceptional influence within Europe such as German (Konrad et al. 2003; Konig et al. 2008) and worldwide such as Australian (Johnston et al. 2006). We hope that indirectly the project will also have an impact on the ethical dimension of SSL use and the perception of deafness in our society.

5. Conclusions and future work

Since we are describing a relatively young and small-scale project, there are currently few conclusions and substantial future work. First of all we plan to finish collecting the materials and in particular proceed with

the transcription, as this activity alone generates fundamental theoretical questions. Transcribing will be performed with iLex primarily by the project members, and in certain stages the deaf students will also be involved.

Next we intend to provide some frequency data on the basic SSL vocabulary, which could be used to update the current SSL dictionary and, in particular, the currently used textbooks. Our next aim is a basic description of SSL grammar, in particular the syntactical structure and the use of spatial placeholders. We plan to experiment with computational techniques such as Machine Learning to infer grammatical rules.

On a yet another level we hope to answer some sociolinguistic questions related to SSL and the factors influencing its development. Some of the questions we plan to explore include the role of education in general and the educational institute in particular, as there is currently only a single school in Slovenia teaching sign language; the development of sign "slang" among youngsters and the impact of other cultures; and the issue of regional/social/age-related variation in SSL use and the perceived need for standardization within the community.

6. References

- Dotter, F. (2011) Sign Languages and Their Communities Now and in the Future. "Multilingualism in Europe. Prospects and Practices in East-Central Europe", Budapest, March, 25-26, 2011
- Hanke, T. and Stolz, J. (2008). iLex – A Database Tool For Integrating Sign Language Corpus Linguistics and Sign Language Lexicography, LREC 2008, May 28 – May 30 2008, poster
- Johnston, T. and A. Schembri (2006). "Issues in the creation of a digital archive of a signed language". In L. Barwick and N. Thiesberger (eds.) *Sustainable data from digital fieldwork*. Sydney: Sydney University Press, 7-16.
- König, S., R. Konrad and G. Langer. (2008). "What's in a sign? Theoretical lessons from practical sign language lexicography". In J. Quer (ed.) *Signs of the time. Selected papers from TISLR 2004*. Hamburg: Signum, 379-404.
- Konrad, R., A. Schwarz, S. König, G. Langer, T. Hanke and S. Prillwitz (2003). *Fachgebärdenlexikon Sozialarbeit/Sozialpädagogik*. Hamburg: Signum. Available at: <http://www.sign-lang.uni-hamburg.de/slex/> (accessed: 30 September 2009).
- Krammer, K., Bergmeister, E., Dotter, F., Hilzensauer, M., Okorn, I., Orter, R. and Skant, A. (2001). The Klagenfurt database for sign language lexicons. *Sign Language and Linguistics*, vol. 4, issue 1, pp. 191-201
- Prinetto, P. et al. (2011). The Italian Sign Language Sign Bank: Using WordNet for Sign Language corpus creation, ICCIT, 2011 International Conference, 29-31 March 2011, pp. 134-137

- Schmaling, C. and T. Hanke (2001). *HamNoSys 4.0*. Available at: <http://www.sign-lang.uni-hamburg.de/Projekte/HamNoSys/HNS4.0/englisch/HNS4.pdf> (accessed: 30 September 2011).
- Tarczay, S. (2010.), Pretpostavke profesionalizacije prevoditelja znakovnoga jezika za gluhe i gluhoslijepe osobe. Magistarski rad. Zagreb: Edukacijsko-rehabilitacijski fakultet Sveučilišta u Zagrebu