Workshop Proceedings

7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining

Language Resources and Evaluation Conference (LREC)
Portorož, Slovenia, 28 May 2016

Editors and Workshop Organizers

Eleni Efthimiou	Institute for Language and Speech Processing, Athens GR
Stavroula-Evita Fotinea	Institute for Language and Speech Processing, Athens GR
Thomas Hanke	Institute of German Sign Language, University of Hamburg, Hamburg DE
Julie Hochgesang	Gallaudet University, Washington US
Jette Kristoffersen	Centre for Sign Language, University College Capital, Copenhagen DK
Johanna Mesch	Stockholm University, Stockholm SE

Workshop Programme Committee

Penny Boyes Braem	Center for Sign Language Research, Basel CH
Annelies Braffort	LIMSI/CNRS, Orsay FR
Onno Crasborn	Radboud University, Nijmegen NL
Athanasia-Lida Dimou	Institute for Language and Speech Processing, Athens GR
Sarah Ebling	Institute of Computational Linguistics, University of Zurich, Zurich CH
Eleni Efthimiou	Institute for Language and Speech Processing, Athens GR
Michael Filhol	CNRS–LIMSI, Université Paris-Saclay, Orsay FR
Stavroula-Evita Fotinea	Institute for Language and Speech Processing, Athens GR
Thomas Hanke	Institute of German Sign Language, University of Hamburg, Hamburg DE
Julie Hochgesang	Gallaudet University, Washington D.C. US
Matt Huenerfauth	Rochester Institute of Technology, Rochester, NY US
Hernisa Kacorri	City University New York (CUNY), New York NY US
Athanasios Katsamanis	Computer Vision, Speech Communication and Signal Processing Group, National Technical University of Athens, Athens GR
Jette Kristoffersen	Centre for Sign Language, University College Capital, Copenhagen DK
John McDonald	DePaul University, Chicago IL US
Johanna Mesch	Stockholm University, Stockholm SE
Carol Neidle	Boston University, Boston MA US
Rosalee Wolfe	DePaul University, Chicago IL US

Table of Contents

Author Index

Editors' Preface

This collection of papers stems from the Seventh Workshop on the Representation and Processing of Sign Languages, held in May 2016 as a satellite to the Language Resources and Evaluation Conference in Portorož.
While there has been occasional attention for sign languages at the main LREC conference, the main focus there is on spoken languages in their written and spoken forms.
This series of workshops, however, offers a forum for researchers focussing on sign languages.
For the fifth time, the workshop had sign language corpora as its main topic, however not surprisingly, since during the past years, sign language corpora became a major trend in sign language research. This time, the focus was on corpus mining.
Once again, the papers at this workshop clearly identify the potentials of even closer cooperation between sign linguists and sign language engineers, and we think it is events like this that contribute a lot to a better understanding between researchers with completely different backgrounds.

The contributions composing this volume are presented in alphabetical order by the first author. For the reader's convenience, an author index is provided as well.

We would like to thank all members of the programme committee who helped us reviewing the submissions to the workshop within a very short timeframe!

Finally, we would like to point the reader to the proceedings of the previous workshops that form important resources in a growing field of research:

- O. Streiter & C. Vettori (2004, Eds.) *From SignWriting to Image Processing. Information techniques and their implications for teaching, documentation and communication.* [Proceedings of the Workshop on the Representation and Processing of Sign Languages. 4th International Conference on Language Resources and Evaluation, LREC 2004, Lisbon.] Paris: ELRA. Available online at http://www.lrec-conf.org/proceedings/lrec2004/ws/ws18.pdf
- C. Vettori (2006, Ed.) *Lexicographic Matters and Didactic Scenarios.* [Proceedings of the 2nd Workshop on the Representation and Processing of Sign Languages. 5th International Conference on Language Resources and Evaluation, LREC 2006, Genova.] Paris: ELRA. Available online at http://www.lrec-conf.org/proceedings/lrec2006/workshops/W15/Sign_Language_Workshop_Proceedings.pdf
- O. Crasborn, E. Efthimiou, T. Hanke, E. Thoutenhoofd & I. Zwitserlood (2008, Eds.) *Construction and Exploitation of Sign Language Corpora.* [Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages. 6th International Conference on Language Resources and Evaluation, LREC 2008, Marrakech.] Paris: ELRA. Available online at http://www.lrec-conf.org/proceedings/lrec2008/workshops/ W25_Proceedings.pdf
- P. Dreuw, E. Efthimiou, T. Hanke, T. Johnston, G. Martínez Ruiz & A. Schembri (2010, Eds.) *Corpora and Sign Language Technologies*. [Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages. 7th International Conference on Language Resources and Evaluation, LREC 2010, Valletta, Malta.] Paris: ELRA. Available online at http://www.lrec-conf.org/proceedings/lrec2010/workshops/W13.pdf
- O. Crasborn, E. Efthimiou, S.-E. Fotinea, T. Hanke, J. Kristoffersen, J. Mesch (2012, Eds.) *Interaction between Corpus and Lexicon*. [Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages. 8th International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey.] Paris: ELRA. Available online at http://www.lrec-conf.org/proceedings/lrec2012/workshops/ 24.Proceedings_SignLanguage.pdf

- O. Crasborn, E. Efthimiou, S.-E. Fotinea, T. Hanke, J. Hochgesang, J. Kristoffersen, J. Mesch (2014, Eds.) *Beyond the Manual Channel*. [Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages. 9th International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland.] Paris: ELRA. Available online at http://www.lrec-conf.org/proceedings/lrec2014/workshops/LREC2014Workshop-SignLanguage%20Proceedings.pdf

The Editors

The SIGNificant Chance Project and the Building of the First Hungarian Sign Language Corpus

Csilla Bartha, Margit Holecz, Szabolcs Varjasi
Research Centre for Multilingualism, Research Institute for Linguistics of the Hungarian Academy of Sciences, 1068 Budapest, Benczúr u. 33.
E-mail: {csillabarthadr, holeczmargit, varjasi.szabolcs}@gmail.com

Abstract

The Act CXXV of 2009 on Hungarian Sign Language and the Use of Hungarian Sign Language recognizes Hungarian Sign Language (HSL) as an independent natural language, moreover it provides the legal framework to introduce bilingual education (HSL-Hungarian) in 2017. In order to establish the linguistic background for bilingual education it was crucial to carry out linguistic research on HSL, which research should be sociolinguistically underpinned and should include corpus-based research. This research also aims to standardize HSL for educational purposes with the highest possible degree of community engagement.
During the SIGNificant Chance project a sign language corpus (approximately 1750 hours) was created. A nation-wide fieldwork was conducted (five regions, nine venues). 147 sociolinguistic interviews and 27 grammatical tests (with 54 participants) were recorded in multiple-camera settings. There were also Hungarian competency tests and narrative interviews conducted with selected participants in order to make the complex description of their different linguistic practices in different discursive contexts possible.
We are using ELAN and three different templates to analyze the collected data for different purposes (sociolinguistic-grammatical template, another for short term project purposes, and one for the dictionary). Some parts of the annotation work has been finished which contributed to the writing of the basic grammar of HSL and the creation of a small corpus-based dictionary of HSL.

Keywords: Hungarian Sign Language, sociolinguistics, SIGNificant Chance Project, corpus building, annotations

1. Introduction

According to some estimates there are 30-40.000, based on other data there are 60.000 D/deaf people living in Hungary which makes them the third largest linguistic minority in Hungary using the Hungarian Sign Language (further on: HSL) as their primary language (Bartha, 2004). The Act CXXV of 2009 on Hungarian Sign Language and the Use of Hungarian Sign Language was an important milestone for the D/deaf community (Bartha et al., 2016). On the one hand, because it recognized HSL as an independent natural language and on the other hand it provides the legal framework for introducing bilingual education in 2017. However, for bilingual education not only theoretical linguistic and applied linguistic researches are necessary but also the standardization of HSL for the purpose of education. We are aware that the notion of standardization (cf. WFD 2014) is quite problematic, however, we conceive standardization as a bottom-up, corpus-based process which is built on data acquired from sociolinguistic sampling. Standardization in our understanding presupposes the widest possible consensus with (and also the involvement of) the signing community and should recognize the diversity of the respective sign language(s). In our bottom-up approach we believe that the standardization process should be based on involving members of the Deaf community. For the linguistic foundation it is essential to have corpus-based, empirical researches applying current sociolinguistic approaches. This is the main goal of TÁMOP 5.4.6/B-13/1-2013-0001 project called Theoretical and practical steps of the standardization of Hungarian Sign Language (SIGNificant Chance).

2. The SIGNificant Chance Project

The corpus created during the project has the following aims: 1. sociolinguistic description of the patterns of language use in the Deaf community 2. analysis of HSL variation and creating an evidence- and corpus-based digital dictionary 3. to provide the grammatical description of the emic categories of HSL 4. measuring competencies of Deaf children in special and mainstream education. Besides providing basic statistic and linguistic (on all linguistic levels) analysis, it also serves educational purposes, like creating educational materials.
The project was materialized between 1st November 2013 and 31st October 2015 involving experts from several fields. Sociolinguists, theoretical linguists, psychologists, sociologist, lawyers and IT professionals helped us among others. Altogether 35 Deaf, hard of hearing and CODA people worked with our colleagues, supporting each other to reach the common goal, under the supervision of Csilla Bartha.

3. The Hungarian Sign Language corpus

3.1 Significance and constitution of the corpus

Recently there are more and more corpus-based sign language researches and descriptions, however, even on international level it is unique to study sign language phenomena using such a vast (more then 1700 hours) corpus which is extremely well documented by sociolinguistic metadata. The participants were selected by applying strict statistical criteria. While there are recent sociolinguistic studies on corpora (see eg. Schembri et al. 2013), in Hungary, there have only been sporadic sign language researches, which were usually based on a small

amount of samples and the analysis was done in a word processor.
In the SIGNificant Chance Project, informants of all the interviews were recorded in the same communicational situation, talking about the same topics; all this provides a unique analysis foundation compared not only to the sign language, but also to the spoken language researches. Since we have comprehensive metadata about the informants, the corpus provides a unique opportunity to conduct qualitative and quantitative analysis as it includes information about gender, age, location, language socialization and many other variable. During our research, each phenomena was analyzed with a bottom-up approach.
For methodological reasons it was necessary to create the sessions without any influence of a hearing person or the Hungarian Spoken Language, therefore no hearing person was allowed in the studio. The fieldworkers of the sociolinguistic interviews, as well as the moderators of the grammatical tests were D/deaf.
During the project it was important to include the Deaf Community: they did not only passively help our hearing colleagues as consultants, they played significant roles as professional supervisors, fieldworkers, annotators, etc.
Together with the Deaf Community, applied linguists, sociolinguists, sign linguists and experts of other fields formed a learning community, all this; spontaneously contributed to raising the prestige of sign language, awakening and increasing the interest towards sign language among the experts involved and the university students. The process of corpus planning and building created a Deaf and Hearing learning community in which our critical and reflexive approach manifested. Furthermore, the level of their metalinguistic and metapragmatic consciousness increased significantly.
The corpus is made of two big parts: sociolinguistic interviews (chapter 3.2) and grammatical tests (chapter 3.3) and the related recordings. During the project recordings were made of each training sessions, including trainings for the fieldworkers and the annotators, workshops and conferences. We used them for documentation and for the purpose of meta analysis. Thus we created a corpus containing a vast amount of material: 1750 hours of recording, which is 6,5 terabyte data.

3.2 Sociolinguistic interviews

In nine venues of 5 regions (Budapest, Szeged, Hódmezővásárhely, Békéscsaba, Debrecen, Kaposvár, Sopron, Győr, Vác) we recorded altogether 147 sociolinguistic interviews (67 metropolitan, 80 rural) which had an average length of 3-4 hours. There were 67 men and 75 women among the informants; the youngest was 21 and the oldest was 82 years old. A statistician helped us with compiling the questionnaire and to choose the right informants in order to acquire data that can be analysed in a quantitative manner. In case of 27 informants both parents, in case of 3 informants only the mother and in case of 1 informant only the father was deaf. The rest was born into a hearing family. During the selection of the participants, the variant of the hearing status of parents was considered to be significant. In order to draw subtle description of the socialization background of the participants, their social networks and the role of languages were mapped. During the fieldwork we recorded sign language users that had different language socialization patterns.
Our questionnaire consisted of 345 questions; we first got some comprehensive data from the results regarding the social situation, language socialization patterns at school and at home, monolingual and bilingual language use and attitude towards the Hungarian Sign Language and different educational programs of Deaf people. Two-third of the questions are closed questions. The rest are narrative, open questions about topics which were mentioned earlier as closed questions at another stage of the interview or topics the informant were keen to speak about. These methods helped in reducing the effect of observer's paradox. The thematic structure of the questionnaire builds from formal to more informal topics so we can also measure the accommodation (cf. Giles et al 1991) between the participants.
During each interview there were 2 fieldworkers and 1 informant. The Fieldworker No.1 followed the questions on a laptop while conducting the interview, thus there was no need of holding the printed version. The Fieldworker No. 2 documented the answers, this way the Fieldworker No. 1 could actively sign and pay attention to the informant. The participants were situated at the three sides of an imaginary square. The informant was on the right side, opposite to the Fieldworker No. 1; the Fieldworker No. 2 documenting the interviews sat a little bit further behind, between the two of them, thus creating a right angle to both participants. We used 3 cameras for recording the interviews: one for the overall picture, one only for the informant and one only for Fieldworker No. 1.

Figure 1: Sociolinguistic interview

Before the live interview we also conducted some pilot tests, which we evaluated. Our aim was not only to avoid technical problems but also to make it sure that the interview is conducted by the fieldworkers in accordance with the guidelines.
In order to prepare the fieldworkers as well as possible and

to get as real data about language use as possible, we organized several trainings for them during which we trained the 16 deaf fieldworkers using different exercises. It was necessary to have 'local' signers conducting the interviews for each venue. Thus made it possible for the informant to clearly understand the fieldworker. On the other hand, dialectal variability was an important analytical aspect in the corpus, and this way we could avoid distortion resulting from speech adaptation (here: signing adaptation). During the fieldwork we asked the fieldworkers to constantly provide feedback for us; we also checked the recordings and we implemented the experiences into the future interviews. Not only the fieldwork but also the whole project was a process of continuous learning for all of us.
For preparing the sociolinguistic interviews we created an online questionnaire; on one hand its purpose was to gain information with the help of the Internet about the most important language use habits of deaf and hard hearing people, on the other hand the data collected contributed significantly to the project as the feedbacks were exceptionally useful in the preparation of the sociolinguistic questionnaire(s). Earlier, there was no research to sum up demographic data (age, gender, professional, education, family relations etc), information about levels of hearing loss, language use habits, attitudes towards Hungarian Languages and Hungarian Sign Language etc. During the trainings – prior to interviews – for the fieldworkers we debated the questionnaire from conceptual, linguistic, wording point of view. The experiences resulting from creating the questionnaire and from the results of the questionnaire provided important research information for the future sociolinguistic field work as well. The questionnaire concluded maximum 66 questions, this number could be smaller depending on the given answers (eg.: question about the child's hearing was asked in case of those who answered to have a child). Each part of the questionnaire (instructions, questions, options etc.) was available in HSL and also in Hungarian. Altogether we received 238 answers, the informants were between the age of 15 and 74; 94 men and 144 women. Based on location there was a big diversity: to the question about the place of residency we received 80 different answers. We succeeded to address the audience based on gender, age, status of hearing etc. It is important to highlight that the online questionnaire was anonymous, it is impossible to identify the informant, therefore we had the chance to reach out to those who would have not answered some questions in real life.

3.2 Grammatical tests

Although previously there were some attempts to describe some parts of its grammar, the first comprehensive, scientific linguistic description of Hungarian Sign Language was created in the Framework of the SIGNificant Chance Project in 2015. It was essential – just like during the whole project– that the grammar should be a result of corpus-based studies using recordings that reflect real language use, it should be based on sign language and avoid applying notions commonly used in spoken languages. Grammar was created by a team consisting of D/deaf colleagues, theoretical linguists and sign linguists. The result of their work was the grammatical test which has eliciting tasks for processing the basic phenomena of sign language. Among others, the following phenomena were tested: WH questions, question words, word order, contrastive topic, quantifiers, negation, etc.
The grammatical test contains 21 exercises. The location of the grammatical fieldwork was Budapest, the informants were always native sign language users; during each session two informants and one deaf fieldworker were present. During the first phase of the grammatical testing we worked with informants from Budapest, 16 recordings were created on 5 cameras. (Besides the 3 cameras used by the sociolinguistic interviews, a bird-eye camera was also used by participant to record the signing from an overhead perspective). During the following phases of the testing we conducted the grammatical tests with rural informants as well. The following criteria played a significant role while choosing the informants: gender, age, education, school type, where they went to school. We conducted the grammatical tests only with such people who have participated in a sociolinguistic interview earlier; this way we received a more complex picture about the language use of each informant adding detailed information about the informants' language socialization and background. At the moment we finished the analysis of the tests recorded with informants from Budapest. Here we recorded 32 informants (15 women and 17 men) in 16 sessions; they were all D/deaf except for one CODA participant. There were four age groups (18-30, 31-45, 46-60 and 61+); it was also necessary to have a proportional distribution not only based on gender and age, but also based on education; and we wanted to have max. 20% professional sign language users (e.g.: sign language teacher). Based on the hearing status of the parents in case of 10 informants both parents were D/deaf and in case of 22 informants both parents were hearing. The whole size of the grammatical corpus is 30 and a half hours.
The fieldworkers during the grammatical testings were deaf people who knew the test well; since they worked on putting them together and since they have some experience in empirical and theoretical linguistics, they could conduct the elicitation exercises in the preferred way, without affecting the natural signing of the informants.

4. Analysis of the corpus

4.1 Preparation and organization of the workflow

We used the ELAN software for analyzing the corpus. On one hand, the results of other international sign language corpus projects proved that this software could help the research aims; on the other hand, looking at the IT competences about operational systems and programs of the researchers and the annotators working with the corpus and the infrastructural conditions, ELAN seemed to be the most appropriate choice.
While elaborating on the annotation methodology of the project, we reviewed the international projects and used

their experiences. We mainly focused on the related works of Johnston (2013) regarding Auslan, but we also gained information from the Dutch (Crasborn et al. 2015) and British (Cormier et al. 2015) practice, furthermore, Ritva Takkinen and her colleagues also supported us. In Hungary we should mention HuComTech's gesture research project where ELAN is used for multimodal analysis (e.g. Abuczki 2013). In Hungary there is no research like this in the field of sign language.
The main aim of the annotation during the SIGNificant Chance Project was to support the sociolinguistic and grammatical work, especially by providing a sample materials and by involving members of the deaf community, because the modern, scientific analysis of sign language is a new experience for most Deaf, therefore it strengthened the positive connection between our colleagues and the community.
Annotators were educated in formal and informal ways. Before the formal education we appointed colleagues with high level of sign language competence, who learnt to use ELAN efficiently in order to support and make the work of the annotators smooth. There were two formal workshops were future annotators learnt the most important aspects of grammatical annotation and the use of the software. In order to adjust to the language skills of the annotators both the trainings and the educational material were available in Hungarian Sign Language and in (written) Hungarian; furthermore we translated the basic ELAN functions into Hungarian.
During the organization we planned and allotted the subtasks based on the individual competencies of the annotators: we prepared a task description, an instruction and we provided constant online (sign language or e-mail) support for the annotators. A work log was written including their notes and impressions.
Most of the annotators did not work in full-time but remotely; however personal meetings were regular in which we documented the feedbacks, we corrected previous works and we coordinated the schedule of the additional work.

4.2 Translating the sociolinguistic interviews

Since several researchers worked on the project who did not sign at all or well enough; besides, our primary aim was not only a grammatical analysis but also the content wise and qualitative analysis of the sociolinguistic interviews, a translation of the interviews was needed.
Our aim was to prepare translations properly segmented in ELAN and to attach them in form of annotations to the videos. However, during the 2 years of the project (including the development of the infrastructure, analysis and preparation of the interviews, development of the dictionary framework etc.) we did not have the chance to fulfill this plan with the limited number of translators, therefore we asked them to insert the translations into a Word chart.
Since we insisted on having CODA or interpreters respected by the members of the deaf community to do the translations to keep the data authentic and accurate, only a small number of translators could work on the recordings during the project, and most of them were from rural areas and they already suffered from a work-overload.
In the future, we plan to integrate these translations into ELAN and to check them.

4.3 Creating the tier structure

Parallel to the fieldworks 3 ELAN templates (sociolinguistic, grammatical and lexical) were created. While establishing these we leaned on the Australian annotation guidelines (Johnston 2013).
We set up 140 tiers for each person (informant and fieldworkers). Besides translation, each linguistic level is represented among them, from phonetics to pragmatics.
The complex tier-structure is a result of constant cooperation of applied linguists, sign linguists and deaf colleagues.
We created controlled vocabularies for certain linguistic types. We defined elements needed to describe handedness, movement (its type, direction and micro movements) and non-manual elements (mouth, eyes, eyebrows, look etc.).
We defined the possible elements based on the results of previous sign language researches and other, non-linguistic, but relevant researches (e.g: emotion and gesture analysis), integrating the feedbacks of the domestic deaf community.
The annotation works started with the lexical (see chapter 5.1) and grammatical (see chapter 5.2) researches; we give a detailed description of the used templates in the relevant chapters. From the conducted sociolinguistic interviews we synchronized 87 and we started the annotation of 76. We started to analyze 15 out of the 16 synchronized grammatical tests.

4.4 Annotation of the corpus

4.4.1 Annotation of the sociolinguistic interviews

Until March 2016, 41 sociolinguistic interviews were fully translated, this means approx. 2500 pages altogether. It will be a very complex task to transfer (and segment) all of them into ELAN and so far we have only been able to check and transfer 5 interviews; we plan to finish the rest in the future.

4.4.2 Processing the grammatical tests

While describing the Hungarian Sign Language grammar, our theoretical linguistic colleagues relied on the sample material created during the so called *focused grammatical research* (annotation), thus while planning the annotation of these parts the main aim was to support the creation of the grammar.
For the annotation of the analyzed phenomena, first the task-based segmentation of full records was needed. This was done by deaf colleagues – who knew the test well enough – on the *interview section* and the *important interview section* tiers; in this case we also used controlled vocabularies for annotation. The next step was the annotation and the segmentation of the elicited phenomena. The segments of the analyzed phenomena reflect in most cases the borders between sign language utterances, however, we did not systematically checked them so far.

Despite the fact, that due to the data-driven approach we used the right hand - left hand differentiation during the annotation (this way, handedness can be defined only from the data), taking the aims of the project into consideration we used passive and active hand tiers during the *focused grammatical research*. In the future, we plan to transfer and check the created annotations to the tier of right and left hand.

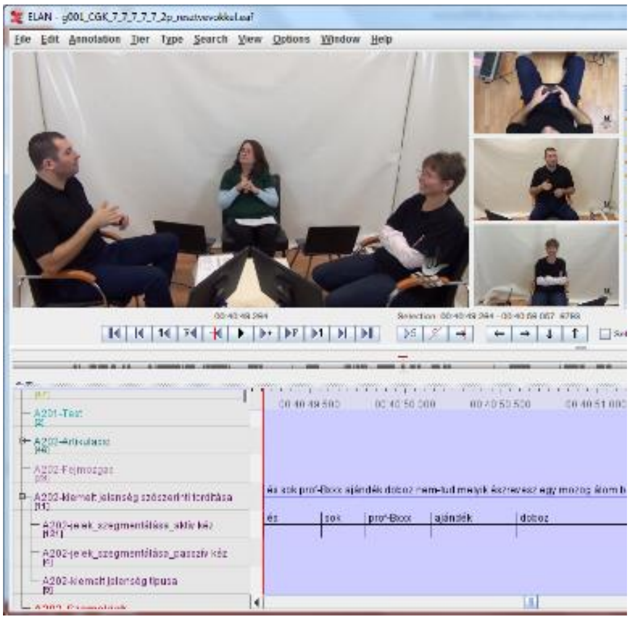


Figure 2: Annotation of the grammatical tests

At the present phase of the analysis we have only worked with actual meanings, however, later we would like to create an ID-gloss database, partly based on the meanings and partly based on the corpus-based online dictionary – to be described in chapter 5.2 – and the previously written Hungarian Sign Language dictionary. Basic grammar mainly focuses on sign language syntax, therefore we did not annotated the grammatical classification of the signs. The reason is that due to the lack of results from proper basic research of Hungarian Sign Language, we wanted to avoid the use of spoken language categories.

Depending on the type of the phenomena we annotated using non-manual components, where we also worked with controlled vocabularies. The annotation and the segmentation was made based on the following tiers: *body movement, head movement, eye-gaze, eye, eyebrow, cheek, chin, articulation, mouth* and *other-non manual*.

As the result of the annotation of the grammatical research, our colleagues created 15363 annotations. They are all approved and checked. Furthermore, during the grammatical annotation of the interviews we segmented each signs occurring in the interval of the *important interview sections*, therefore we created 34440 segments and their filling is going to be an important task in the future. This will serve as the secondary sign material of the ID-gloss database.

4.4.3 Annotation of the dictionary

From the 5 regions we analyzed 6 sociolinguistic interviews from each (altogether 30). Our aim was to have a ratio of 50-50% for women and men in each region, as well as in case of old and young. This criteria was not fulfilled in one region because there we had only a smaller number of interviews thus the proportional selection was not possible.

We annotated 209 pre-defined expressions (their occurrences in the interviews) that are essential in everyday life. Knowing the interview questions it was almost certain that they will be used (e.g. mother, father, and language). We created a separate template for the annotation of the dictionary. The sociolinguistic template would have also been appropriate to gain the needed information for the dictionary, but due to the shortness of the project, and in order to make the annotators' job faster and easier, we used a revised and simplified template. After the segmentation we annotated the following levels: 1. Hungarian translation equivalent. 2. Type of the sign (one handed, two-handed, mirror-symmetrical etc.) 3. Dominant hand, handshape 4. Non-dominant hand, handshape 5. Region, location of signing 6. Type of movement. The above mentioned linguistic information was completed with the code of the informant (enabling us to track other metadata later) and the city.

We did not use controlled vocabularies but we created a virtual keyboard similar to the keyboards on mobile phones and we depicted some elements (handshape, type of movement etc.) by pictograms. By pushing a particular button for a long time, options, such as possible elements of the signing location appeared. We used this method so that the deaf annotators could analyze the material faster, and the pictograms used helped them to rely on their visual competences and not on written language input.

5. Results of the project

5.1 Research on sociology of sign languages and sociolinguistic studies

From the results of the research we first got some comprehensive data about the social status of the Deaf, about language socialization patterns at home and at school, about monolingual and bilingual language use, about their attitude towards the Hungarian language and different educational programs etc. Besides, the recordings also made the corpus-based analysis (qualitative and quantitative) on different levels of sign language use possible. Moreover further researches can be conducted exploiting the database of sociolinguistic metadata.

5.2 Corpus-based dictionary

During the project we created the beta version of the corpus-base dictionary reflecting the dialectal diversity of HSL. We can search with the help of all the annotated characteristics (handshape, location, type of sign, type of movement, direction of movement, sign language – regional – variety); therefore search is not only possible from a spoken language perspective (keyword, topic, grammatical category, first letter) but also from the sign language perspective. Search in sign language was made easier with the pictogram-based search criteria. There is a definition for each sign and we can search for the English equivalent as well (English translation of all the signs has not been finished yet.).

5.3 Sign language grammar

The first comprehensive, scientific linguistic description of Hungarian Sign Language was created in the framework of the SIGNificant Chance Project in 2015; its script is accessible in the Research Institute for Linguistics of the Hungarian Academy of Sciences. It is outstanding in the sense that it is based on results of analysis of corpora, that reflect real language use, is sign language based it avoids applying notions commonly used in spoken languages to HSL.

5.4 Educational use

Bilingual education is a long-term objective, for which a lot of research is still needed. Since the main aim of SIGNificant Chance Project is to conduct a research that is essential for the establishment of bilingual education, it was very important for us to be able to use it in the field of education. The Hungarian Sign Language corpus by itself can be used as an educational material: it provides an access to authentic texts signed by native users of sign language. Furthermore, ELAN makes possible to subtitle videos faster, example sentences and helping materials can be exported which can be useful in deaf education, education of interpreters and in sign language courses. The corpus can be used as a source at courses focusing on the analysis of grammar phenomena in sign language (see Mesch-Wallin 2008).

6. Further tasks

By having a tier structure, we already have a framework for analyzing most levels of sign language, however, these should be revised from time to time. The next step – based on the pragmatic and discourse analysis experience of Research Center for Multilingualism – will be to work out the tier structure of sign language discourse and pragmatic researches. (The recent structure already includes a rudimentary version).

The creation of an ID-gloss database for corpus analysis (Johnston 2010) is also among our future aims, similarly to the integration of metadata into the corpus. We should regularly use a version-tracking software for documenting the annotations.

It is necessary to deepen the annotation of the corpus, to conduct more corpus-based researches in all levels of the sign language, because researches prior to the SIGNificant Change Project used a non-corpus based approach. Another aim of analyzing the sociolinguistic interviews is to make sign language accessible for those learning the sign language or learning in sign language – regardless of hearing status. In order to have an accurate description of sociolinguistic, dialectal and other levels of Hungarian Sign Language and about the lives of Deaf people, their experiences and language use it is necessary to further annotate the corpus and to publish the materials based on the results for the Deaf community, Sign Language Institutions, for hearing parents of deaf children and for those interested.

7. Acknowledgements

The study could not be realized without the support of SIGNificant Change Project (Támop 5.4.6/B-13/1-2013-0001) and the Methodology Project of the Hungarian Academy of Sciences (SZ-050/2014). We would like to thank the creators of the SIGNificant Project, our deaf and hearing colleagues.

8. Bibliographical References

Abuczki, Á. (2013). A *mondjuk* nem konceptuális használatának vizsgálata multimodális kontextusban. In: Váradi, T. (ed.): Alknyelvdok7. MTA Nyelvtudományi Intézet. Budapest. 3-17.

Bartha, Cs. (2004). Siket közösség, kétnyelvűség és a siket gyermekek kétnyelvű oktatásá-nak lehetőségei. In: Ladányi M., Dér Cs., Hattyár H. (eds.): „…még onnét is eljutni túlra…". Nyelvészeti és irodalmi tanulmányok Horváth Katalin tiszteletére. Tinta Könyvkiadó, Budapest 313–332

Bartha, Cs., Varjasi, Sz., Holecz, M. (2016). A magyar jelnyelvi korpusz létrehozásának és annotálásának kihívása. In: Tanács, A. Varga, V. Vincze V. (eds.) XII. Magyar Számítógépes Nyelvészeti Konferencia. Szegedi Tudományegyetem, TTIK, Informatikai Intézet, Szeged. 207–219.

Cormier, K., Fenlon, J., Gulamani, S., Smith, S. (2015). BSL Corpus Annotation Conventions (2015) http://www.bslcorpusproject.org/wp-content/uploads/BSLCorpus_AnnotationConventions_v2_-Feb2015.pdf

Crasborn, O., Bank, R., Zwitserlood, I., van der Kooij, E., de Meijer, A., Sáfár, A.: Annotation Conventions for the Corpus NGT. (2015) http://www.bslcorpusproject.org/wp-content/uploads/CorpusNGT_AnnotationConventions_v3_Feb2015.pdf

Giles, Howard – Coupland, Justine – Coupland, Nikolas (eds.) 1991. Contexts of Accomodation. Cambridge University Press. Cambridge–New York–Melbourne.

Johnston, T. (2010). From archive to corpus: transcription and annotation in the creation of signed language corpora. International Journal of Corpus Linguistics, 15(1), 104-129.

Johnston, T. (2013). Auslan Corpus Annotation Guidelines. Online: http://media.auslan.org.au/attachments/AuslanCorpusAnnotationGuidelines_Johnston.pdf

Mesch, J., Wallin, L. (2008). Use of sign language materials in teaching. Online: http://www.ling.su.se/polopoly_fs/1.68035.1325063806!/menu/standard/file/MeschWallin_LREC2008.pdf

WFD (2014). Statement on Standardized Sign Language. http://wfdeaf.org/wp-content/uploads/2014/05/WFD-Statement-on-Standardization-of-Sign-Languages-FINAL-20-MAY-2014.pdf

Collecting and Analysing a Motion-Capture Corpus of French Sign Language

Mohamed-El-Fatah Benchiheub[1,2], Bastien Berret[2], Annelies Braffort[1]
[1]LIMSI-CNRS, University of Paris-Saclay, [2]CIAMS, Univ. Paris-Sud, University of Paris-Saclay
[1]Campus d'Orsay, bât 508 F-91405 Orsay cx, France , [2]Campus d'Orsay, bât 335 F-91405 Orsay cx, France
mohamed-el-fatah.benchiheub@u-psud.fr, bastien.berret@u-psud.fr, annelies.braffort@limsi.fr

Abstract

This paper presents a 3D corpus of motion capture data on French Sign Language (LSF), which is the first one available for the scientific community for pluridisciplinary studies. The paper also exhibits the usefulness of performing kinematic analysis on the corpus. The goal of the analysis is to acquire informative and quantitative knowledge for the purpose of better understanding and modelling LSF movements. Several LSF native signers are involved in the project. They were asked to describe 25 pictures in a spontaneous way while the 3D position of various body parts was recorded. Data processing includes identifying the markers, interpolating the information of missing frames, and importing the data to an annotation software to segment and classify the signs. Finally, we present the results of an analysis performed to characterize information-bearing parameters and use them in a data mining and modelling perspective.

Keywords: French Sign Language, Motion Capture, Mocap, Animation, Annotation, Movement Analysis.

1. Introduction

Sign languages (SLs) are languages used to communicate with and among the Deaf communities. They are natural languages based on visuo-gestural modalities. Recent advances in computer graphics and animation have allowed the possibility to create and display 3D content in SL, by using a virtual signer (or signing avatar), i.e. a 3D character expressing itself in SL. This method allows the broadcasting of messages to Deaf people in an anonymous and modular way. However, generating 3D models based on actual knowledge of SL kinematics is still a challenge for computer scientists.

French Sign Language (LSF), as many other SLs, is still little described, particularly for what concerns the movement of articulators, and the existing models or representations in computer science are very simplified. Most of the studies in SL processing are interested in modelling linguistic properties, but few are interested in understanding the kinematics or dynamics of the movement itself and how it improves the comprehensibility of the generated signing. The rare ones have been applied on video corpora that do not allow estimating accurately and reliably velocities and accelerations (Segouat and Braffort, 2009; Lefebvre-Albaret, 2010).

Getting a better account of SL motion data thus requires novel resources. Recording 3D kinematics will allow designing more accurate models and improving knowledge in all scientific disciplines related to SL. However, the availability and the accessibility of the necessary technologies, which is scarce and expensive, make 3D corpora still rare especially for LSF.

Existing studies based upon such 3D corpora showed that they are of great value for all applications: generation, analyse of the movements (kinematic and dynamic) as well as linguistic analysis. For example, an American Sign Language (ASL) corpus has been used to compare animations generated by motion capture (mocap) and by generation algorithms. It was found that the animation based on mocap data generates movements that are more natural (Lu and Huenerfauth, 2010). Another study using LSF mocap data has been dedicated to automatic segmentation of the hand movement based on principal component analysis (Héloir et al., 2006). This method proved to be effective to solve high-level segmentation. 3D corpora are also used for linguistic analysis. For example, a study focused on identifying the type of verb (Telic and atelic), which seem to be distinguishable on the basis of speed and acceleration parameters on ASL corpus (Malaia et al., 2008).

There exists 3D corpora for LSF (Duarte and Gibet, 2010), but either they are not available or they do not meet the requirements for multidisciplinary research as we envision it, which is animation replay with a virtual signer, 3D data analysis of both body and facial movements, and linguistic annotation.

For these reasons, we started to create APlus, a 3D corpus of LSF available to the scientific community for multidisciplinary studies[1]. Our paper presents the steps of the data recording, data processing (labelling, gap-filling), and annotation. We also demonstrate how we perform and may exploit kinematic analysis on 3D data.

2. Content of the corpus

This paper describes the first part of the corpus, which represent about one hour of data. Six LSF native signers were involved in this part. They present various socio-linguistic profiles and signing styles, in order to have some insights on inter-signer variability.

Signers were asked to describe pictures in a spontaneous way. Each signer had a look at each picture during a few minutes before beginning the recording session. The elicitation material consisted of a set of 25 pictures showing many objects with peculiar geometrical properties (e.g. horizontal or vertical arrangements etc.) as in Figure 1. For the subject, the task thus consisted of describing successively the images.

The second part of the corpus, including various tasks is not described in this paper. More details can be found in

[1]More details on corpus characteristics here: https://tals.limsi.fr/corpus

Figure 1: Example of the described pictures

(Braffort et al., 2015).

3. Data recording

Motion capture is the process of recording the movement of objects or people. The recorded mocap data is transformed into a digital format for further processing and analysis or mapped on a digital model in 3D software. The recording provides a numerical coordinate matrix that can be used as a source of data for analysing the movements of the body parts from a kinematic perspective.

All the recordings have taken place in our studio in the Complexité, Innovation and Activités Motrices and Sportives laboratory (CIAMS) at the University of Paris-Sud, France. The CIAMS laboratory focuses on the study of motor control from biomechanical, neurophysiological and psychological perspectives. The studio hosts a 10 camera optical motion capture system (OptiTrack S250e). The frame rate of the cameras is 250 Hz, which is a sufficient resolution for our purpose. It allows managing a correct amount of markers with various sizes, attached to the body but also to the face, where miniaturized markers are needed. Given the resolution of the cameras, we have designed a setup with 40 markers of various sizes allowing to track the motion of the limbs but also movements occurring on the face (eyebrow, eye lead, cheeks and mouth movements). However, our system does not allow for accurate tracking of all the fingers. Only coarse information is given on finger movements. In addition to the mocap cameras, we have also used a digital video camera that provides a classical video to be used in the annotation software.

The very first step of this work was to design the best setup for the camera and marker locations. For that, we have conducted evaluations such that we can record sufficient details of the human performance for our various needs: animation replay with virtual signer, 3D data analysis, and linguistic annotation. One of the most important questions in mocap data recording is marker locations: where to attach the markers on the body, and why? This issue is important because the location of markers affects their visibility in the system: covered markers are not recorded. Marker locations are also important from the point of view of potential post-processing steps such as transforming the three-dimensional marker data into joint or segment representations. Furthermore, markers that are placed inappropriately might make it difficult for the signer to properly articulate signs. Finally, marker location must allow us to track as much as possible all the useful movements from a linguistic point of view.

Figure 2 shows the setup of the forty markers that we have used. There are 4 markers on top of the head: 2 in the front and 2 in the back. The torso contains 7 markers: 4 on the upper part (sternum, clavicle, two on the back (C7: Spinous process of the 7th cervical vertebrae, and T10: Spinous process of the 10th thoracic vertebrae), the other 3 markers on the pelvis. Each arm has 5 markers placed on the main joint positions (shoulder, inner and outer elbow, wrist ulnar and radial) and one on the triceps. There are also 2 markers on each palm. A set of 13 markers is used for the face: eyebrows, eyelids, cheeks, chin and mouth (below, above, left and right).

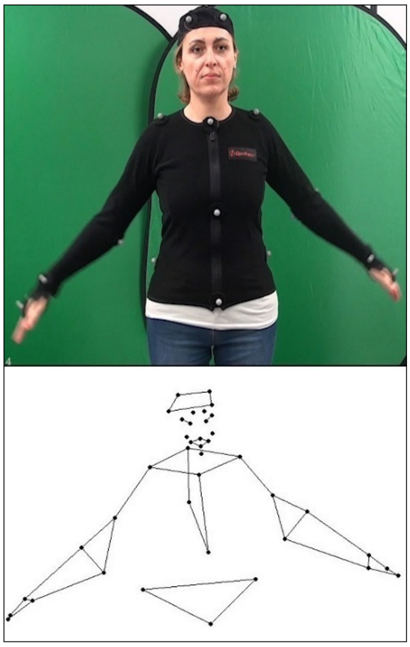


Figure 2: Up: markers attached on a subject, Down: markers connected by segments

The positions in our configuration were chosen so that the markers are maximally visible and identifiable by the system, and so that they capture the main global movements of the hands, arms, upper torso, and head. The location of the 6 markers on each arm was chosen in a way to be able to reconstruct the orientation (joint angles) of the 2 segments of the arm (upper arm and forearm). The rule is that there must be at least 3 markers on a rigid body to define its 3D orientation. We have put markers on the pelvis to differentiate the movements of lower part of the torso and those of the upper part. We have done tests with markers on the fingers but with the 10 camera system, finger markers overlap between them. For that reason, we have used 2 markers on each hand palm that allow us to have at least movement and rotation of the hands. The markers on the face allow us to have the eyebrow movements, winks, movement of the cheeks and of the mouth.

In comparison with other recent mocap studies, the total number of markers in our setup is fairly comprehensive: (Jantunen et al., 2012) used 20 markers (7 on each arm and hand, 4 on the head , and 2 on the upper torso). (Tyrone et al., 2010) used 30 markers (7 on each arm, 7 on the head, and 9 on the torso) and (Duarte and Gibet, 2010) whose additional goal is to use the data to create animated signing, avatars employed 98 markers (43 facial markers, 43 body markers, and 6 on each hand).

Figure 3 shows our optimal camera setup. We have used 4 cameras facing the signer and at the same height of the signer's head, 2 cameras on each side, and 2 cameras behind the signer. The cameras ahead allow a very good capture of the face markers, the cameras on the sides and behind allowing to capture the markers of the arms and head, the cameras behind are sufficient to capture the 2 markers placed on the back (C7 and T10). The digital video camera is placed in front of the signer. This setup allowed us a very good capture with minimum losses and overlapping of markers during recording.

Figure 3: Setup of the ten-camera optical motion capture as well as the HD video camera in our studio

In comparison with other recent mocap studies, (Jantunen et al., 2012) used eight-camera optical motion capture system (Qualisys ProReflex MCU120), while (Lu and Huenerfauth, 2012) used the Animazoo IGS-190 system to capture the movement of the arms and torso, with Intersense IS-900 to capture the movement of the head together with the two Immersion Cyber Gloves and eye tracker to capture the hands and eye movements respectively. (Duarte and Gibet, 2010) used twelve-camera optical motion capture system (Vicon MX).

We have included in our tests the use of a Tobii eye tracker. This is a device that incorporates illumination, sensors and processing to track eye movements and gaze point. This device allows to record gaze direction. This kind of device is not satisfactory because it hides the eyebrow and eye-lids movements. A better device remains to be found in order to include eye gaze in our data.

4. Data processing

Once raw data are recorded, there are several essential steps that must be done before the data can be exported and exploited. Due to the possible occlusion between the various parts of the body, and because the markers are not identified and may appear identical (marker swapping), a post-processing is needed to clean up the data.

The use of a high number of markers (40 for this corpus) has a drawback, which is the amount of gaps in the data as well as the overlapping between the markers which are close, or which will be close during the signing. This disadvantage may be overcome by the use of a larger number of mocap cameras (here we used 10 cameras). At least 2 cameras must see a marker at each frame for its instantaneous three-dimensional location to be recovered. If we had a system with more cameras (18 or 20) we would not have the gaps in data or overlap between markers, as there would be seen at any time by at least 2 cameras.

The Optitrack Motive software[2] gives several setups of markers. Using these predefined setups, we could obtain directly the markers labelled at the end of the recording. Unfortunately, these setups do not take into consideration the face, so we did not use them, and we add a step of identification (labelling) of the markers. Each time a marker is lost it must be re-labelled.

When all markers are labelled, we move to the second stage, which consists of removing noise. Indeed, at the end of the recording, there are fake/phantom markers, which are due to noise or reflection during the recording. This step can be done automatically, by removing all remaining non-labelled markers.

The third step is the gap-filling (filling the missing frames). This step can be done in Matlab software after export of the data with specific toolboxes such as (Burger and Toiviainen, 2013).But these toolboxes give quite arbitrary results when the gap is too long. To solve this problem, an option of the Motive software has been used before the export, with several methods (linear interpolation, cubic interpolation or interpolation based on other markers). The interpolation relative to other markers (markers that are in the same segment and which are fixed to each other) is the best method for the gaps that are relatively long, the cubic interpolation was used for the gaps that are in circular movements, and linear interpolation was used for the gaps that are in linear movements.

The next step is to check if there is an overlap between markers (errors on the marker identification). There is an overlap when two markers are too close between them during recording sequences, and the system confuses the two markers and reverse their identity. There are two cases in the overlap:

- The first case, the identity of markers remains reversed after the overlap.

[2]http://www.optitrack.com/products/motive/

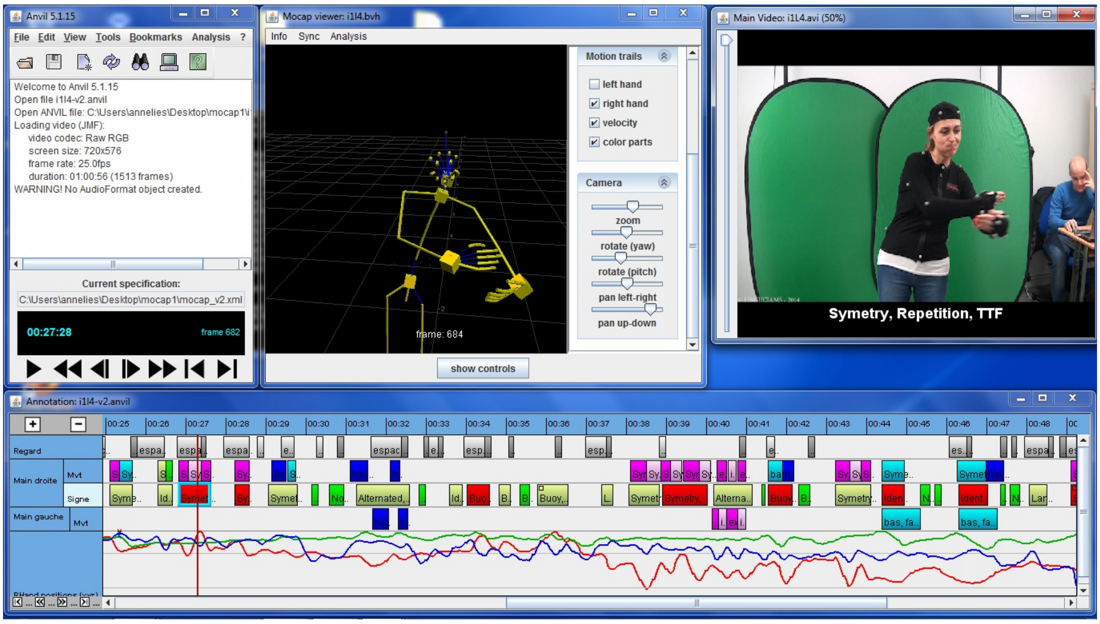


Figure 4: ANVIL screenshot shows the annotation using video, 3D skeleton and the mocap data

- The second case, the system reverses the identity of the markers only during the overlap, i.e. when the markers move away from each other their identities returns correct.

The verification is done marker by marker throughout the recording sequence. When there is an overlap between two markers, we delete the data of the two markers during the overlap. Then, as we said above, we have two cases. In the first case where the identity of markers remains reversed, we remove the labelling which is after the overlapping of these two markers, and we re-labelled them with the good identifications. In the second case the markers have the correct identification.

Now that we have markers with the good identifications, we fill the gaps that we made during the correction of the overlapping by using one of the three types of interpolation (defined above) depending on the case of movement and the sizes of the gaps.

At the end of these steps, and before exporting the data, we did an audit of the data by checking that all markers were labelled and that there were no gaps throughout the recording. This verification is done by running the animation in Motive software and looking at the colour of the markers. If they are all white during the animation, it means they are labelled throughout the recording. If a marker's colour changes from white to orange, this means that it is not labelled during these frames. To verify that there are no gaps in all markers, we verify that no marker is lost during all frames, but there is another easier way by selecting all markers and verify if there is no holes in displacement curves (X, Y and Z). When all these steps were achieved, the data were exported in c3d format with a frequency of 250 Hz.

The last step consists of making the data usable for the annotation software. ANVIL annotation software[3] was chosen because it can display the 3D data in addition to the video. For that, the c3d format was transformed into a bvh format by adding a skeleton hierarchy using the 3ds Max software.

[3]http://www.anvil-software.org/

5. Annotation

To make the data usable by linguists and also to analyse the movements, the bvh files and videos were imported to ANVIL annotation program. At this moment, the annotation is composed of three tracks, the first for gaze direction, the second for the type of movement (e.g. main direction), and the third for the linguistics annotation (see Figure 4). So far, the annotated movements are the linear ones in the three main axis up-down, medial-lateral, and anterior-posterior of both hands.

6. Data analysis

The 3D corpus enables an accurate quantitative analysis, allowing us to compute multiple parameters that characterize the movement: position, speed (mean velocity, peak of velocity), acceleration, angles between articulators, etc. As our ultimate aim is to develop models to generate LSF movements, we have first to identify the information-bearing parameters to reproduce in priority those critical parameters and get meaningful LSF. Indeed, it is currently difficult to expect that a model will reproduce all kinematic features of LSF given the complexity and the large number of degrees of freedom of the human body. Moreover, from a motor control viewpoint, the laws of motion used by signers when producing LSF movements are still poorly known, especially in comparison to non-LSF movements produced by other individuals. Intriguing and unresolved questions pertain to the existence of invariant and peculiar features in the kinematics of SL movement, and how they compare to non-SL movements.

We present here some preliminary results of our study related to velocity, and explain how this kind of corpus can be exploited for the study of motor control in SL and how simple process on 3D data can allow for automatic computation of metadata related to the signer.

6.1. Mean velocity: linked to the degree of control?

A parameter that has been analysed here is the mean velocity of the movement of the dominant arm, for lexical signs and depicting signs that describe the size and the shape of entities (SASS) which are very frequent in this description task. It was found that this parameter varies extensively between different subjects: For instance, the mean velocity in lexical signs and SASS respectively of a subject was around 0.51 m/s and 0.60 m/s with standard deviation of 0.3 m/s and 0.22 m/s across the entire session, while the mean velocity in lexical signs and SASS respectively of another subject was around 0.91 m/s and 0.97 m/s with standard deviation of 0.31 m/s and 0.23 m/s.
This drastic change of movement pace gives some hints about the underlying control laws used in LSF. In particular, it allows to assert that the mean velocity of movement of the dominant arm between different subjects does not influence the understanding of LSF. In other words, the mean velocity of the signer does not carry information about linguistic meanings of the movements.
The other result is the difference in the mean velocity of movement of the dominant arm between lexical signs, SASS, and transitions as shown in the histogram of Figure 5. The mean of the mean velocity of the four subjects in the lexical signs, SASS, and the transitions are respectively 0.72 m/s, 0.81 m/s and 0.91 m/s with standard deviation of 0.2 m/s, 0.2 m/s and 0.2 m/s. Thus, the mean velocity in the lexical signs is lower than in the SASS, which is lower than in the transitions.

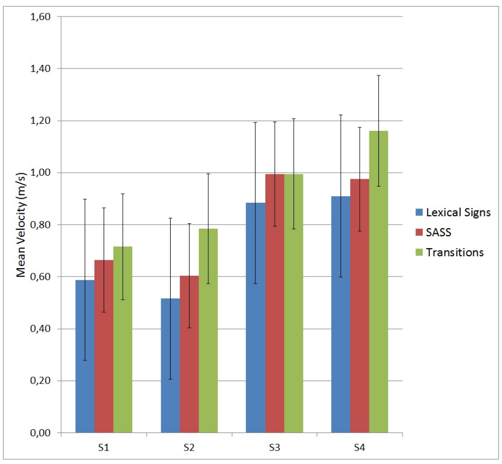


Figure 5: This histogram shows the mean velocity of the four subjects for lexical signs, SASS and transitions

Concerning the transition velocity, an explanation could be that transitions do not convey any message (information) and then need less control, being faster to perform.
Concerning the difference between the lexical signs and the SASS velocity, an hypothesis could be that most of the time, eye gaze is accompanying SASS depicting signs (Braffort, 2016), which is not the case for lexical signs. That could help performing these gestural units in an easier and then faster way.
Of course, these hypotheses should be confirmed by other studies.
This result is also confirmed by the peak of the maximal velocity. The average of the peak of the speed for lexical signs, SASS, and transition respectively is 1.12 m/s 1.25 m/s and 1.36 m/s with standard deviation of 0.26 m/s, 0.32 m/s and 0.26 m/s. This confirms that the velocity in lexical signs and SASS is lower than in transitions.
In conclusion, we can assume that there is more control over arms movements during the signs, being lexical or iconic (SASS) than during transitions.
Therefore, models should be able to produce signs at various paces while preserving the same spatiotemporal organisation. These speeds should also take into account the difference between the types of signs and the transitions, which may be done by means of two parameters that could be tunable in our models in order to change the overall pace of LSF movements.

6.2. Motor control in LSF

Another analysis in progress is to check whether classical laws established in the human motor control literature apply to LSF. That means that we ask the following question: Do classical invariants remain valid during LSF movements? If these laws apply in LSF, one may conceivably assume that classical motor control principles, such as minimum effort or maximum smoothness criteria, may have shaped LSF and must be incorporated into LSF production models. Alternatively, it is possible that LSF requirements led signers to deviate from such classical principles in order to produce very peculiar kinematics of the hands and deliver linguistic meaning. Ongoing investigations will attempt to answer such questions, which is made possible thanks to the creation of a corpus of 3D data of LSF.
One other current focus is related to the law of up-down asymmetries, which states that point-to-point upward movements decelerate for a longer time compared to downward movements, in particular due to the integration of gravity in the motor command driving the limb's motion (Papaxanthis et al., 1998; Gaveau and Papaxanthis, 2011).

6.3. Detection of the dominant hand

An application of using 3D data is the automatic detection of the dominant hand in LSF.
This can be achieved based on the computation of the distances covered by the two hands. By comparing these distances, we can automatically detect the strong hand, which is more active. This computation could be used to automatically feed the metadata related to the signers in annotation software.
We have also studied the variability of this difference across the subjects, by calculating the ratio r between the two distances.

$$r = \frac{D_{weakhand}}{D_{stronghand}} \quad (1)$$

Histogram 2 shows that the ratio (r) is quite stable across the subjects. The global average is r = 0.768 with standard deviation of 0.014.

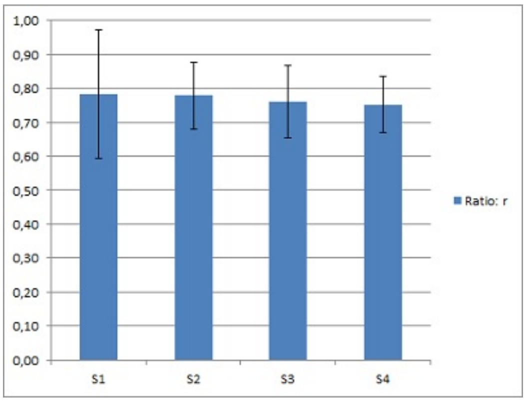


Figure 6: Histogram showing the ratio between the distance covered by the strong hand and the distance covered by the weak hand

7. Conclusion

This paper described the different stages of the constitution of APlus, the first available 3D corpus of LSF, which will be usable in several disciplines. The potential power of analyses based upon the 3D corpus was illustrated. Their main advantage is that they allow to quantify and identify the information-bearing parameters of LSF movements with the aim to use them in the modelling of movements in LSF.

At this moment, the initial part of the corpus, corresponding to the picture description task, has been recorded and fully annotated. The targeted analyses are being completed using the above-mentioned fundamental questions. The first part of the corpus is available on request from authors. The second part has been recorded and annotated.

8. Bibliographical References

Braffort, A., Benchiheub, M.-E.-F., and Berret, B. (2015). APLUS: A 3d corpus of french sign language. In Yeliz Yesilada et al., editors, *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility, ASSETS 2015, Lisbon, Portugal, October 26-28, 2015*, pages 381–382. ACM.

Braffort, A. (2016). Eye gaze in french sign language: a study on depicting signs. In *The 12th Conference on Theoretical Issues in Sign Language Research, 4 - 7 January 2016*, Melbourne, AUSTRALIA.

Burger, B. and Toiviainen, P. (2013). Mocap toolbox - a matlab toolbox for computational analysis of movement data. In *Proceedings of the Sound and Music Computing Conference 2013, SMC 2013, Logos Verlag Berlin*, pages 172–178, Stockholm, Sweden. Logos Verlag Berlin.

Duarte, K. and Gibet, S. (2010). Heterogeneous data sources for signed language analysis and synthesis: The signcom project. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta*, pages 461–468.

Gaveau, J. and Papaxanthis, C. (2011). The temporal structure of vertical arm movements. *PLoS ONE*, 6(7):e22045.

Héloir, A., Gibet, S., Multon, F., and Courty, N. (2006). Captured motion data processing for real time synthesis of sign language. In Sylvie Gibet, et al., editors, *Proceedings of 6th International Gesture Workshop*, volume 3881 of *Lecture Notes in Computer Science*, pages 168–171. Springer.

Jantunen, T., Burger, B., Weerdt, D. D., Seilola, I., and Wainio, T. (2012). Experiences collecting motion capture data on continuous signing. In Onno Crasborn, et al., editors, *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions Between Corpus and Lexicon, The 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 75–82, Istanbul, Turkey, May. Paris: ELRA.

Lefebvre-Albaret, F. (2010). *Traitement automatique de vidéos en LSF Modélisation et exploitation des contraintes phonologiques du mouvement*. Theses, Université Paul Sabatier - Toulouse III, October.

Lu, P. and Huenerfauth, M. (2010). Collecting a motion-capture corpus of american sign language for data-driven generation research. In *Proceedings of the NAACL HLT 2010 Workshop on Speech and Language Processing for Assistive Technologies*, pages 89–97, Los Angeles, California, June. Association for Computational Linguistics.

Lu, P. and Huenerfauth, M. (2012). Cuny american sign language motion-capture corpus: First release. In Onno Crasborn, et al., editors, *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, The 8th International Conference on Language Resources and Evaluation (LREC 2012)*, pages 109–116, Istanbul, Turkey, May. Paris: ELRA.

Malaia, E., Borneman, J., and Wilbur, R. B. (2008). Analysis of asl motion capture data towards identification of verb type. In Bos Johan et al., editors, *Semantics in Text Processing. STEP 2008 Conference Proceedings*, volume 1 of *Research in Computational Semantics*, pages 155–164. College Publications.

Papaxanthis, C., Pozzo, T., and Stapley, P. (1998). Effects of movement direction upon kinematic characteristics of vertical arm pointing movements in man. *Neuroscience Letters 253*, pages 103–106.

Segouat, J. and Braffort, A. (2009). Toward the study of sign language coarticulation: methodology proposal. *Advances in Computer-Human Interactions*, pages 369–374.

Tyrone, M. E., Nam, H., Saltzman, E., Mathur, G., and Goldstein, L. (2010). Prosody and movement in american sign language: A task-dynamics approach. Speech Prosody 2010, 100957, pages 1–4.

Visualizing *Lects* in a Sign Language Corpus: Mining Lexical Variation Data in Lects of Swedish Sign Language

Carl Börstell[1] & Robert Östling[2]
[1]Dept. of Linguistics, Stockholm University
S-106 91 Stockholm, Sweden
calle@ling.su.se
[2]Dept. of Modern Languages, University of Helsinki
FI-00014 Helsinki, Finland
robert.ostling@helsinki.fi

Abstract

In this paper, we discuss the possibilities for mining lexical variation data across (potential) *lects* in Swedish Sign Language (SSL). The data come from the SSL Corpus (SSLC), a continuously expanding corpus of SSL, its latest release containing 43 307 annotated sign tokens, distributed over 42 signers and 75 time-aligned video and annotation files. After extracting the raw data from the SSLC annotation files, we created a database for investigating lexical distribution/variation across three possible lects, by merging the raw data with an external metadata file, containing information about the age, gender, and regional background of each of the 42 signers in the corpus. We go on to present a first version of an easy-to-use graphical user interface (GUI) that can be used as a tool for investigating lexical variation across different lects, and demonstrate a few interesting finds. This tool makes it easier for researchers and non-researchers alike to have the corpus frequencies for individual signs visualized in an instant, and the tool can easily be updated with future expansions of the SSLC.

Keywords: Swedish Sign Language, sign language, corpus, lexical variation, data visualization, interface

1. Introduction

Lexical variation is a topic that has received a fair amount of attention in sign language linguistics (Lucas, 2006; Schembri and Johnston, 2012). However, it is only recently that sign language corpora have come about, meaning that the study of lexical variation now has access to a larger, more varied dataset than ever before. To date, sign language corpora are available for a number of sign languages (see Börstell et al. (2014b) for a non-exhaustive list) with more under way, but their size in terms of tokens is far from that of spoken languages. Although sign language corpora are not big by token count, they do require a substantial space for data storing, since sign language data is necessarily recorded in video format. Perhaps because of this, most sign language corpora are not easily accessible to non-researchers, seeing as they often require downloading of heavy bundles of video and annotation files, and mostly render corpus search results in a strictly numerical form (i.e. without any type of graphical visualization). Thus, with this study, we looked to mine and re-compile the data from a sign language corpus by adding signer metadata for sociolinguistic factors known to interact with lexical variation directly into a searchable database, but also create a simpler graphical user interface (GUI) that directly visualizes the output of any corpus search without depending on video files, in an attempt to make the corpus data more accessible in a lightweight format.

2. Background

2.1. Lexical Variation

Variation in sign language has been a topic researched since the early days of sign language linguistics (Lucas, 2006). The specific focus of the research has varied, with different studies looking at variation on levels ranging from sublexical to discourse units, and the explanations for which factors are responsible for the variation have included region, age, gender, and ethnicity (Bayley et al., 2015). A well-known work on the issue of lexical variation is the book *What's your sign for* PIZZA*?* (Lucas et al., 2003), which presents the findings of a large-scale project on lexical variation in American Sign Language (ASL) across the United States. More recently, with the advent of true sign language corpora, some studies have been conducted looking at variation in British Sign Language (BSL), such as Fenlon et al. (2013) investigating the contextual and sociolinguistic factors affecting the shape of the 1-hand configuration, and Stamp et al. (2014) investigating the regional variation of color signs. This second study made use of corpus data, but specifically a subset of corpus data consisting of lexical items elicited using word lists. For Swedish Sign Language (SSL), the only previous study concerning variation is Nilsson (2004), which looked at the form variations of the first-person pronoun PRO1 in discourse data, although not from a sociolinguistic perspective. However, the online dictionary of SSL (Björkstrand, 2008) does contain some information about sociolinguistic features of signs, such as regional distribution of particular signs, as well as signs seen as old-fashioned, but this dictionary is not linked to, or based on, corpus data (Mesch et al., 2012a).

2.2. The SSL Corpus

The SSL Corpus (SSLC) is a corpus of naturalistic, dyadic signing of Swedish Sign Language. The SSLC data were collected over three years (2009–2011), and comprises 300 video recordings distributed over 42 signers (Mesch et al., 2012b), with the signers selected in order to approximate a balanced and representative sample in terms of age groups, genders, and regional distribution (Mesch, 2012; Mesch et al., 2012a; Wallin and Mesch, 2015).[1] To date, 75 (i.e. 25%) of the video files have been edited, glossed, and translated (Mesch et al., 2015). The video files are annotated using the ELAN software, producing annotation files (.eaf) that are underlyingly XML files, allowing for multiple annotation tiers time-aligned to a media file (Wittenburg et al., 2006). Currently, the SSLC annotation files consist of two main tier types: sign gloss annotations; and Swedish translations. The only segmentation that has been done for the SSL data is on the lexical level, with sign glosses being entered into annotation cells corresponding to the duration of individual signs on the time-axis, though the possibility of introducing a syntactic/prosodic segmentation has been investigated (Börstell et al., 2014a). Apart from the sign glosses—i.e. the labels uniquely identifying each sign in the corpus (Mesch and Wallin, 2015; Wallin and Mesch, 2015)—the SSLC has also recently been tagged with parts of speech, using a semi-automatic tagging procedure (Östling et al., 2015).

3. Methodology

3.1. Aim

In the SSLC, the participants are grouped according to three different variables, as provided by the signer metadata documented during the collection of the primary (i.e. sign language) data. These three group variables are: (a) **Region**, the regional affiliation of the signers based on the *landsdelar* (lit. 'country parts') of Sweden—Norrland, Svealand, and Götaland; (b) **Age group**, the categorization of signers into six age groups; and (c) **Gender**, female or male.[2] Furthermore, the individual files in the SSLC are categorized into three different text types—conversation, narrative, and presentation, respectively. However, the signer metadata and the text type information are not available directly in the SSLC annotations to be used with ELAN as the user interface. The raw metadata files themselves contain information about individual signers and are thus not publicly available. In this project, we used the metadata files to match the anonymous signer-IDs to each group variable, such that the resulting database does not contain neither personal details about individual signers, but rather sign frequency data for groups of signers (or text types).

The aim of this work was two-fold: firstly, we wanted to link the group variables of the signer metadata directly to the lexical data in the SSLC, storing it as a type of database; secondly, we wanted to create methods for mining interesting data, either by using computational search methods for research purposes, or as an custom-built, easy-to-use interface for which researchers and non-academics alike could search this database and get instant visual representations of the lexical frequency distributions across all group variables.

[1] `http://www.ling.su.se/teckenspraakskorpus`

[2] Though additional metadata such as educational background and age of onset for sign language acquisition have been documented during the data collection, this information was not available to us for each signer as the other metadata, thus restricting our study to the selected variables.

In this paper, we also make a short evaluation of the data and our search interface, and provide a few examples of how the tool can be used for quick visualizations of lexical distributions.

3.2. Data

For this study, we used the data from the latest version of the SSLC. This version comprised 75 annotation files, consisting of 43 307 sign tokens. However, many tokens are tagged with any of the suffixes `@x` or `@z`, marking that the sign gloss is uncertain or the sign unidentifiable (Wallin and Mesch, 2015), hence such signs were excluded from our dataset. Thus, we arrived at a dataset of 39 733 sign tokens, distributed over 4 676 sign types. However, since the SSLC is still being annotated, the corpus is not (yet) balanced in terms of the distribution of annotated tokens within each group variable in the metadata. In order to account for the imbalance in token frequency across groups, we based all results on relative frequencies (see 3.2.1. and 3.3.). The distribution of sign tokens within each of the three group variables is given in Tables 1, 2, and 3, and the distribution of sign tokens across text types is given in Table 4.

Region	Signers	Tokens
Norrland	4	5 310
Svealand	24	24 605
Götaland	14	9 818

Table 1: Distribution of signers and tokens according to region.

Age group	Signers	Tokens
20–29	9	4 225
30–39	6	11 680
40–49	7	10 646
50–59	8	3 007
60–69	8	7 756
70–100	4	2 419

Table 2: Distribution of signers and tokens according to age.

Gender	Signers	Tokens
female	20	15 862
male	22	23 871

Table 3: Distribution of signers and tokens according to gender.

It should be noted that the crude division of regions into *landsdelar* does not correspond to Deaf schools, for which there have traditionally been seven: one in Norrland; four in Svealand; and two in Götaland (see Figure 1).[3]

[3] NB: Some cities had more than one Deaf school.

Text type	Files	Tokens
Conversation	56	34 071
Narrative	14	3 525
Presentation	5	2 137

Table 4: Distribution of files and tokens according to text type.

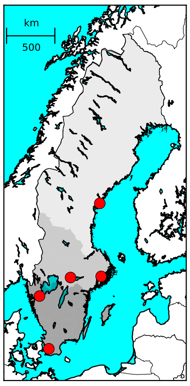


Figure 1: The *landsdelar* of Sweden—Norrland (light gray), Svealand (gray), Götaland (dark gray)—with the locations of the deaf schools (red dots).

3.2.1. Extracting and reading the relevant data

All sign data were extracted from the ELAN annotation files and then matched to the external metadata on signers, so that we end up with a count $c_{s,g}$ representing the number of times sign s was used by any signer from group g. Then, we can compute the relative frequency among all the groups in a category G (e.g. age) using the maximum-likelihood estimate:

$$r_{s,g} = \frac{c_{s,g}}{\sum_{g' \in G} c_{s,g'}}$$

3.3. Identifying Unevenly Distributed Signs

Rather than just obtaining the social and geographic distribution of particular signs, we are also interested in *finding* the signs that are used significantly more often by some groups than by others.

We compute three rankings, one each for the categories of region, age, and gender. Signs are ranked by the Bayes factor between the hypothesis of separate categorical distributions versus an identical categorical distribution, assuming a Dirichlet prior for the categorical parameters:

$$b_s = \frac{B(x_s + \alpha)B(t - x_s + \alpha)}{B(t + \alpha)}$$

where x_s is a vector representing the distribution of the sign s and t is the distribution vector of all signs, and $B(x)$ is the multinomial Beta function:

$$B(x) = \frac{\sum_i \Gamma(x_i)}{\Gamma\left(\sum_i x_i\right)}$$

We use a uniform prior for the distributions, setting $\alpha = 1$.

3.4. Constructing a Visual Interface

For the visual interface, we wrote a program that took the input sign objects read from the datafile and waited for a user input, in this case asking for a specific sign gloss to be plotted. When a sign gloss was entered into the interface, the program would plot it using the Matplotlib module (Hunter, 2007). A bar chart was subsequently created for each of the group variables—region, age group, and gender—as well as one for text type, presenting the sign's relative frequencies in tokens per 100 signs. This interface was implemented as a web script and made accessible online.[4]

4. Results and Evaluation

4.1. Evaluating the Data Visualization

The obvious problem with the SSLC data is its small scale. Even after balancing out the skewed token distribution within variables, the fact remains that ≈40 000 tokens is insufficient for estimating reliable statistics for anything but the most high-frequent items. The most frequent sign in the SSLC is PRO1 (Börstell et al., Submitted). The graphs in Figure 2 show the distribution of relative token frequencies for PRO1 across each group variable.

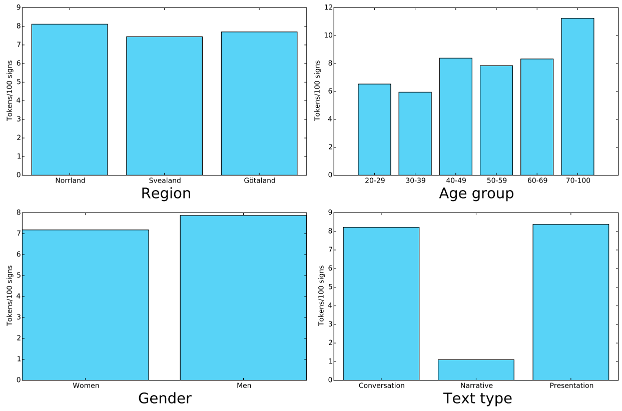


Figure 2: The distribution of the sign PRO1 (n = 3 018).

As is visible from these graphs, the relative frequencies are more or less even for each group variable. This is to be expected from a sign that is highly frequent. Unsurprisingly, it is for text type that the sign PRO1 shows a skewed distribution, with the sign being relatively uncommon in the narrative texts, which in the SSLC are mainly elicited narratives (as opposed to self-experienced narratives). However, we also wanted to see if specific items do exhibit a distribution that reflects *lectal* lexical variation.

For region, we take the example of the sign ÄLG(Jb) ('moose'), which is listed as a regional northern sign in the SSL dictionary (Björkstrand, 2008).[5] Figure 3 shows the distribution of the seven tokens found for this sign, supporting the claim that this sign is associated with Norrland, with

[4] `http://mumin.ling.su.se/cgi-bin/sslects.py`

[5] Suffixed tags in round brackets indicate a specific form for meanings for which there are sign variations. The letters within the brackets describe the handshape.

all tokens coming from this region. As for the identification of unevenly distributed signs, the sign ÄLG(Jb) does in fact appear in the top (15^{th} place) of signs with an uneven distribution across regions, showing that the method correctly identifies this sign as a sign with a skewed regional distribution (in this case, being associated with a specific region, viz. the north). Unfortunately, the non-northern sign for 'moose' (ÄLG(5)) is not yet attested in the SSLC.

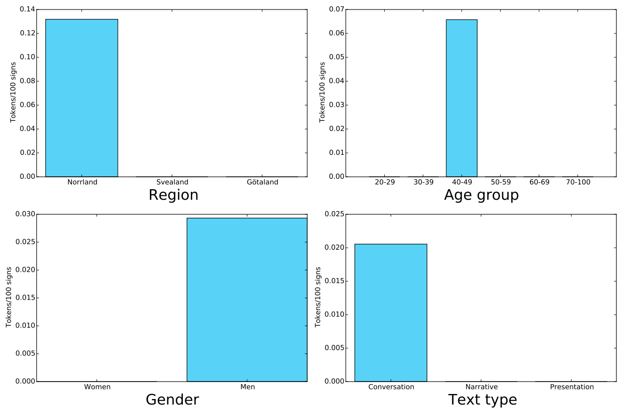


Figure 3: The distribution of the sign ÄLG(Jb) ('moose') ($n = 7$).

For age, there are not many signs marked as typical for younger or older signers in the SSL dictionary that also occur in the SSLC. However, there are signs generally perceived as more typical to a certain generation or age group. One such sign is TYP@b ('kind of', lit. 'type'), which is said to be more typical among younger signers, as it is a borrowing from spoken Swedish (where it is also associated with younger speakers).[6] Figure 4 appears to support this idea, with the 77 tokens of the sign being largely distributed over the younger age groups. Furthermore, the sign TYP@b appears in the very top (5^{th} place) of signs with an uneven distribution across age groups, showing that the method again correctly identifies this sign as a sign with a skewed distribution (in this case, being associated with younger signers).

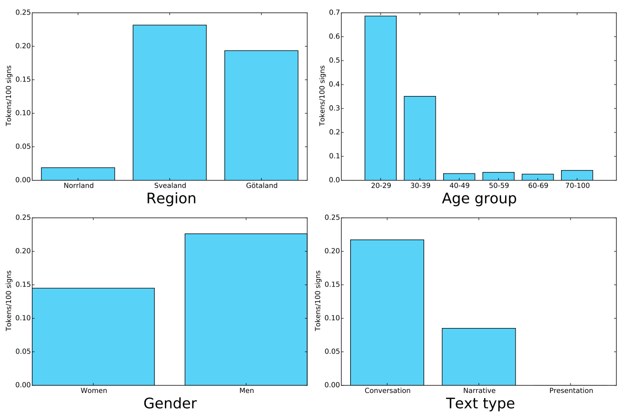


Figure 4: The distribution of the sign TYP@b ('kind of') ($n = 77$).

Finally, for gender, there is one pair of signs often claimed to be in a gendered complementary distribution, namely the signs SNYGG@b and SNYGG(H), both meaning 'attractive', but the former said to be used by women and the latter by men. Figures 5 and 6 seem to support this, although it should be noted that the graphs are based on very few absolute tokens (3 and 1, respectively)—also, the few tokens make these signs hard to identify statistically as showing an uneven distribution.

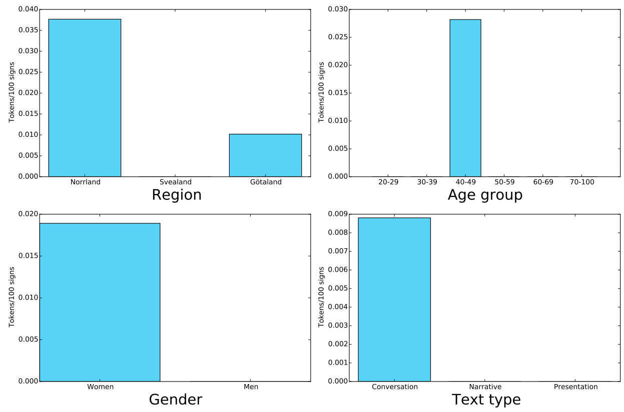


Figure 5: The distribution of the sign SNYGG@b ('attractive') ($n = 3$).

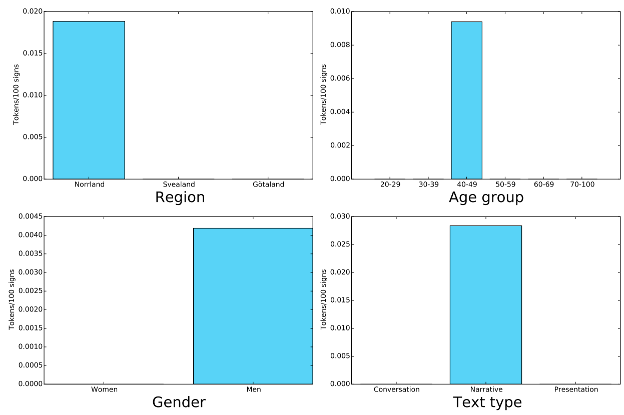


Figure 6: The distribution of the sign SNYGG(H) ('attractive') ($n = 1$).

4.2. Evaluating the Method Identifying Unevenly Distributed Signs

The output of the method identifying unevenly distributed signs (described in 3.3.) shows potential. Although the SSLC suffers from a quite limited amount of data in terms of token size—as do all sign language corpora—the method correctly identifies the signs that we selected from prior knowledge (albeit anecdotal, in some cases) about their lectal distribution. Thus, it shows potential as a method of automatically identifying signs with a skewed distribution based on lectal lexical variation. However, with the limited amount of data available in the current version of the SSLC, many signs identified as showing a skewed distribution are, as confirmed after a manual check, merely skewed due to conversation topics of individual signers rather than

[6]The tag @b indicates that the sign is fingerspelled.

as cases of lexical variation (i.e. a certain sign is skewed towards a specific group because of a single signer talking about a related topic and making it seem as though the group "overuses" the sign). In some cases, this points to interesting differences in conversation topics, as with the sign MAN(H) ('husband') being heavily skewed towards being used by female signers, whereas the sign FRU ('wife') is skewed towards male signers. Similarly, certain toponyms are, unsurprisingly, used more by signers from that region. Nonetheless, with an expansion of the corpus, we are optimistic of the possibilities that this method brings.

5. Conclusion

In this study, we have described the procedure of extracting data from raw corpus annotations, matching them to signer metadata, and constructing a database for investigating lexical distribution (and possible variation) based on the factors region, age, and gender, as well as the creation of a web-based data visualization tool that we have made publicly available, for researchers and non-researchers alike. We also utilize a method for automatically identifying uneven distributions, and find that it correctly identifies several signs that are expected to exhibit a skewed distribution based on lectal variation. Though the SSLC is still too small to do any large-scale investigations of lexical variation—simply based on the fact that the there are too few tokens as well as signers—we can still visualize some of the known or previously assumed cases of lexical variation in SSL, and more instantly than previously possible thanks to our database and GUI. With the expansion of the SSLC in terms of data, the database will get richer, and thus more adequate for research purposes on lexical variation. A larger corpus would also give the automatic identification of unevenly distributed signs a better dataset on which to conduct its calculations, for which we are confident it could serve as a useful tool for pinpointing interesting sociolinguistic variation. Also, making the web interface available online with direct access to and visualization of the SSLC data should make the corpus as a resource more available to the general public and more specifically the SSL community.

6. Acknowledgments

We wish to thank Johanna Mesch for providing us with the metadata files from the SSLC.

7. Bibliographical References

Bayley, R., Schembri, A. C., and Lucas, C. (2015). Variation and change in sign languages. In Adam C. Schembri et al., editors, *Sociolinguistics and Deaf Communities*, pages 61–94. Cambridge University Press, Cambridge.

Björkstrand, T. (2008). Swedish Sign Language Dictionary online. `teckenspraakslexikon.su.se`.

Börstell, C., Mesch, J., and Wallin, L. (2014a). Segmenting the Swedish Sign Language Corpus: On the possibilities of using visual cues as a basis for syntactic segmentation. In Onno Crasborn, et al., editors, *Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel [Language Resources and Evaluation Conference (LREC)]*, pages 7–10, Paris. European Language Resources Association (ELRA).

Börstell, C., Sandler, W., and Aronoff, M. (2014b). Sign Language Linguistics. In Mark Aronoff, editor, *Oxford Bibliographies Online: Linguistics*. Oxford University Press.

Börstell, C., Hörberg, T., and Östling, R. (Submitted). Distribution and duration of signs and parts of speech in Swedish Sign Language.

Fenlon, J., Schembri, A., Rentelis, R., and Cormier, K. (2013). Variation in handshape and orientation in British Sign Language: The case of the '1' hand configuration. *Language and Communication*, 33(1):69–91.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95.

Lucas, C., Bayley, R., and Valli, C. (2003). *What's your sign for PIZZA?* Gallaudet University Press, Washington, DC.

Lucas, C. (2006). Sign language: Variation. In Keith Brown, editor, *Encyclopedia of Language & Linguistics*, number 1993, pages 354–358. Elsevier, Oxford.

Mesch, J. and Wallin, L. (2015). Gloss annotations in the Swedish Sign Language Corpus. *International Journal of Corpus Linguistics*, 20(1):103–121.

Mesch, J., Wallin, L., and Björkstrand, T. (2012a). Sign Language Resources in Sweden: Dictionary and Corpus. In Onno Crasborn, et al., editors, *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon [Language Resources and Evaluation Conference (LREC)]*, pages 127–130, Paris. European Language Resources Association (ELRA).

Mesch, J., Wallin, L., Nilsson, A.-L., and Bergman, B. (2012b). Dataset. Swedish Sign Language Corpus project 2009–2011 (version 1).

Mesch, J., Rohdell, M., and Wallin, L. (2015). Annotated files for the Swedish Sign Language Corpus. Version 3.

Mesch, J. (2012). Swedish Sign Language Corpus. *Deaf Studies Digital Journal*, 3. `http://dsdj.gallaudet.edu/index.php?issue=4&section_id=2&entry_id=128`.

Nilsson, A.-L. (2004). Form and discourse function of the pointing toward the chest in Swedish Sign Language. *Sign Language & Linguistics*, 7(1):3–30.

Östling, R., Börstell, C., and Wallin, L. (2015). Enriching the Swedish Sign Language Corpus with part of speech tags using joint Bayesian word alignment and annotation transfer. In Beáta Megyesi, editor, *Proceedings of the 20th Nordic Conference on Computational Linguistics (NODALIDA 2015), NEALT Proceedings Series 23*, pages 263–268, Vilnius. ACL Anthology.

Schembri, A. and Johnston, T. (2012). Sociolinguistic aspects of variation and change. In Roland Pfau, et al., editors, *Sign language: An international handbook*, pages 788–816. De Gruyter Mouton, Berlin/Boston, MA.

Stamp, R., Schembri, A., Fenlon, J., Rentelis, R., Woll, B., and Cormier, K. (2014). Lexical variation and change in British sign language. *PLoS ONE*, 9(4).

Wallin, L. and Mesch, J. (2015). Annoteringskonventioner för teckenspråkstexter. Forskning om teckensprk (FOT-rapport) XXIV.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, pages 1556–1559.

Towards an Annotation of Syntactic Structure in the Swedish Sign Language Corpus

Carl Börstell, Mats Wirén, Johanna Mesch & Moa Gärdenfors
Department of Linguistics
Stockholm University
SE-106 91 Stockholm
{calle, mats.wiren, johanna.mesch, moa.gardenfors}@ling.su.se

Abstract

This paper describes on-going work on extending the annotation of the Swedish Sign Language Corpus (SSLC) with a level of syntactic structure. The basic annotation of SSLC in ELAN consists of six tiers: four for sign glosses (two tiers for each signer; one for each of a signer's hands), and two for written Swedish translations (one for each signer). In an additional step by Östling et al. (2015), all glosses of the corpus have been further annotated for parts of speech. Building on the previous steps, we are now developing annotation of clause structure for the corpus, based on meaning and form. We define a clause as a unit in which a predicate asserts something about one or more elements (the arguments). The predicate can be a (possibly serial) verbal or nominal. In addition to predicates and their arguments, criteria for delineating clauses include non-manual features such as body posture, head movement and eye gaze. The goal of this work is to arrive at two additional annotation tier types in the SSLC: one in which the sign language texts are segmented into clauses, and the other in which the individual signs are annotated for their argument types.

Keywords: clause segmentation, annotation, syntactic structure, Swedish Sign Language, corpus

1. Introduction

The number of corpora available for sign languages around the world is constantly increasing, and many of the already existing corpora are expanding, both in terms of token size and in terms of the detail and amount of linguistic annotations that they contain. What seems to be a shared feature of most sign language corpora today is that they minimally contain (i) a lexical segmentation of the sign language texts into individual signs, labeled with sign glosses, and (ii) a written or spoken (audio recorded) translation of the texts. However, segmentations on a clausal level and the inclusion of annotations of the syntactic structure of clauses appear to be lacking from all but the Auslan corpus (Johnston, 2008; Johnston, 2014). This paper deals with the first steps towards such a segmentation and annotation of the Swedish Sign Language Corpus (SSLC).

1.1. Background

Basic syntactic structure has been a topic of research on a number of different sign languages. For instance, establishing a basic constituent order (i.e. SOV, SVO, etc.) as part of the description of individual languages has been done for quite a few sign languages around the world (see Napoli and Sutton-Spence (2014) for a summary). Many such studies have made use of elicited sign language data, often based on a picture-based elicitation task. Even though the procedure has been to use primarily elicited rather than conversational data, the analysis of the data is often not completely straightforward, and a consistent set of criteria to be used in analyses across languages does not exist (Johnston et al., 2007).

Some problems that arise when analyzing a syntactic feature such as constituent ordering include the topic–comment structure found in many sign languages, ellipsis, the splitting of transitive events into multiple intransitive clauses, and the repetition of verbs, sometimes labeled "verb sandwiches" (Fischer and Janis, 1990; Jantunen, 2008; Jantunen, 2013). Furthermore, trying to analyze sign language data from the the assumption of a linear syntax is somewhat problematic, seeing as the gestural–visual modality allows for a higher degree of simultaneity than the spoken modality (Vermeerbergen et al., 2007). This simultaneity also leads to some modality-specific features of the prosody of signed language, such that the various manual and non-manual articulators work together to mark the boundaries of phrases and clauses by prosodic means (Sandler, 1999). Using prosody as visual cues for segmenting sign language utterances has been investigated for some sign languages (Fenlon et al., 2007; Crasborn, 2007). Although using prosodic segmentation as a means of achieving a basic syntactic segmentation of a sign language corpus has been attempted for the SSLC, this was deemed to be too time-consuming and inaccurate to be practical (Börstell et al., 2014). Furthermore, some of the previous research on Swedish Sign Language (SSL) was conducted on the topic of sentence structure, but this was based on a much smaller dataset than the one available today using the SSLC (Bergman and Wallin, 1985). However, in order to conduct further such research on SSL using the SSLC, the data need to be segmented on a clausal level, and the only sign language corpus that does feature such a segmentation and syntactic annotation today, appears to be the Auslan corpus, with the work done entirely by hand (Johnston, 2014).

1.2. The Problem

Many research questions on the structure and use of SSL depend on a linguistic segmentation of the data above the lexical level. This does not only concern research on syntactic structure, but also questions about the lexicon, such as the distribution of certain lexical items in specific contexts. The goal of the project presented here is threefold: first, criteria are formed on which to base the segmentation and annotation work in order to arrive at conventions for conducting this annotation work; second, the SSLC data is segmented into "clauses", in order to achieve a linguistic segmentation above the lexical level; third, the constituents within the clausal segmentations are annotated for syntactic arguments assigned by the predicates in order to get information about argument structure and basic syntactic structure such as constitutent ordering. The work process for the three steps is by no means strictly linear, but rather cyclic, in the sense that the criteria for segmenting and annotating partly arise from the actual segmentation/annotation process, and vice versa. Thus, this paper aims to discuss some of the methodological problems that appeared along the way, as well as some preliminary results of the annotations.

2. Data

The Swedish Sign Language Corpus (SSLC) is a corpus consisting of a collection of sign language texts in `.mpg` format (Mesch et al., 2012b) and its accompanying annotation files in `.eaf` format (Mesch et al., 2015). The texts consist of naturalistic, dyadic signing, the majority of the data coming from conversational type texts, and a smaller part coming from elicited narratives. In total, 300 texts have been recorded, distributed over 42 different signers (Mesch, 2012; Mesch et al., 2012a). These texts are being made available through regular updates online as the video files are being edited and the annotation files completed. The annotation files contain six main tiers: four for the sign glosses (i.e. one for each of the hands of the two signers); two for written Swedish translations (i.e. one tier for each signer) (Mesch and Wallin, 2015). All annotations are made with the ELAN software (Wittenburg et al., 2006), producing annotation cells on tiers time-aligned with the video files. The most recent update of the SSLC contains 48 690 tokens, spanning just over 6 hours of video data, distributed across 85 files and 42 signers. Within the current project, 12 of these files (comprising 3 664 sign tokens in approximately 30 minutes of video data) have thus far been segmented and annotated for syntactic structure.
Besides the sign glosses and translations, the SSLC also features part of speech (PoS) tags, which are attached to the sign gloss annotations on the sign gloss tier (e.g. "PRO1[PN]"). The tagging procedure was initially based on a semiautomatic method on an earlier version of the corpus (Östling et al., 2015), and subsequent expansions have been manually tagged. The PoS tagging is done on the type, rather than token, level, using the PoS categories described in Table 1.

PoS	Tag
Noun	NN
Verb	VB
Adjective	JJ
Adverb	AB
Numeral	RG
Pronoun	PN
Conjunction	KN
Preposition	PP
Verb (depicting)	VBAV
Verb (stative)	VBS
Verb (CA)	VBCA
Verb (locative)	VBPP
Interjection	INTERJ
Point	PEK
Noun classifier	NNKL
Buoy	BOJ
Uncertain	?

Table 1: PoS tags used in the SSLC.

3. Annotation of Clauses

3.1. Segmenting SL Text into Clauses

The first step in working towards a syntactic annotation of the SSLC is to segment the data into clausal units. For this project, we are using the descriptions of basic syntactic structure in Role and Reference Grammar as proposed by Van Valin Jr. and LaPolla (1997) and Van Valin Jr. (2005), in which a clause consists of a predicate, core (obligatory) arguments assigned by the predicate, and a periphery (optional modifiers). The peripheral elements are not part of the syntactic annotation at this stage, however, leaving us with the annotation of the core of the clause, i.e. predicate and obligatory arguments (see section 3.2.). Furthermore, we are currently only annotating the smallest clausal units (with a single semantic predicate per clause). Thus, we do not keep track of the relations between matrix and subordinated clauses, or between coordinated clauses.
It is important to acknowledge the fact that signed language has certain features that do not readily fit into the syntactic structure of spoken language, namely that signed language has the option to *show* situations/events/actions rather than to *tell* about them. Thus, our notion of a clause is very similar to that of Johnston (2014) in that both lexically described situations, and depicted or enacted situations can be instances of clauses (or, in Johnston's terminology *clause-like units*, CLUs). Minimally, our definition of a clause is that it must contain a predicate (verbal, depicted, enacted, or non-verbal). If there are adjacent arguments or obligatory complements associated to a predicate, they are also included in the clause of that predicate. When it comes to the issue of multiple repetitions of arguments or predicates, we follow the criteria of Meir et al. (Submitted) in that multiple predicates are included in the same clause only if (i) they are repetitions of the same sign (with or without morphological alterations such as reduplication (Fischer and Janis, 1990; Bergman and Dahl, 1994)), or (ii) they are semantically related, or near-synonyms, describing the same event/action, such as 'grab' and 'take' (serial predicates).
Apart from these syntactic and semantic criteria, we also include prosody as a way of distinguishing a clause, such that the elements included into a clausal unit should be linearly adjacent within a prosodically uniform sequence. Since prosodic breaks appear on many levels (Sandler, 1999), we allow for smaller prosodic units to differ within a clause

Tag	Description
S	Single intransitive argument
A	Transitive Actor
P	Transitive Undergoer
T	Ditransitive Theme
R	Ditransitive Recipient
V{1,2,3}	Verb (numerals denote order in chain)
Aux	Auxiliary verb
nonV	Non-verbal predicate
Loc	Obligatory locative complement

Table 2: Argument tags used in the SSLC.

(such as a topic–comment structure), but may use layered boundary markers as a criterion for a syntactic break (Börstell et al., 2014). However, since we are only identifying the smallest clausal unit, we do allow for a syntactic break to split a larger prosodic unit, such as dividing a subordinate clause from its matrix clause.

3.2. Annotating Predicates and Arguments

The (single or multiple) predicates of a clause are distinguished according to the criteria in Section 3.1. Our inventory of arguments is based on categories commonly used in comparative and descriptive linguistics, as well as a few ones that were added underway to reflect the particular properties of SSL. The categories are shown in Table 2 and exemplified below in Examples (1)–(6), with annotated clauses obtained from the SSLC.[1]

(1) PRO1 PLAY-BADMINTON
S V
'I played badminton.' (SSLC01_322)

(2) OFTEN PRO1 CALL INTERPRETER
A V P
'I often call for an interpreter.' (SSLC01_322)

(3) POINT.PL GIVE OBJPRO1.PL DISCOUNT
A V R T
'They give us a discount.' (SSLC01_302)

(4) LIE-DOWN(G)@*ca* SLEEP TOSS-AND-TURN
V1 V2 V3
'[He was] tossing and turning.' (SSLC01_332)

(5) SO PRO1 *think-gesture@g* PERF ALWAYS
A Aux
FOR-EXAMPLE PU@*g* GO-INTO STORE
V Loc
'If I have, for instance, gone into a store.' (SSLC01_322)

[1]The sign glosses have been translated into English for the convenience of the reader. The original sign glosses in the SSLC are in Swedish.

(6) PRO1 SNOW^MAN
S nonV
'I am a/the snowman.' (SSLC01_332)

In the past, the S, A, P, T and R categories have been used by different authors alternately for distinguishing universal syntactic functions and thematic/semantic roles (Haspelmath, 2011). Our criteria involve both dimensions; more specifically, while the goal is to annotate syntactic functions, these functions are to a large extent semantically motivated, following Van Valin Jr. and LaPolla (1997) and Van Valin Jr. (2005).
Among the additional categories, V{1,2,3} denotes multiple predicates in the same clause as described in Section 3.1., with labels adopted from the *Auslan Corpus Annotation Guidelines* (Johnston, 2014, 71–72). However, repeated instances of the same predicate will not result in a numeral suffix unless other predicates are part of the same clause. Instead, a repeated predicate will receive the same Argument tier label as the first occurrence, such that it is clear that it is an instance of repetition rather than verbal chains (see Example (7)).

(7) DOG WAG-TAIL HAPPY WAG-TAIL
S V nonV V
'The dog was happy, wagging its tail.' (SSLC01_331)

Similarly, repetitions of arguments are dealt with in the same way, i.e. using the same label for both repetitions. This is also true of cases where multiple *different* signs refer to the same argument referent, a pattern most commonly found in cases in which the signer uses a lexical sign *and* a pointing sign to refer to a certain argument.

3.3. Criteria for Distinguishing Clauses

A summary of the established criteria for distinguishing clauses is as follows:

- A clause is distinguished on semantic grounds as a unit that minimally contains a predicate and its arguments. Syntactically, this corresponds to the *core* in the terminology of Role and Reference Grammar.
- Optional modifiers (peripheral elements) are included in the clause unless they form independent clauses themselves through subordination or coordination.
- Multiple predicates are included in the same clause only if they are formally and/or semantically related and describing the same situation.
- The elements of a clause should fall within a uniform prosodic unit.

These criteria could be contrasted with those for spoken languages such as English or Swedish, where a clause is typically seen as a unit containing at most one finite verb (Ejerhed, 1988), a notion not manifested in signed languages.

Sign order	Tokens
V	476
S V	154
nonV	86
V P	80
S nonV	46
A V P	35
P V	24
Aux V	17
V S	14
nonV P	13
Other	154

Table 3: The most common sign orders in the SSLC.

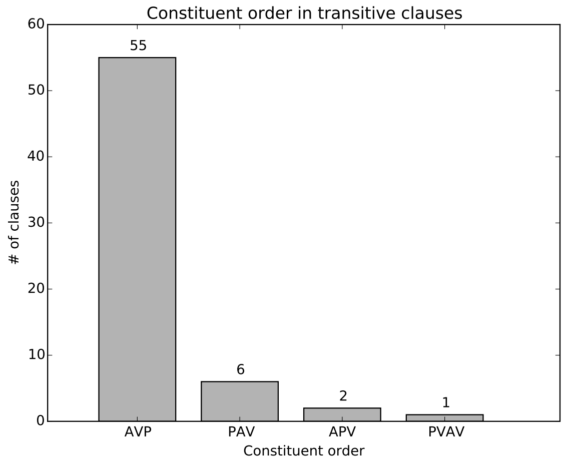


Figure 2: Constituent order in the explicitly two-argument transitive clauses (out of 64 two-argument clauses in total).

3.4. Tiers in ELAN

This annotation work has resulted in the addition of two new tier types in the SSLC: CLU and Argument, respectively. The CLU tier is the tier on which the text is segmented into clauses, and its annotation cells are currently empty, serving only to create a cell that spans the sign gloss annotations on the timeline that are analyzed as constituting a clausal unit. The Argument tier features cells that align with the sign glosses that serve one of the core syntactic functions as given in Table 2. The CLU tier type is used for two tiers in the annotation, one for each signer, and the Argument tier type is used for four tiers, one for each of the signers' hands. Figure 1 illustrates the annotation tiers as they appear in the ELAN interface, with the visible clause being the same as illustrated in Example (2).

3.5. The Structure of Some Basic Clauses in SSL

Having completed a clausal segmentation and syntactic annotation of 12 files of the SSLC thus far, we wanted to do a preliminary investigation of constituent ordering on this small portion of the SSLC data. We wrote a Python script that extracted the annotations contained within clauses (i.e. cells on the CLU tier), combined the Argument tier cells into linear strings showing the ordering of constituents, and tallied the encountered orders. The results were that out of the 1099 clauses segmented in the data, there were 150 distinct orders of predicate–argument tags. In order to clean up the data, we let the script collapse juxtaposed occurrences of the same type, such that the order A V1 V2 P would be rendered as simply A V P, reducing the number of distinct orders to 69. Of these 69 orders, the ten most common ones are listed in Table 3, showing that the most frequent structure is simply a predicate (consisting of one or more verbal signs) without any explicit arguments. This, together with the fact that there are instances of transitive type arguments showing up in clauses without an explicit second argument (e.g. V P), suggests that ellipsis is quite common, such that arguments are readily left out if co-referent with or implied from adjacent clauses.

In an additional step, we wanted to see the structure of transitive clauses for the sake of looking at the basic sign/constituent order in terms of frequency. In order to do this, we further cleaned the data by collapsing the Aux category with V, and extracting only those clauses which contain both an A and a P argument. Looking specifically at the 64 clauses that contain two explicit arguments, we find a strong preference for the A V P order (see Figure 2), which corroborates earlier claims of SSL being a predominantly SVO language (Bergman and Wallin, 1985).

4. Discussion

In this on-going project, we have tried to apply previous research on both spoken and signed language to arrive at a template and well-defined criteria for segmenting and annotating clauses in the SSLC. Some of the potentially problematic cases that we had identified prior to the start of the project, through previous research, were found to be easily dealt with, whereas others are still under discussion and may require further revisions to our criteria and annotation structure. For instance, the simultaneity of manual signs is easily dealt with using the ELAN software, by simply allowing each hand to be associated with its own annotation tier. However, when we wish to extract such data (e.g. for constituent ordering investigations), we have to rely on a linear (temporal) ordering, which we have solved by letting the onset of each element decide the linear ordering. The issue of repetitions of elements (such as "verb sandwiches") and distinguishing same verb repetitions from serial verbs, is handled by using identical or enumerated labels on the Argument tier, respectively, a method which we—at least partly—have adopted from Johnston (2014). The issue of ellipsis seems to pose more of a challenge, and the question of how to deal with this is yet to be solved. In our current annotation scheme, we do not mark cases of ellipsis in any way, although we find the phenomenon to be ubiquitous in our data. An updated annotation scheme under discussion includes the addition of Argument tier labels that function as place-holders for arguments that are explicitly expressed in a text, but not in all clauses for which the ar-

Figure 1: Screen shot of ELAN with the sign gloss, clause segmentation, syntactic annotation, and translation tiers.

gument is co-referent. Such annotations could help resolve some questions with regard to constituent ordering, but also the argument structure of individual verbs.

5. Conclusion

We have described our preliminary annotation of syntactic structure in the SSLC, thus far comprising segmentation of clauses as well as annotation of predicates and obligatory arguments in 12 files of the corpus. In addition to annotating more data, we plan to extend this work by including optional modifiers (elements of the syntactic periphery) on the Argument tier, and by introducing an additional tier on which the relations between matrix and subordinated clauses on the one hand and coordinated clauses on the other are annotated. The ultimate goal of this work is to arrive at a syntactic annotation which is sufficiently well worked out to allow for a mapping to a standard formalism in language technology, such as dependency grammar (Tesnière, 1959). In addition to being a functional formalism, and thus akin to Role and Reference Grammar, this is currently being subject to standardization for the purpose of multilingual treebank annotation in the form of Universal Dependencies (`http://universaldependencies.org`). So far, this has been used for around 50 spoken languages, and would constitute an interesting touchstone for the work on syntactic annotation attempted here.

6. Acknowledgements

This work has been supported by an infrastructure grant from the Swedish Research Council (SWE-CLARIN, project 821-2013-2003).

7. Bibliographical References

Bergman, B. and Dahl, Ö. (1994). Ideophones in Sign Language? The place of reduplication in the tense-aspect system of Swedish Sign Language. In Carl Bache, et al., editors, *Tense, Aspect and Action. Empirical and Theoretical Contributions to Language Typlogy*, pages 397–422. Mouton de Gruyter.

Bergman, B. and Wallin, L. (1985). Sentence structure in Swedish Sign Language. In William C. Stokoe et al., editors, *Sign language research '83*, pages 217–225, Silver Spring, MD. Linstok Press.

Börstell, C., Mesch, J., and Wallin, L. (2014). Segmenting the Swedish Sign Language Corpus: On the possibilities of using visual cues as a basis for syntactic segmentation. In Onno Crasborn, et al., editors, *Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel [Language Resources and Evaluation Conference (LREC)]*, pages 7–10, Paris. European Language Resources Association (ELRA).

Crasborn, O. (2007). How to recognise a sentence when you see one. *Sign Language & Linguistics*, 10(2):103–111.

Ejerhed, E. (1988). Finding clauses in unrestricted text by finitary and stochastic methods. In *Proceedings of the Second Conference on Applied Natural Language Processing*, pages 219–227. ACL.

Fenlon, J., Denmark, T., Campbell, R., and Woll, B. (2007). Seeing sentence boundaries. *Sign Language & Linguistics*, 10(2):177–200.

Fischer, S. D. and Janis, W. (1990). Verb sandwiches in

American Sign Language. In Siegmund Prillwitz et al., editors, *Current trends in European sign language research*, number 2, pages 279–293, Hamburg. Signum Verlag.

Haspelmath, M. (2011). On S, A, P, T, and R as comparative concepts for alignment typology. *Linguistic Typology*, 15(15):535–567.

Jantunen, T. (2008). Fixed and free: Order of the verbal predicate and its core arguments in declarative transitive clauses in Finnish Sign Language. *SKY Journal of Linguistics*, 21(1):83–123.

Jantunen, T. (2013). Ellipsis in Finnish Sign Language. *Nordic Journal of Linguistics*, 36(3):303–332.

Johnston, T., Vermeerbergen, M., Schembri, A., and Leeson, L. (2007). 'Real data are messy': Considering cross-linguistic analysis of constituent ordering in Auslan, VGT, and ISL. In Pamela M. Perniss, et al., editors, *Visible variation: Cross-linguistic studies in sign language structure*, pages 163–205. Mouton de Gruyter, Berlin/New York, NY.

Johnston, T. (2008). The Auslan Archive and Corpus. In David Nathan, editor, *The endangered languages archive*. Hans Rausing Endangered Languages Documentation Project, School of Oriental and African Studies, University of London, London.

Johnston, T. (2014). Auslan Corpus Annotation Guidelines. Auslan Signbank. `http://new.auslan.org.au/about/annotations/`.

Meir, I., Aronoff, M., Börstell, C., Hwang, S.-O., Ilkbasaran, D., Kastner, I., Lepic, R., Lifshitz Ben Basat, A., Padden, C., and Sandler, W. (Submitted). The effect of being human and the basis of grammatical word order: Insights from novel communication systems and young sign languages.

Mesch, J. and Wallin, L. (2015). Gloss annotations in the Swedish Sign Language Corpus. *International Journal of Corpus Linguistics*, 20(1):103–121.

Mesch, J., Wallin, L., and Björkstrand, T. (2012a). Sign Language Resources in Sweden: Dictionary and Corpus. In Onno Crasborn, et al., editors, *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon [Language Resources and Evaluation Conference (LREC)]*, pages 127–130, Paris. European Language Resources Association (ELRA).

Mesch, J., Wallin, L., Nilsson, A.-L., and Bergman, B. (2012b). Dataset. Swedish Sign Language Corpus project 2009–2011. Version 1. `http://www.ling.su.se/teckensprakskorpus`.

Mesch, J., Rohdell, M., and Wallin, L. (2015). Annotated files for the Swedish Sign Language Corpus. Version 3. `http://www.ling.su.se/teckensprakskorpus`.

Mesch, J. (2012). Swedish Sign Language Corpus. *Deaf Studies Digital Journal*, 3. `http://dsdj.gallaudet.edu/index.php?issue=4&section_id=2&entry_id=128`.

Napoli, D. J. and Sutton-Spence, R. (2014). Order of the major constituents in sign languages: Implications for all language. *Frontiers in Psychology*, 5:1–18.

Östling, R., Börstell, C., and Wallin, L. (2015). Enriching the Swedish Sign Language Corpus with part of speech tags using joint Bayesian word alignment and annotation transfer. In Beáta Megyesi, editor, *Proceedings of the 20th Nordic Conference on Computational Linguistics (NODALIDA 2015), NEALT Proceedings Series 23*, pages 263–268, Vilnius. ACL Anthology.

Sandler, W. (1999). Prosody in two natural language modalities. *Language and Speech*, 42(2-3):127–142.

Tesnière, L. (1959). *Eléments de syntaxe structurale*. Klinksieck, Paris.

Van Valin Jr., R. D. and LaPolla, R. J. (1997). *Syntax: Structure, meaning, and function*. Cambridge University Press, Cambridge.

Van Valin Jr., R. D. (2005). *Exploring the syntax-semantics interface*. Cambridge University Press, New York, NY.

Myriam Vermeerbergen, et al., editors. (2007). *Simultaneity in signed languages: Form and function*. John Benjamins, Amsterdam/Philadelphia, PA.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, pages 1556–1559.

Preventing Too Many Cooks from Spoiling the Broth: Some Questions and Suggestions for Collaboration between Projects in iLex

Penny Boyes Braem* and Sarah Ebling**

*Center for Sign Language Research, Basel
E-mail: boyesbraem@fzgresearch.org

**Institute of Computational Linguistics, Univ. of Zurich
E-mail: ebling@cl.uzh.ch

Abstract

Collaborative development of sign language resources is fortunately becoming increasingly common. In the spirit of collaboration, having one shared lexicon for sign language projects is a big advantage. However, this poses challenges to aspects pertaining to consistency of data, privacy of informants, and intellectual property. This contribution points out some problems that arise, especially if the common data comes from projects of different institutions. We describe what we have found to be a sustainable legal framework for our collaborative iLex corpus lexicon, giving an overview of the different kinds of partners involved in the creation and exploitation of a shared iLex corpus lexicon and providing our answers to the questions we faced along with an outlook for the future.

Keywords: iLex, collaboration, framework, users, legal agreements

1. Introduction

Collaborative development of sign language resources is fortunately becoming increasingly common. With all the advantages that such cooperation clearly brings, there are also new problems that arise, especially if the common data comes from projects of different institutions. In Switzerland, research on Swiss German Sign Language (DSGS) is dispersed, with several smaller projects being or having been carried out at different institutions.

In the spirit of collaboration, having one shared lexicon for these projects is, of course, a big advantage. However, this poses challenges to general issues of consistency of data, privacy of informants, and intellectual property.

The main research database for DSGS has recently been migrated from FileMaker to iLex, a software tool for creating and analyzing sign language lexicons and corpora (Hanke & Storz, 2008). In iLex, the sign tokens introduced as part of transcripts of individual projects appear as corpus evidence of sign types in a shared lexicon. For example, Figure 1 shows the tokens corresponding to the type GLAUBEN_1A [*BELIEVE_1A]* in the DSGS instance of iLex (henceforth referred to as "iLex-DSGS").

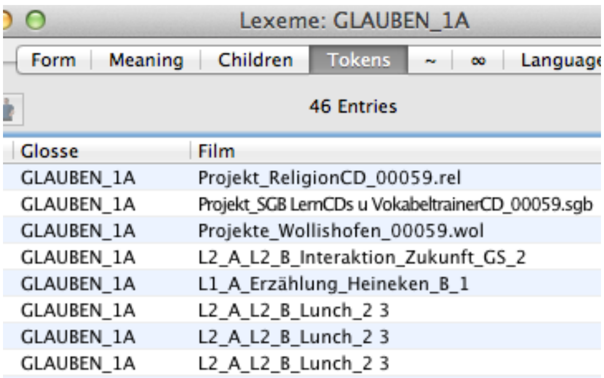


Figure 1: Tokens corresponding to the type GLAUBEN_1A [*BELIEVE_1A*] in iLex-DSGS

While it is technically possible to restrict the display of tokens from the linked corpus of annotated videos to members of a specific project, this would drastically reduce the benefit of a corpus-based lexicon, which is to obtain information on the use of a sign type in different contexts (i.e., from the data from different projects).

The advantage of a shared lexicon in the iLex software over other sign language lexicon and/or corpus tools, however, also means that in the relational database of iLex, changes of sign types in the shared lexicon affect all tokens of the same type in all linked corpus transcripts from different projects. For example, if a member of a hypothetical Project P were to change the gloss GLAUBEN_1A introduced by a previous project to MEINEN_1 [*SUPPOSE_1*], all 28 tokens of the sign in all transcripts (shown in Figure 1) would automatically be changed to that new gloss label. This is an aspect to be clearly explained to partner projects, even though in iLex there is the possibility of storing the old gloss for a sign in the metadata.

The prospect of combining data over a long stretch of time from different projects, many of them from different institutions, into the iLex-DSGS database has brought to the forefront several more general questions related to collaborative resource production and exploitation, including the following:

- What happens after a project ends? Will its members still have access to iLex-DSGS?
- How can we make sure that data that has been created in iLex-DSGS stays there after a project has ended?
- How can we ensure that proper informed consent is available for all informant-related data imported into iLex-DSGS?
- How can we make sure that minimum standards pertaining to data creation are adhered to in iLex-DSGS?
- How can we make sure that minimum standards pertaining to data security are followed?
- Who can *use* which data in iLex-DSGS when and under which conditions for publications, presentations, teaching, etc.?

- Who can *modify* which data in iLex-DSGS when and under which conditions?
- Who can *delete* which data in iLex-DSGS when and under which conditions?
- How can we have control over who accesses iLex-DSGS, while still allowing users to share data with their colleagues of the same institution in a low-threshold manner?
- Who decides over all of the above questions?

Here, we describe what we have found to be a sustainable legal framework for our collaborative corpus lexicon. In the following sections, we give an overview of how the different kinds of partners are involved in the creation and exploitation of iLex-DSGS (Section 2) and provide our answers to our posed questions (Section 3) as well as an outlook for the future (Section 4).

2. Kinds of Partners Involved in the Collaboration

We found it necessary to define the following different types of partners who we foresee being involved in collaborative use of iLex-DSGS (Figure 2):

- **Data producers and users**: These are partners who are creating data in iLex, i.e., producing textual data (notations, annotations, metadata, etc.) as well as introducing references to videos (of signed utterances, individual signs, etc.) and images (illustrations, supporting materials), etc. These partners would also like to use data of other projects in iLex-DSGS. They need both reading and writing privileges in iLex-DSGS.
- **Data contributors**: These are partners who have previously created DSGS video and other data outside of iLex and have agreed for their data to be included in iLex-DSGS, but who do not wish to access iLex-DSGS themselves. Examples for possible data contributors for us would be the Swiss German Sign Language interpreter and teacher training programs, an on-line television program for Swiss German Deaf persons, as well as students who have completed research projects at the BA, MA and PhD levels.
- **Data users**: These are partners who would like to use existing data from iLex-DSGS for their projects while not creating additional data. These partners require read-only access to iLex. Foreseeable data users here are sign language teacher trainers, interpreter trainers, students, and researchers.
- A small group of experienced sign language researchers responsible for technical maintenance and quality assurance of iLex-DSGS, which in our framework constitute the **oversight committee.**

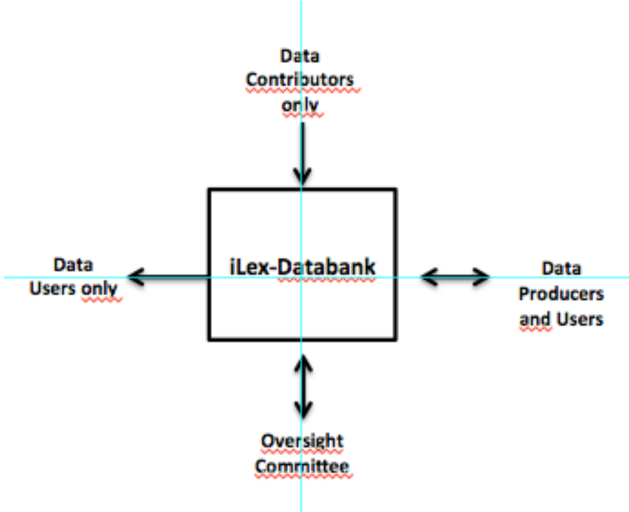


Figure 2: Kinds of Partners involved in the collaboration

3. Our Current Solutions for Our Questions

In order to answer the general questions posed previously, we have established a framework for collaboration (Figure 3). This framework consists of three tiers: a consortium at the top, the individual collaborating projects at the bottom, and an oversight committee in-between.

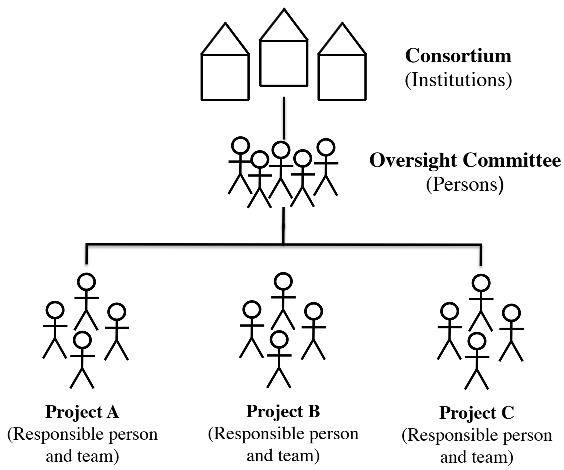


Figure 3: Structure of the collaborative iLex-DSGS framework.

The consortium, oversight committee, and the participating projects all have specific rights and duties, which are discussed below. These rights and duties are specified in the legal agreement forms we have drawn up with legal advice from the university where the iLex-DSGS database is hosted, with slightly different agreement forms depending on the type of user. (Examples of these forms are available by request to the authors.)

3.1 The Consortium

At the apex of the framework we have designed for the iLex-DSGS database is a consortium composed of the three main institutions of higher education or research who are presently the major contributors to and users of the database. All the participating institutions sign an

agreement to be in this consortium.
The database is housed on a server of one of these institutions, which is called the "leading house" in the legal agreements. This institution's personnel for server maintenance, backups, and updating also provide these services for iLex-DSGS.

3.2 Collaborating Projects

3.2.1. Contributors and Users of Data Created Within iLex-DSGS

Projects wishing both to create and to use data within iLex-DSGS need to provide a list of users who should have both reading and writing access to iLex-DSGS. Sharing data from iLex-DSGS within the partner institution of the project is possible with prior consent from the oversight committee. Project coordinators are asked to inform their members that iLex-DSGS login details should be treated confidentially and, in particular, not be sent through unencrypted e-mail exchange. The project coordinator will also be asked to confirm that proper informed consent is present for all data incorporated into iLex-DSGS. The project coordinator will be asked for permission for the data to remain in iLex-DSGS after the end of a project. In exchange, the members of the project will have access to iLex-DSGS for a specific number of years beyond the lifetime of a project itself (agreements are renewable). As mentioned in Section 3.1, modification or deletion of data from other projects in iLex-DSGS requires prior consent of the iLex oversight committee. Using data from other projects for publications, presentations, teaching, etc. requires prior consent of the person listed as responsible for the other project. This person will also determine how the data is to be cited and can ask for any anonymization of the data that might be necessary.

3.2.2. Contributors of Data Created Outside of iLex-DSGS

Partners who are contributing data created outside of iLex-DSGS will be asked for permission to permanently store the data in iLex-DSGS. They need to confirm that the necessary informed consent has been obtained. The contributors do not incorporate the data into iLex-DSGS themselves; this is done by members of the oversight committee.

3.2.3. Users of Data in iLex-DSGS

These are partners who are granted read-only access to iLex. Since they do not produce data in iLex-DSGS, many of the precautions mentioned earlier do not need to be taken in any agreements made with them.

3.3. Oversight Committee: Technical Maintenance and Quality Assurance

The interface between the consortium and individual projects is a small iLex-DSGS oversight committee. This group is responsible to the consortium for technical maintenance and quality assurance. The committee is composed of Deaf and hearing researchers experienced in iLex. It includes computational and sign language linguists as well as sign language teachers and interpreter trainers experienced in research. All committee members are both producers and users of iLex-DSGS.

The oversight committee has the following responsibilities:

- Creation of iLex-DSGS user accounts;
- Definition of the maximum amount of disk space available for each project on the server on which the iLex-DSGS database resides;
- Organization of obligatory training courses that contribute to quality assurance through the following guidelines, which are to be made available through training courses for the project team and through an iLex-DSGS on-line Wiki:
  - General introduction to iLex and iLex-DSGS;
  - Explanation of (an-)notation conventions, especially glossing and form notation conventions;
  - Creation and explanation of informed consent form proposed for use in all projects;
  - Explanation of quality standards for primary and secondary data created in iLex-DSGS. (The quality standards include, for example, a four-eyes principle for notations of sign forms.)
- Giving of final approval of the changing or deleting of lexicon data contributed by other users. This is necessary as new project team members might not be aware of all the existing signs in the lexicon, which might have slightly different glosses. It is also necessary to check that any new glosses conform to the iLex-DSGS glossing conventions, particularly for different types of variants;
- Incorporation of existing data from external contributors to the iLex-DSGS database;
- Correspondence about iLex-DSGS with the Deaf community, outside researchers, and other interested parties.

4. Outlook

Underway now are projects that involve adding to the iLex-DSGS corpus lexicon older DSGS data that had been annotated in Excel. This will necessitate, of course, manual checking that the information – especially the glossing – conforms to the iLex-DSGS conventions. The recent reprogramming of an existing on-line lexicon for technical terms based on iLex has greatly facilitated the correcting and updating of this product (see Ebling & Boyes Braem in the proceedings of this workshop). We plan to expand this use of iLex-DSGS as a base for on-line lexicons for a wider range of terms (linguistic, place and proper names, jurisprudence, medicine).
There is also the possibility of expanding iLex-DSGS such that it becomes "iLex-CH", which would include all three sign languages used in Switzerland (Swiss German, Swiss French, and Swiss Italian sign languages).
High on our agenda is the investigation of appropriate financing for a sustainable collaborative corpus lexicon. We are in the process of specifying how to share the costs, particularly of the work of the oversight committee, be-

tween participating projects and outside financing.
The framework of this collaboration, as well as the agreements we have formulated, including the measures they involve, will be tested over the next few years as new and different kinds of partners join the collaboration.
We already have received feedback that, in addition to the different kinds of agreements we have prepared, a financial agreement that secures the long-term maintenance of iLex-DSGS would be wished by some cooperating projects. The question has also arisen of whether all the videos, which a project intends to annotate should be stored on a leading house server or locally. Also desired would be more details about what happens when the leading house does not fulfill its obligations (due perhaps to a change in personnel). The process of 'gearing up' for working with iLex can also entail expenses for necessary infrastructure, server space, Internet connection, and general technical guidance in the process of setting up the project. Additional information must be provided from the side of the oversight committee concerning these technical questions.
While our questions and our current solutions, as well as the still open questions, are tailored to our local situation, they might also be helpful suggestions for research teams in other countries who are facing similar problems in this exciting but challenging new age of digital humanities.

5. Acknowledgements

We are grateful to legal experts at the university hosting iLex-DSGS for their help in the final stages of formulating the solutions we have discussed into a set of separate agreement forms for the different kinds of partners.

6. Bibliographical References

Hanke, T. Storz, J. (2008). iLex: A database tool for integrating sign language corpus linguistics and sign language lexicography. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC) Marrakech, Morocco,* pp. 64-76.

Community Input on Re-consenting for Data Sharing

Deborah Chen Pichler[1,3], Julie Hochgesang[1,3], Doreen Simons[2,3], Diane Lillo-Martin[2,3]
[1]Gallaudet University; [2]The University of Connecticut, [3]Haskins Laboratories
[1]Washington, DC 20002; [2]Storrs, CT 06269-1145
E-mail: {deborah.pichler, julie.hochgesang}@gallaudet.edu; {diane.lillo-martin, doreen.simons}@uconn.edu

Abstract

Development of large sign language corpora is on the rise, and online sharing of such corpora promises unprecedented access to high quality sign language data, with significant time-saving benefits for sign language acquisition research. Yet data sharing also brings complex logistical challenges for which few standardized practices exist, particularly with regard to the protection of participant rights. Although some ethical guidelines have been established for large-scale archiving of spoken or transcribed language data, not all of these are feasible for sign language video data, especially given the relatively small and historically vulnerable communities from which sign language data are typically collected. Our primary focus is the process of re-consenting participants whose original informed consent did not address the possibility of sharing their video data. We describe efforts to develop ethically sound, community-supported practices for data sharing and archiving, summarizing feedback collected from two focus groups including a cross-section of community stakeholders. Finally, we discuss general themes that emerged from the focus groups, placing them in the wider context of similar discussions previously published by other researchers grappling with these same issues, with the goal of contributing to best-practices guidelines for data archiving and sharing in the sign language research community.

Keywords: language documentation and long-term accessibility for sign language data; experiences in building sign language corpora, ASL, child acquisition

1. Introduction

Development of large sign language corpora is on the rise, and online sharing of such corpora promises unprecedented access to high quality sign language data. For researchers studying early language development, having ready access to longitudinal video data means that many research questions can be tested immediately, on data from multiple children, without the time-consuming prerequisite of subject recruitment, filming and video annotation over the relevant age range. Considering the time and effort required to collect and process longitudinal data from just a single child, the time-saving benefits of shared online corpora clearly has potential to revolutionize the way sign language acquisition research is conducted ("economization of resources" as described in Himmelmann, 2006).

Yet the same long-term data infrastructure that promises such accessibility also brings with it complex logistical challenges for which few standardized practices currently exist. Some of the greatest challenges revolve around the protection of participant rights. Although some ethical guidelines have been established for large-scale archiving of spoken or transcribed language data (e.g. the CHILDES database; MacWhinney, 2000), not all of these are feasible for sign language video data, especially given the relatively small and historically vulnerable communities from which sign language data are typically collected.

Our primary focus in this paper is the process of re-consenting participants whose original informed consent did not address the possibility of sharing their video data. We describe our efforts to develop ethically sound, community-supported practices for data sharing and archiving. Our discussion is focused on video data collected two decades ago from a longitudinal spontaneous production study of the acquisition of American Sign Language (ASL), but the issues and recommendations outlined here are equally relevant to any situation in which video data are shared with a wider audience than initially intended. Below, we introduce the set of longitudinal video data that we plan to share, and outline the anticipated steps for obtaining re-consent from filming participants. We then summarize outcomes of two focus group events in which we sought feedback from a cross-section of community stakeholders. Finally, we discuss general themes that emerged from the focus groups, placing them in the wider context of similar discussions previously published by other researchers grappling with these same issues, with the goal of contributing to best-practices guidelines for data archiving and sharing in the sign language research community.

2. Background

Our immediate context for addressing the issues of this discussion is a body of naturalistic video footage collected longitudinally from four deaf children and their deaf families, between ages 1;05-4;02 (years; months) (Lillo-Martin and Chen Pichler, 2008). The children were filmed in their homes or other familiar locations at intervals ranging from one week to two months. Because all four children were under the age of 5 at the time of filming, their parents provided signed consent for the children's participation. The video data have been painstakingly annotated in different ways over the past twenty years, and "basic transcription" will soon be available for a large portion of the sessions, including ID glosses for individual signs and free translations for all utterances by the target children and their various interlocutors. A screenshot of an example transcript for our project along with text balloons exemplifying our annotation conventions is shown in Figure 1. These basic transcriptions, along with their accompanying video files, are slated for digital archiving in the future at a databank that will be monitored and restricted to academic use, from where they can be shared with researchers pursuing a wide variety of topics related to sign language development.

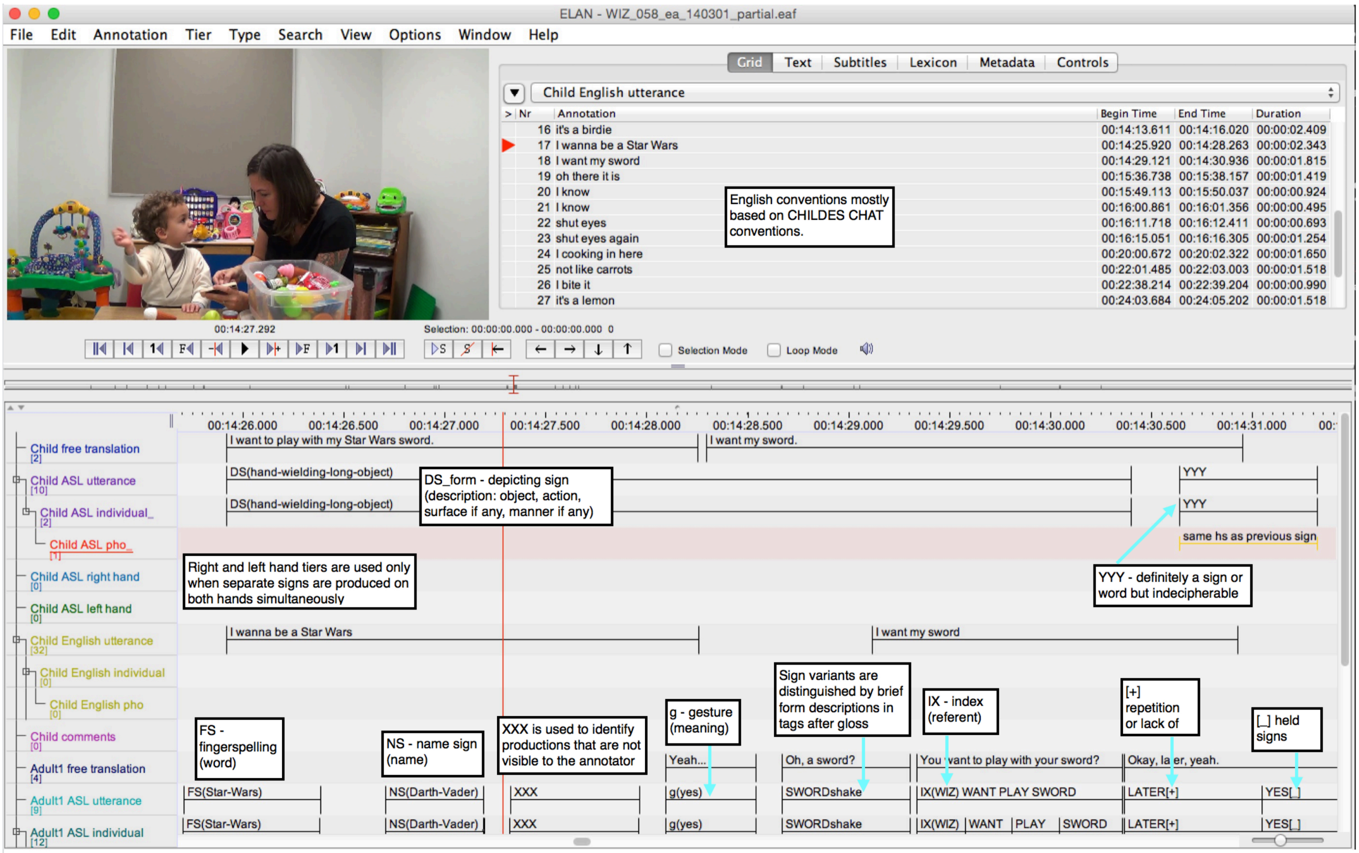


Figure 1. Example of our project's "basic transcript" with ID glosses and free translations

Subsequently, other similar video data of sign language acquisition may also be shared in this way.

However, before sharing the video data and basic transcripts, we must first locate and obtain consent from individuals appearing on video. This re-consenting is no trivial task, given that the data were collected between 1991 and 1999, since which time the target children have grown to adulthood and moved away. The task is further complicated by the many individuals who interacted with the target children on our video footage, ranging from research assistants and the children's immediate family members, to occasional friends and neighbors who appear only sporadically on camera – and in some cases only a portion of their bodies may be visible because they are largely out of the camera's range. Informed consent procedures at the time did not require signed consent from anyone beyond the target children (or parents granting consent in place of target children) so we do not have contact information for most of these "incidental appearances." Thus we must also establish guidelines for determining who requires (re-)consent and what must be done if individuals can not be located or do not grant consent for their video footage to be archived and shared. And finally, we need to determine what measures are deemed necessary by the stakeholder community before they will be comfortable with data archiving and sharing. Individual preferences vary widely, and it is clearly not possible to accommodate the wishes of everyone. Nevertheless, stakeholders in the Deaf community have traditionally had little input on issues of how their video data are used and shared in the long term, so their inclusion in this discussion is critically important.

3. Focus Groups

In view of the questions raised here, we convened two focus groups to collect community feedback on issues of data sharing and re-consent. The focus groups took place at Gallaudet University in Washington, D.C., and the American School for the Deaf in West Hartford, Connecticut. Participants were selected from the following groups, identified as stakeholders because of their participation (actual or potential) as subjects or parents of longitudinal filming, and/or their interest as researchers in collecting or analyzing sign language longitudinal data:

1. Deaf of Deaf adults who participated or could have participated in longitudinal video collection for research purposes when they were children
2. Signing family members of Deaf or Koda children
3. Researchers interested in sign language video data
4. Current and former research assistants on projects related to sign language

Each focus group began with a summary of the important role longitudinal data have played in acquisition research and the value of sharing data more widely. It was emphasized that the videos would be shared through online archives maintained by academic institutions, in stark contrast to unmonitored online sharing on YouTube or other forms of social media. Then participants were presented with question prompts targeting selected issues surrounding video data sharing:

- How comfortable are people in the Deaf community with the idea of their videos appearing online? Has the rise of social media made people become more accepting of their videos online or more cautious?

- What types of footage from longitudinal, spontaneous filming might potentially be embarrassing for subjects? How should such cases be addressed?
- If longitudinal videos are shared beyond the original research team, who should have access to them, and what are some ways in which researchers could potentially use them?
- When an outside researcher requests access to shared longitudinal video data, what information should be collected from them? How much of this information do you think should be available to subjects appearing in the shared videos?
- Of the many individuals aside from the target children who appear in longitudinal videos, who should be contacted for re-consent? Does any appearance on video warrant re-consent, or only those that exceed a specific level of frequency and prominence? Should members of the research team (including assistants hired to film and interact with children on video, but not necessarily to analyze the resulting footage) be contacted for re-consent, or is their consent tacit in their role as filmers and experimenters?
- At what age should children be expected to give (re-)consent for longitudinal filming and/or sharing of previously filmed video data?
- What should be done with "orphan works" (O'Meara and Good, 2010), or data from an individual that has died, or otherwise cannot be reached, or does not give consent for video to be used? Should procedures differ for primary subjects (e.g. target children and those with whom they most often interacted) and incidental subjects (e.g. classmates in the background, occasional visitors)?
- What measures and safeguards should researchers establish for Deaf community members to feel comfortable with longitudinal video collection and sharing?

4. General Findings (Themes)

The two focus group discussions yielded a wide range of opinions concerning some of the target topics listed in (2), and almost no opinions concerning others, with high variability across viewpoints. This was all expected, given the disparate composition of the focus groups. Nevertheless, several important points emerged for which a consensus or dominant opinion could be identified across one or both focus groups or subsets of focus group participants. In this section, we summarize discussion on three of those points.

4.1 General Acceptance of Video Data Sharing

Video data sharing was generally viewed positively, with participants regarding research as valuable to the community. Participants acknowledged the utility of sharing data and were in principle supportive. At the same time, many participants emphasized that Deaf communities are small and participants are never anonymous on film, so sharing of sign language data requires a higher level of precautions than is typical in the majority spoken language community. Not surprisingly, younger (i.e. 18-25 years old) participants reported less anxiety about the idea of their videos being online, and parents reported much more anxiety about their young children's videos being online than their own. Having personal contact with researchers and periodic updates emerged as a crucial mitigating factor; several parents reported that they trusted the researchers who filmed their children because these researchers had met with them in person to explain what their research goals were and periodically presented updates on their findings. Periodic updates were regarded as more than simple professional courtesy, as expressed by one parent who recalled, "On the consent forms, I checked 'Yes, yes, yes…' straight down the whole page, everything was fine with me. But I expect to be contacted every now and then with updates. Don't come and film us then just disappear." Periodic updates from researchers not only inform participants on what has been done with their data and keeps contact information updated, they also provide a tangible illustration of how the video data benefit their community.

4.2 Deciding Who Should be Re-consented and When

Opinions were split on when children should be allowed or expected to give consent. On the one hand, parents were in favor of respecting the wishes of their children, even those who are still minors, if they chose to withdraw their video data from research analysis. At the same time, participants recognized that children and teenagers may not yet fully understand or appreciate the importance of research, so they suggested not destroying any data in these cases, but simply suspending further analysis of them until the subject reached 18 years and had the opportunity to revise their preferences on the consent forms. Opinions were also mixed on whether research assistants on longitudinal filming projects should be re-consented in the same way as target children and their families, or whether individuals in the former group a priori give consent for their video footage to be analyzed and shared when they accept their positions as filmers and experimenters. Clearly, this topic warrants much further discussion; a conservative approach would include them in the re-consenting process.

4.3 Measures That Increase Comfort Level with Video Data Sharing

Some participants were willing to allow data to be used by authorized researchers in any scientifically appropriate way. However, others expressed the opinion that videos in which they or their children appear should only be used for research towards certain goals (e.g., promotion of the use of a sign language). Those participants supported the idea of data archives collecting information from any researchers requesting access to video data and making it available to subject families. Suggested information included the researcher's name and institution, research history, and involvement in the Deaf community. This information might be posted on a user list associated with the data archive, to which participants could have access.

Preferences expressed by families on their original video consent forms regarding such questions as whether clips can be used in scientific presentations should naturally extend to any researcher obtaining the video data from an archive, and families should have the option to change or update those preferences at any time (Harris, Holmes and Mertens, 2009). One parent declared, "It's an exchange. We as research participants give up our privacy in allowing ourselves to be filmed, but in return, the researchers must respect our preferences and wishes." Focus group participants also emphasized the responsibility that researchers have to train their students and assistants in responsible conduct of research, again citing the small size of the Deaf community as reason to be extra sensitive when sharing data from signing families.

As for embarrassing moments on video (e.g. children's temper tantrums (Figure 2); parents losing their temper; parents or children caught on tape in various states of undress; occasional voicing from individuals who normally refrain from using their voice in public, etc.), participants generally agreed that families should have the right to request that certain segments of the video be excised, particularly if those clips have little research significance.

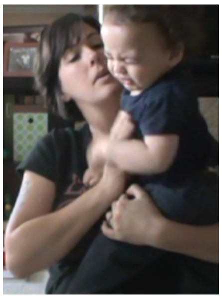


Figure 2. Temper tantrum visible on camera

Since researchers must perform a general review of all video footage as a prerequisite for archiving in a shared database, focus group participants agreed that identification and deletion of embarrassing segments could be undertaken at the same time, and would not necessarily need to involve participant families. Families could indicate the types of activities they would like to have excised (e.g., breastfeeding), and they would then trust the judgment of the researchers in finding and excising appropriate segments.

Another practice that was heavily favored by focus group participants was to include three options in the video release form regarding permission for video clips to be shared or shown in public: a) broad permission; b) only after the subject has the opportunity to view them and give consent for each one; or c) no public viewing (see example item in (1)).

> (1) May we show short clips of video footage including you as part of scientific publications resulting from this research?
> ___ Yes, you may do so without further approval from me.
> ___ Yes, but only with my prior approval of each clip that you plan to share.
> ___ No.

This option was included in the video release form signed by all focus group participants, and many commented that it made them more comfortable about giving consent for their data to be shared. One participant stated, "If the only choices I have on a video release form are "Yes" and "No," that's tough to make a decision...That third option gives me the opportunity to say, "Oh yeah, that clip is fine, you can do whatever you want with it, " or "No, no, that clip is embarrassing, I'd rather keep it private." Having that option eliminates a lot of deliberation, it's really nice."

5. Preliminary Recommendations, With Consideration of Previous Proposals

The goal of the two focus group discussions was to sample the varied opinions surrounding archiving and sharing of potentially sensitive video data involving Deaf children and their families, in the hope that they would direct us in the development of guidelines for best practices in this area. As mentioned earlier, other researchers have previously raised similar issues, providing us with a broader context in which to consider our focus group findings. In the field of spoken language acquisition, large databases of archived longitudinal data already exist, perhaps the most notable being the CHILDES database (MacWhinney, 2000). Digital archives of sign language video data are also increasingly common, including at least two that include longitudinal data from child signers (VALID Data Archive (Klatter et al., 2014) and the IPROSLA data sets (Crasborn et al., 2015)). Outside the domain of language acquisition, many researchers working with endangered spoken languages have established archives of digital language resources as part of language documentation and maintenance efforts (Himmelmann, 2006). All of these groups have wrestled with issues of re-consent and data sharing, and in this section, we will comment on how our findings fit with the discussions that have already emerged from those other groups.

Some of the issues listed in (2) are relatively circumscribed, and for those, it is fairly easy to identify existing practices that are directly relevant for our purposes. For example, with respect to (re-)consenting minors, researchers archiving NGT (Sign Language of the Netherlands) data follow the practice supported by our focus groups of not collecting consent from children until they are 18 years old or older; until then, children's parents give consent for them. Baker (2012) notes that once children reach 18, they must have the right to withdraw consent for their data to be used, if they so choose, also in line with the sentiments of our focus groups.

Our focus group participants' views also generally aligned with previous proposals on the topic of data anonymization. In fact, there appears to be virtually unanimous agreement that total anonymization, long taken as a standard practice for medical data, is not feasible for language data that include audio and/or video components. Distortion of faces or voices compromises the usefulness of language data too dramatically to be a viable option (Crasborn, 2010; O'Meara and Good, 2010; Baker, 2012). Crasborn (2010) suggests that development of life-like sign avatars may offer a solution in the future, but for now, a better solution is to accept the fact that sign participants' identities will not be anonymous and establish guidelines to ensure that participants are aware of this fact, and have options to deal with various related eventualities. Crasborn (2010) details a series of steps taken by researchers on the NGT corpus project to ensure that participants were fully aware of the implications of their video being freely available online for perpetuity before consenting. Additionally, participants were given a DVD copy of their video data after filming, with instructions to review it carefully and inform researchers if there were any segments that they wished to be excluded from sharing. In theory this practice could be applied to the cases of embarrassing footage discussed by our focus groups, although asking participants to review all their footage, which might involve over a hundred videos from a single family, would be less feasible for longitudinal data. An attractive alternative proposed during our focus groups was discussed earlier in the section on "Measures that increase comfort level with video data sharing".

Establishing graded access levels for shared data is another current best practice that would address potential concerns about data sharing, since participants may accept their footage being viewed by the researcher(s) who originally collected the data, but object to the same footage being shared with a wider audience. Among our focus groups, the idea of graded access levels was suggested as a measure that would increase stakeholders' acceptance of large-scale data sharing. Figure 3 shows an example of graded access, taken from the Endangered Languages Archives at SOAS (http://elar.soas.ac.uk, last accessed March 2016).

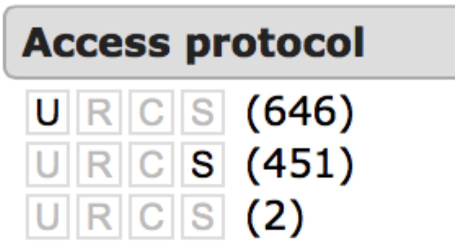


Figure 3. Screenshot of "Access Protocol" from Endangered Languages Archives at SOAS

In the access protocol featured in Figure 2, "U" is for "ordinary user", "R" for "researcher", "C" for "community member", and "S" for "subscriber" (see http://www.elar-archive.org/using-elar/access-protocol.php for more, last accessed March 2016).

One approach to applying graded access is to focus on the qualifications of the researcher requesting access to the data. This is the type of system that was discussed at our focus groups, and it appears to be a common option for other language archives, too (as demonstrated in Figure 2). However, this kind of graded access typically prioritizes access for individuals with ties to the community from which the data were recorded, raising the thorny issue of defining "community" (Leopold, 2013; O'Meara and Good, 2010; Harris, Holmes and Mertens 2009). As a case in point, some focus group participants suggested that only researchers with verifiable ties to the Deaf community (e.g. issuance from a Deaf family, ability to sign, history of working on projects with Deaf community members) should receive full access to archived sign data, presumably because such individuals are most highly aware of the potential harm that could come to the community if the data are misused. But like any social construct, the boundaries of "the Deaf community" are fluid, depending on who defines them, so determining which researchers possess the requisite community ties will not be straightforward.

The notion of "community" is still relevant, but possibly less problematic, under the second approach to applying graded access, which involves categorizing the data themselves into different levels of accessibility. This is the approach proposed by the CHILDES database (http://childes.psy.cmu.edu, last accessed March 2016), in which each corpus is assigned to one of nine levels, with data at the least restrictive level being fully public and viewable/downloadable without prior registration. At the more restricted levels, researchers may be required to register, submit nondisclosure forms, obtain explicit approval from the original data collectors, or view the data only under direct supervision of someone from the original data collection team (MacWhinney, 2000). For data at any given level, researchers all have the same access, regardless of qualifications. Requirements for data at more restricted levels are made explicit in the database, so prospective researchers can consider them when deciding which corpora to request access to. However graded access is implemented, it would be helpful to establish an Advisory Board to work together with researchers to develop these guidelines. Also, regardless of which type of graded access is instituted in the end, video release forms should also offer the various options for public sharing of specific segments of the data, mentioned earlier.

For the remaining topics in list (2), there was still very little consensus after the focus group discussions: how to deal with "orphaned works;" how to define incidental appearances and whether or not the same re-consent procedures extend to them; whether or not to extend the same re-consent procedures to former research assistants. Opinions on these topics varied widely, some of it probably reflecting age and location. Continued dialogue on these topics is an important step towards developing clear, diversified and actionable protocols, especially since many focus group participants felt that the community has traditionally had very little input on the collection or use of sign language data by researchers. Indeed, the importance of sustained and transparent communication between researchers and research participants can not be overstated, as it is lays the basis for joint efforts across these groups to develop guidelines for video archiving and sharing that are culturally sensitive and balance the benefits of increased access for sign language research with the need to protect individual participant rights.

6. Acknowledgments

Research reported here was supported in part by the National Institute on Deafness and other Communication Disorders of the National Institutes of Health under Award Number R01DC013578. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We also extend our gratitude to our focus group participants and the many children and their families who

have participated in our research in the past, as well as the research assistants who have worked with us.

7. References

Crasborn, O. (2010). What Does" Informed Consent" Mean in the Internet Age?: Publishing Sign Language Corpora as Open Content. *Sign Language Studies*, *10* 2), pp. 276--290.

Crasborn, O. van Zuilen, M., Fikkert, P., van den Bogaerde, B. & Baker, A. (2015). IPROSLA: Data sets on the acquisition of NGT as a first language. Pposter presented at 2nd ICSLA (International Congress on Sign Language Acquisition), Amsterdam.

Harris, R., Holmes, H. M., & Mertens, D. M. (2009). Research ethics in sign language communities. *Sign Language Studies*, *9*(2), pp. 104--131.

Himmelmann, N. (2006). Language documentation: What is it and what is it good for? In J. Gippert, N.P. Himmelmann, & U. Mosel (Eds.), *Essentials of Language Documentation*. New York: Mouton de Gruyter, p 1--30.

Klatter-Folmer, J., Van Hout, R., Van den Heuvel, H., Fikkert, P., Baker, A., De Jong, J., Wijnen, F., Sanders, E. and Trilsbeek, P. (2014). Vulnerability in acquisition, language impairments in Dutch: Creating a VALID data archive. In *Proceedings of LREC 2014: 9th International Conference on Language Resources and Evaluation (http://www.lrec-conf.org/proceedings/lrec2014/index.html)*. pp. 357--364.

Lillo-Martin, D., & Chen Pichler, D. (2008). Development of sign language acquisition corpora. In *Proceedings of LREC 2008: 3rd Workshop on the Representation and Processing of Sign Languages, 6th Language Resources and Evaluation Conference (http://www.lrec-conf.org/proceedings/lrec2008/)*. pp. 129--133.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. 3rd Edition. Vol. 2: The Database*. Mahwah, NJ: Lawrence Erlbaum Associates.

O'Meara, C., & Good, J. (2010). Ethical issues in legacy language resources. *Language & Communication*, *30* (3), pp. 162--170.

8. Language Resource References

Baker, A. (2012) Ethics issues in the archiving of sign language acquisition data. Project deliverable from IPROSLA (Integrating and Publishing Resources of Sign Language Acquisition) Archive (http://www.ru.nl/publish/pages/637809/iprosla_deliverable_d2_v3.pdf).

Digging into Signs:
Emerging Annotation Standards for Sign Language Corpora

Kearsy Cormier[1], Onno Crasborn[2], Richard Bank[2]
[1]Deafness Cognition and Language Research Centre, University College London
49 Gordon Square, London, WC1H 0PD, United Kingdom
[2]Centre for Language Studies, Radboud University Nijmegen
PO Box 9103, NL-6500 HD Nijmegen, The Netherlands
E-mail: k.cormier@ucl.ac.uk; {o.crasborn, r.bank}@let.ru.nl

Abstract

This paper describes the creation of annotation standards for glossing sign language corpora as part of the Digging into Signs project (2014-2015, http://www.ru.nl/sign-lang/projects/digging-signs/). This project was based on the annotation of two major sign language corpora, the BSL Corpus (British Sign Language) and the Corpus NGT (Sign Language of the Netherlands). The focus of the gloss annotations in these data sets was in line with the starting point of most sign language corpora: to make general corpus annotation maximally useful regardless of the particular research focus. Therefore, the joint annotation guidelines that were the output of the project focus on basic annotation of hand activity, aiming to ensure that annotations can be made in a consistent way irrespective of the particular sign language. The annotation standard provides annotators with the means to create consistent annotations for various types of signs that in turn will facilitate cross-linguistic research. At the same time, the standard includes alternative strategies for some types of signs. In this paper we outline the key features of the joint annotation conventions arising from this project, describe the arguments around providing alternative strategies in a standard, as well as discuss reliability measures and improvement to annotation tools.

Keywords: Annotation standards, Glossing, Corpora, Lexical database, Signbank, ELAN annotation software

1. Introduction

The relatively recent advances in computer technology and digital video have made it possible to collect and store large datasets of sign language video recordings. Describing these videos, however, still has to be done manually and is extremely time consuming. Partly due to the fact that sign languages lack a commonly used writing system, annotation of lexical signs involves assigning a unique gloss to each sign: the ID-gloss (Johnston, 2008) As Johnston (2014a) emphasises, there are good arguments for prioritising annotation over transcription. These ID-glosses are stored in a lexical database so that signs in the corpus can consistently be identified. However, this leaves many complexities to deal with in annotation as not all manual signs (or manual articulations more generally) are lexicalized, such as classifier constructions.

Although several sign language corpus projects have provided guidelines for annotation (e.g. Crasborn, Mesch, Waters, Nonhebel, Van der Kooij, Woll, & Bergman, 2007; Crasborn & Zwitserlood, 2008a; Johnston, 2014b; Cormier & Fenlon, 2014; Mesch & Wallin, 2015), there is no general agreement on annotation standards. Recent arguments for standardising sign language corpus annotation have been made by Johnston (2008) and Schembri & Crasborn (2010). More agreement on how to gloss not only lexical but also partly-lexical and non-lexical manual actions will facilitate the access to corpus data by other researchers, and will thus contribute to the empirical study of sign languages in general and to comparative analyses in particular.

The current paper describes some aspects of our proposal for such annotation standards for glossing sign language corpora. They are the results of the Digging into Signs project (2014-2015, http://www.ru.nl/sign-lang/projects/digging-signs/). Our proposal includes a universal standard for some aspects of glosses, while offering alternatives for others. We will therefore also outline some motivations for offering alternatives when needed. The full proposal is published as a PDF document online (Crasborn, Bank & Cormier, 2015).

The focus of the project was in line with the starting point of most sign language corpora: to make general corpus annotation maximally useful regardless of the particular research focus. Therefore the joint annotation guidelines that were the output of the project focus on basic annotation of hand activity, and ensure that annotations can be made in a consistent way for all sign language corpora, providing annotators with the means to create consistent annotations for various types of signs that in turn will facilitate cross-linguistic research.

The aforementioned project not only aimed at setting a standard for the field of sign language studies throughout the world, but also to make significant advances toward two of the world's largest machine-readable datasets for sign languages – specifically the BSL Corpus (British Sign Language, http://bslcorpusproject.org) and the Corpus NGT (Sign Language of the Netherlands, http://www.ru.nl/corpusngt). We start by describing these corpora in section 2, then discussing some aspects of the annotation standard and the related issue of reliability in sections 3 and 4, respectively. Section 5 briefly characterises our efforts thus far to promote the standard. Finally, section 6 describes some related new functionalities in the annotation tool that is used for creating and exploiting these corpora: ELAN. Section 7 provides a brief conclusion.

2. Data Description

We will briefly outline the form and contents of the NGT and BSL corpora, and the previous annotation practices for both datasets.

2.1. The Corpus NGT

The Corpus NGT (Crasborn, Zwitserlood & Ros, 2008) was recorded between 2006 and 2008. It contains 72 hours of dialogues by 92 signers of various age groups from 5 regions in the Netherlands (Crasborn & Zwitserlood, 2008b), and includes both elicited narratives (fables) and free conversation. The great majority of video and annotation files are publicly available at The Language Archive (TLA) of the Max Planck Institute for Psycholinguistics (MPI) in Nijmegen (see https://hdl.handle.net/1839/00-0000-0000-0004-DF8E-6@view). The recent third public release of the annotation files (June 2015) includes tiers for gloss annotations, mouth action annotations, sentence level translations, and a tier for examples referred to in publications. About 20% of the corpus is annotated for the hands, less so for the mouth. The corpus specific annotation guidelines (that can be found on the TLA website as well: https://hdl.handle.net/1839/00-0000-0000-0020-B7CA-4@view) cover all aspects of annotation of the Corpus NGT, not just the publicly available part.

2.2. The BSL Corpus

The BSL Corpus is a collection of around 125 hours of signing by deaf native and near-native BSL signers from 8 regions around the UK (Schembri, Fenlon, Rentelis, & Cormier, 2014; Schembri, Fenlon, Rentelis, Reynolds, & Cormier, 2013). It was published as a partly open-source, partly restricted-access video collection in 2011, and is hosted by UCL CAVA (Human Communication Audio-Visual Archive for UCL). The narrative and lexical elicitation data are open access, while the conversation and interview data are restricted to registered researchers only. Further information about the movies, the annotations and the restrictions can be found on the BSL Corpus web site, http://www.bslcorpusproject.org/cava/. Both CAVA and a version of this corpus for a general audience can be found on the BSL Corpus Data page: http://www.bslcorpusproject.org/data/.

As of 2016, there are around 100 files that have been annotated primarily for manual activity at the lexical level (on right hand and left hand tiers) and that are available on CAVA: 25 each from Birmingham, Bristol, London and Manchester from the conversation data. A substantial part of this annotation work has been carried out for a lexical frequency study (Fenlon, Schembri, Rentelis, Vinson, & Cormier, 2014) with the remainder done as part of a study on directional verbs (Cormier, Fenlon, & Schembri, 2015; Fenlon, Schembri, & Cormier, under review). Additionally, under the Digging into Signs project, an additional 50 files have been annotated at the lexical (ID gloss) level: 25 each from Belfast and Glasgow from the narrative data. Annotation guidelines for manual activity used for all of these files can be found on the BSL Corpus website (http://www.bslcorpusproject.org/cava/).

3. Annotation Standards

Some core features of the gloss annotation guidelines for these two corpora are shared with most researchers in the field: glosses are written words in the standard orthography for a spoken language that uniquely identify a sign form (that is, they function as formal identifiers rather than as translations), they are written in capital letters, and when multiple words are needed for the ID they are separated by hyphens. Moreover, the language of the glosses is trivial in a sense: while it makes most sense to use the spoken language best known to the signers and annotators, it has also been argued that the glosses should match the language of the publication in the case of the citation of examples (Frishberg, Hoiting & Slobin, 2012). As ELAN allows for multilingual controlled vocabularies (Crasborn & Sloetjes, 2014), gloss annotations can be added in one language and displayed in another to other users. The Corpus NGT glosses were created in Dutch, but can also be displayed in their English form.

For most other aspects of glosses, however, a lot of variation can be observed. Our main goal was to develop annotation standards for glosses of signs in sign language corpora, particularly for partly-lexical or non-lexical material. A comprehensive description of the annotation guidelines that were the output of the Digging into Signs project can be found in Crasborn, Bank & Cormier (2015). To summarise, we identified 22 categories (see Table 1) and extensively compared and adapted our (former) annotation practices for both the NGT and BSL corpora (Crasborn, Bank, Zwitserlood, Van der Kooij, De Meijer, & Sáfár, 2015, and Cormier, Fenlon, Gulamani, & Smith, 2015, respectively). This was achieved by several rounds of pilot annotation of small amounts of data from both corpora.

1	Basic gloss	12	Number incorporation
2	Two-handed signs	13	Ordinal numbers
3	Buoys	14	Sign names
4	Lexical variants	15	Fingerspelling
5	Repetition	16	Pointing signs
6	Compounds	17	Classifier/depicting signs
7	Manual negative incorporation	18	Type-like classifier/depicting signs
8	Directional verbs	19	Shape constructions
9	Plurality	20	Gestures
10	Numbers	21	Palm up
11	Number sequences	22	Manual constructed action

Table 1: 22 categories on which agreement was sought

On most of these items we reached agreement, although there are some minor notational issues. For example, as prefixes to indicate certain items, the NGT team prefers to use special characters for brevity, whereas the BSL team prefers abbreviations to provide some memory support, such as in Table 2.

	BSL prefix	NGT prefix
Fingerspelling	FS:	#
Sign name	SN:	*
Gestures	G:	%

Table 2: examples of notational differences

When doing cross-linguistic comparisons, however, these differences are easily overcome by a simple search and replace in the annotation software or a text editor. Similarly trivial notation differences can be found for the glossing of (cardinal and ordinal) numbers and number sequences.

We realised, however, that some aspects of the NGT and BSL corpora are (and will remain) different. These include not only annotation conventions, but also file and tier naming conventions. Also, for specific projects with particular research questions, additional tiers will be needed in order to describe different phenomena related to the manual articulators. As these will contain properties or classifications particular to certain research questions and are more likely to reflect specific theoretical perspectives, standardisation will be more of a challenge. However, it is important to be aware that even general glossing conventions also come with linguistic assumptions (Cormier, Crasborn, & Bank, 2016).

A point in case is the annotation of some types of buoys (meaningful perseveration) vs. meaningless perseveration, and this was a reason to suggest two alternative approaches to glossing instances of non-dominant hand spreading. The notion of buoy as first proposed by Liddell (2003) characterises spreading of the non-dominant hand that fulfils the discourse function of highlighting information. Depending on the sign that is held in its final position, different types of buoys are distinguished: theme, pointer or fragment buoys (list buoys behave differently and are not included here). For both the time alignment and the content of the annotation, different options are proposed. The spreading behaviour of a sign can either be annotated by adding a separate gloss annotation for the hold part of the sign, or the length of the annotation for the sign can be so long as to include both the movement part and the long hold at the end. For the content of the gloss annotation, one can opt for only the gloss of the source sign, include an explicit labelling of that sign as functioning as a buoy, or add a categorisation of the type of buoy. Depending on the amount of linguistic analysis one wants to include in the gloss tier (and thus require from annotators), either a more phonetic or a more functional approach will be attractive. In some cases, it may be possible to use the corpus annotations to test which approach works best (Cormier, Crasborn, & Bank, 2016).

1a:
GlossL MOVE+O BE+O-------------------------------------
GlossR CAT MOVE+2

1b:
GlossL MOVE+O FBUOY FBUOY
GlossR CAT MOVE+2

Figure 1: Example from NGT with non-dominant hand annotated as perseveration (1a) versus as a fragment buoy (1b). MOVE+O and MOVE+2 are depicting constructions which include movement with an O-handshape and 2-handshape respectively. BE+O is a depicting construction with no movement and an O-handshape.

4. Reliability and Validity

Near the end of the project, in order to address an additional aim of testing reliability of these annotation standards, we also conducted a small reliability study of each corpus, with 2 annotators independently annotating a sample of BSL data and 3 annotators independently annotating a sample of NGT data. (Cross-linguistic reliability was not possible because none of the annotators knew both sign languages.) Reliability of the BSL data (around 200 annotations, content of annotations only) was 75% across the 2 annotators. Reliability of the NGT data (around 150 annotations, content of annotations only) was an average of 71% across the 3 pairs of annotators. A content analysis of the present annotation data is taking place at the time of writing. We further plan to develop and apply more detailed measures of reliability in the near future. This will include measurements on alignment of annotations, which was outside the scope of the Digging into Signs project.

5. Improvement of Annotation Software

One of the aims of the project was to improve software for sign language corpus annotation. This project exploited the most widely used multimedia annotation tool in sign language research: ELAN (tla.mpi.nl/tools/tla-tools/elan), developed by the Max Planck Institute for Psycholinguistics (Wittenburg, Brugman, Russel, Klassmann & Sloetjes, 2006). MPI tools are open source software which are well documented and supported. The multilingual user interface of ELAN (like that of other annotation tools) allows access to the software for research assistants with limited knowledge of English, like some of the deaf annotators in the Dutch team. Version 4.9.0 of ELAN was released in May 2015, and included an improvement in the use of External

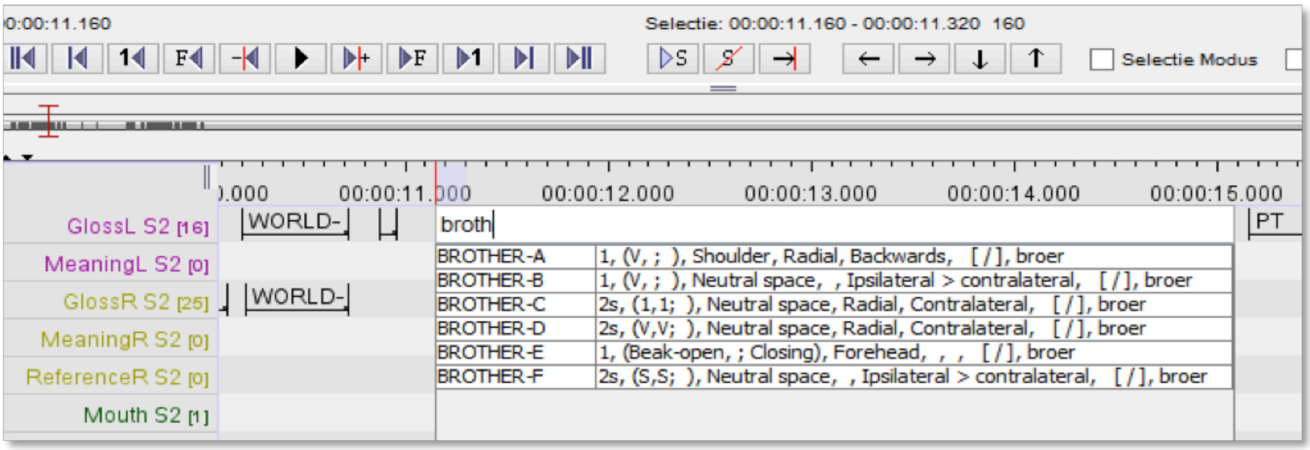


Figure 2: creating a new annotation from an ECV. The user is presented with a drop-down list with possible choices of ID gloss based on the first few letters that are typed in, along with basic phonological information about each to help with identification.

Controlled Vocabularies (ECVs). An ECV provides the annotator with a list of choices (Figure 2) that are based on a lexicon (see below). Working with a lexicon-based ECV eliminates spelling errors, and greatly reduces the number of choices our annotators have to make. The ECV file is no different in format than a regular controlled vocabulary, but because of its size is stored externally, on a server, rather than in each annotation file. The ECV file is an XML file that stores a value and a description for each lexical entry in one or more languages, just like the inline vocabularies in ELAN documents. The list of glosses taken from a lexicon (the values) can thus include extra information (in the description field), which can contain for instance phonological information of the citation form, or information about the semantics.

With release 4.9.0 of ELAN, this description field can be shown at the time of selecting a new gloss. In Figure 1, a screenshot of this drop-down list is shown, with in the second column phonological information on the all glosses in the lexicon that start with 'broth-'. The format is as follows: handedness, (strong hand, weak hand, handshape change), location, absolute orientation: movement, movement direction, movement shape, [number of occurrences / number of signers], keywords/translation equivalents. By displaying phonological information about an ID-gloss at the time of creating a new annotation, annotators can assure themselves that indeed they are selecting the right ID-gloss for the right form, without necessarily having to look up the gloss and video in the lexicon itself every time.

Additionally, a Tier Set function has been created (in beta testing at the time of writing), by which a different selections of tiers can each be assigned a name, after which the user can quickly hide and show groups of tiers in the timeline viewer and other menus. With the large number of tiers that are created for many corpora, it is a challenge to present all and only the desired information at any given time. The Tier Set function allows users to quickly display a specific (pre-defined) set of tiers for a specific purpose, for instance in order to make a quick annotation on a tier that a user is not normally working on. Annotators that are normally focussing on the gloss and mouth tiers can thus quickly show the handshape tiers to annotate a deviant handshape and then hide it again, or quickly hide or show translation tiers depending on the annotator's needs. This results in an uncluttered workspace with easier access to relevant tiers.

The lexicons that form the basis of these lists of glosses are the NGT Signbank (http://signbank.science.ru.nl) and BSL Signbank (http://bslsignbank.ucl.ac.uk/), forks of the original Auslan Signbank (http://www.auslan.org.au/). In future it is possible that ECVs within ELAN could be adapted to work with lexical databases unrelated to these.

6. Conclusion

In summary, the Digging into Signs project provided some much needed improvement to sign language annotation software tools and also brought the field of sign language corpus research one step closer to achieving cross-linguistic annotation standards for sign language data.

However, several challenges remain. Changing existing annotations in a corpus to conform with changed annotation standards is a lot of work, and unfortunately we haven't yet been able to implement all proposed standards into our existing annotations. However, all annotations added to our corpora in current projects make use of the new standards, and older annotations will follow in due time. Also, as annotation standards are implemented and evaluated, it is possible that some changes may be needed, resulting in a need to revisit and change the standards. Open access, sharing and transparency across annotators and projects will help ensure these issues can be addressed and resolved as this field of corpus sign linguistics moves forward.

7. Acknowledgements

This research was possible thanks to the Digging into Data Challenge which for this project included joint funding from Joint Informations Systems Committee (JISC), Arts and Humanities Research Council (AHRC) and Economic and Social Research Council (ESRC) in the UK (grant 3152) and the NWO in the Netherlands (grant 640.006.011). We thank our project teams: Jordan Fenlon, Sannah Gulamani, Sandra Smith in the UK and Merel van Zuilen and Anique Schüller in the Netherlands.

8. References

Cormier, K., Crasborn, O., & Bank, R. (2016). *The role of theory in sign language corpus annotation.* Paper presented at the 12th Conference on Theoretical Issues in Sign Language Research, Melbourne, Australia, , 4-7 January 2016.

Cormier, K. & Fenlon, J. (2014). BSL Corpus Annotation Guidelines, v. 1. Deafness, Cognition and Language Research Centre, University College London. http://www.bslcorpusproject.org/wp-content/uploads/BSLCorpusAnnotationGuidelines_23October2014.pdf.

Cormier, K., Fenlon, J., Gulamani, S., & Smith, S. (2015). BSL Corpus Annotation Conventions, v. 2.1. Deafness, Cognition and Language Research Centre, University College London. http://www.bslcorpusproject.org/wp-content/uploads/BSLCorpus_AnnotationConventions_v2.1_July2015.pdf

Cormier, K., Fenlon, J., & Schembri, A. (2015). Indicating verbs in British Sign Language favour motivated use of space. *Open Linguistics, 1*(1), 684-707. doi:10.1515/opli-2015-0025

Crasborn, O., Bank, R., & Cormier, K. (2015). Digging into Signs: Towards a gloss annotation standard for sign language corpora. Nijmegen: Radboud University & London: University College London. http://www.ru.nl/publish/pages/723853/dis_annotation_guidelines_4may2015.pdf. DOI: 10.13140/RG.2.1.2468.5840.

Crasborn, O., Bank, R., Zwitserlood, I., Van der Kooij, E., De Meijer, A., & Sáfár, A. (2015). Annotation conventions for the Corpus NGT. Version 3, June 2015. Radboud Universiteit Nijmegen. https://hdl.handle.net/1839/00-0000-0000-0020-B7CA-4@view

Crasborn, O., Mesch, J., Waters, D., Nonhebel, A., Van der Kooij, E., Woll, B., & Bergman, B. (2007). Sharing sign language data online: Experiences from the ECHO project. *International Journal of Corpus Linguistics*, *12*(4), 535–562.

Crasborn, Onno, & Slöetjes, Han. (2014). Improving the exploitation of linguistic annotations in ELAN. In N. Calzolari, K. Choukri, & et al. (Eds.), *Language Resources and Evaluation 2014*. Paris: ELRA. http://www.lrec-conf.org/proceedings/lrec2014/pdf/567_Paper.pdf

Crasborn, O., & Zwitserlood, I. (2008a). Annotation of the video data in the Corpus NGT. http://www.ru.nl/publish/pages/527859/corpusngt_annotationconventions.pdf

Crasborn, O., & Zwitserlood, I. (2008b). The Corpus NGT: an online corpus for professionals and laymen. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood, & E. Thoutenhoofd (Eds.), *Construction and exploitation of sign language corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages (LREC)* (pp. 44–49). Paris: ELRA. http://www.lrec-conf.org/proceedings/lrec2008/workshops/W25_Proceedings.pdf

Crasborn, O., Zwitserlood, I., & Ros, J. (2008). Corpus NGT. An open access digital corpus of movies with annotations of Sign Language of the Netherlands (Video corpus). Centre for Language Studies, Radboud University Nijmegen.

Fenlon, J., Schembri, A., & Cormier, K. (under review). Modification of indicating verbs in British Sign Language: A corpus-based study.

Fenlon, J., Schembri, A., Rentelis, R., Vinson, D., & Cormier, K. (2014). Using conversational data to determine lexical frequency in British Sign Language: The influence of text type. *Lingua*, *143*, 187–202. doi:10.1016/j.lingua.2014.02.003

Frishberg, N., Hoiting, N., & Slobin, D. I. (2012). Transcription. In R. Pfau, M. Steinbach, & B. Woll (Eds.), *Sign language: An international handbook* (pp. 1045-1075). Berlin: Mouton de Gruyter.

Johnston, T. (2008). Corpus linguistics and signed languages: No lemmata, no corpus. In O. Crasborn, T. Hanke, E. D. Thoutenhoofd, I. Zwitserlood, & E. Efthimiou (Eds.), *Construction and exploitation of sign language corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages (LREC)* (pp. 82–87). Paris: ELRA. http://www.lrec-conf.org/proceedings/lrec2008/workshops/W25_Proceedings.pdf.

Johnston, T. (2014a). The reluctant oracle: using strategic annotations to add value to, and extract value from, a signed language corpus. *Corpora*, *9*(2), 155–189.

Johnston, T. (2014b). Auslan Corpus Annotation Guidelines. Sydney: Macquarie University. http://media.auslan.org.au/attachments/Johnston_AuslanCorpusAnnotationGuidelines_14June2014.pdf

Liddell, S. K. (2003). *Grammar, gesture and meaning in American Sign Language*. Cambridge: Cambridge University Press.

Mesch, J., & Wallin, L. (2015). Gloss annotations in the Swedish Sign Language Corpus. *International Journal of Corpus Linguistics*, *20*(1), 103–121.

Schembri, A., & Crasborn, O. (2010). Issues in creating annotation standards for sign language description. In P. Dreuw, E. Efthimiou, T. Hanke, T. Johnston, G. Martinéz Ruiz, & A. Schembri (Eds.) *Corpora and Sign Language Technologies. Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages. Language Resources and Evaluation Conference (LREC)* (pp. 212–216). http://www.sign-lang.uni-hamburg.de/lrec2010/lrec_cslt_01.pdf

Schembri, A., Fenlon, J., Rentelis, R., & Cormier, K. (2014). British Sign Language Corpus Project: A corpus of digital video data and annotations of British Sign Language 2008-2014. London: Deafness, Cognition and Language Research Centre, University College London.

Schembri, A., Fenlon, J., Rentelis, R., Reynolds, S., & Cormier, K. (2013). Building the British Sign Language Corpus. *Language Documentation and Conservation*, *7*, 136–154.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk, D. Tapias (Eds.), *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)* (pp. 1556–1559). Genoa. http://www.lrec-conf.org/proceedings/lrec2006/pdf/153_pdf.pdf

Linking Lexical and Corpus Data for Sign Languages: NGT Signbank and the Corpus NGT

Onno Crasborn, Richard Bank, Inge Zwitserlood, Els van der Kooij, Anique Schüller,Ellen Ormel, Ellen Nauta, Merel van Zuilen, Frouke van Winsum, Johan Ros

Radboud University, Centre for Language Studies

PO Box 9103, NL-6500 HD Nijmegen, The Netherlands

E-mail: {o.crasborn, r.bank, i.zwitserlood, e.vanderkooij, a.schuller, e.ormel, e.nauta, m.vanzuilen, f.vanwinsum, j.ros}@let.ru.nl

Abstract

How can lexical resources for sign languages be integrated with corpus annotations? We answer this question by discussing an increasingly frequent scenario for sign language resources, where the lexical data are stored in an online lexical database that may also serve as a sign language dictionary, while the annotation data are offline files in the ELAN Annotation Format (EAF). There is by now broad consensus on the need for ID-glossesin corpus annotation, which in turn requires having at least a list of ID-glosses with a description of the phonological form and meaning of the signs. There is less of a consensus on standards for glossing, on practices of sign lemmatisation, and on the types of information that need to be stored in the lexical database. This paper contributes to the establishment of standards for sign language resources by discussing how two data resources for Sign Language of the Netherlands (NGT) are currently being integrated, using the ELAN annotation software for corpus annotation and an adaptation of the Auslan Signbank software as a lexical database. We discuss some of the present relations between two large NGT data sets, and outline some future developments that are foreseen.

Keywords: sign language resources, multimodal annotation, Signbank, lexical database ELAN annotation software

1. Introduction

How can lexical resources for sign languages be integrated with annotated video corpora? In this paper we aim to answer this question by discussing an increasingly frequent scenario for sign language resources, where the lexical data are stored in an online lexical database, while the annotation data are offline files in the ELAN Annotation Format (EAF).

Lexical databases for sign languages often originated from the purpose of creating sign language dictionaries (Johnston, 2001). These dictionaries were created in a variety of contexts, ranging from language technology or linguistics departments within academia to deaf associations. The varying demands and facilities have led to a diversity of proprietary databases and some open source solutions. A standard even for data structures in this domain is not within view. It is therefore extra important to document the existing solutions, as we do in this paper.

In terms of annotating and retrieving lexical signs for linguistic research, there is by now broad consensus on the need for ID-glosses (Johnston, 2008, 2010) in corpus annotation, which in turn requires having at least a list of ID-glosses with a description of the phonological form and meaning of the signs. There is less of a consensus on standards for glossing (Crasborn, Bank & Cormier, 2015), on practices of sign lemmatisation (Fenlon, Cormier & Schembri, 2015), and on the types of information that are to be stored in the lexical database.

This paper contributes to the establishment of standards for sign language resources by discussing how two data resources for Sign Language of the Netherlands (Nederlandse Gebarentaal; NGT) are currently being integrated, using the ELAN annotation software for corpus annotation (Wittenburg et al., 2006) and an adaptation of the Auslan Signbank[1]software (Johnston, 2001, 2010) as a lexical database.

2. Two Existing Data Sets

This section describes first the Corpus NGT and then NGT Signbank. While not the only option, as we will discuss in the conclusion, this type of combination of data sets is getting more common in the domain of sign language resources.

2.1 Corpus NGT

The Corpus NGT (Crasborn & Zwitserlood, 2008; Crasborn, Zwitserlood, & Ros, 2008) is a collection of video and annotation data of 92 prelingually deaf signers, recorded in dyads, who retell video clips and picture stories and discuss issues related to deafness, deaf education and sign language.Annotation of the corpus is on-going; the latest (third) public release of Corpus NGT annotationsthat was published in June 2015 (Crasborn et al., 2015) contains over 145,000 glosses for the left and right hands. At present the production version of the corpus contains almost 370,000 annotations for different levels of transcription and analysis, from sentence-level translations to degree of thumb extension.The latest public release of Corpus NGT annotations can be viewed in and downloaded from The Language Archive[2], and as a single compressed file from the corpus website.[3] The production version of the corpus is stored on a private SVN server. This allows working on offline copies of EAFs, with locally stored video files.

[1]http://www.auslan.org.au

[2]http://hdl.handle.net/1839/00-0000-0000-0004-DF8E-6@view

[3]http://www.ru.nl/corpusngtuk/methodology/annotation

2.2 NGT Signbank

Earlier projects that aimed to link NGT corpus data with a lexical database (e.g. Crasborn, Hulsbosch & Sloetjes, 2012) involved the LEXUS lexical database. However, as the further development and support of LEXUS was discontinued in 2015, our pilot efforts in this direction had to be aborted. As an alternative, we chose to adopt the Signbank lexical database software. Signbank has originally been developed for Australian Sign Language[4] (Auslan; Johnston, 2001, 2010), and has since also been implemented for British Sign Language[5] (BSL; Fenlon et al., 2015), NGT[6] (Crasborn et al., 2014).

Versions of the Signbank software are currently also implemented for Finnish Sign Language[7] and American Sign Language (ASL). Given the fact that the sign language corpora developed in these countries are all annotated in ELAN, the discussion in this paper also pertains to the development of those Signbanks.

While the NGT Signbank software is open source and its development can be tracked online,[8] the NGT data set will not be published under an open access license for at least another two years. In this period, intensive annotation of the Corpus NGT leads to weekly additions to NGT Signbank. At the time of writing, the phonological description of the existing 3,200 entries is still being double-checked, and the recording of citation form videos is almost completed.

3. Existing Relations between Data Sets

This section describes two types of relationships between corpus and lexical database that are already implemented, while section four will focus on some further interactions between the data sets that will make exploitation of the data richer and easier. All interactions are visualised in Figure 1.

3.1 The Lexical Database as a Vocabulary of Gloss Types for Annotation

Tofacilitate video annotation in ELAN, an externalcontrolled vocabulary (ECV; Crasborn, Hulsbosch & Sloetjes, 2012) is used. AnECV contains the full list of ID-glosses in Signbank, to label lexical signs with, as well as phonological information about those signs (e.g. handshapes, location, movement direction) and Dutch translation equivalents that serve to clarify their meaning. When deciding on an annotation, the annotator choosesan entry from the ECV to be included in the EAF file. This facilitates decision-making and reduces the occurrence of typing errors. The ECV is centrally stored on a web server (hence the E for external), allowing for central updating of the ECV with changed or added glosses, phonological information and meaning. The ECV is automatically reloaded each time an EAF file is opened on a local computer with an internet connection. Annotation values of the glosses are then updated, if applicable, to reflect the current information in the ECV. This ensuresthat the annotators always work with the latest version.

As described in Crasborn & Sloetjes (2014), starting from version 4.7.0, ELAN allows for the creation of multilingual ECVs. This means that each entry in an ECV can have multiple values, one per language. Annotations based on an ECV entry will display one of these values, depending on the language selected by the user in ELAN.

A bilingual ECV is generated directly from Signbank. A nightly server-side script generates an updated ECV, including all changes made to the Signbank databasein the previous day. Users can also choose to manually update the ECV, so that any changes to the database during the annotation process immediately become available to themselves and to other annotators. As was already explained above, such changes to the ECV will not be immediately visible in open EAF files: these have to be closed and reopened.

As the production version of the Corpus NGT is stored on an SVN server, all annotations are made in local copies of the corpus, which are then committed to the central server. Consequently, any changes in gloss values in Signbank (and thus, nightly or manually, in the ECV) are not propagated to the whole corpus, but only in locally opened, saved and committed files. In order to update all the gloss values (and hence provide accurate search results when doing a corpus wide search), a nightly script (currently in beta testing) changes gloss values in the corpus on the basis of their ECV-links, and also checks whether all glosses that lack an ECV-link but whose value occurs in the ECV (and thus in Signbank) receive the proper ECV-link.

The relation between Signbank and EAF files through the ECV is currently one-way: items are added to Signbank and then displayed in ELAN. It is not yet possible to harvest new items in ELAN files and add these to Signbank, for instance, or to manually add a new item to Signbank from within the ELAN interface. Both of these options will be explored in the near future. Ideally, the ECV file would only serve as a cache for offline work, while whenever a user is online, there is a live link between Signbank and the gloss tiers in ELAN.

Currently, ID-gloss fields exist in NGT Signbank and in the ECV for Dutch and English, while the whole system is being internationalised. Both the interface and most of the data have been translated into Dutch, facilitating all team members who work with the software. Further translations can be easily made in the future. A Mandarin Chinese translation is underway.

Translation equivalents are presently only added in Dutch, and while it is in principle possible to add translation equivalents in English or another language, optimal quality would require that translation equivalents are added on the basis of the meaning of the sign, rather than by translating the Dutch translation equivalents one

[4] https://github.com/Signbank/Auslan-signbank
[5] http://bslsignbank.ucl.ac.uk
[6] http://signbank.science.ru.nl
[7] https://github.com/Signbank/FinSL-signbank
[8] See the Github repository at https://github.com/Signbank/NGT-signbank

by one. This requires high-level knowledge of both NGT and English, which not every team member has.

NGT Signbank currently contains more than 3,200 ID-glosses. While annotating, one often encounters signs that do not yet have an ID-gloss. Several criteria are used for the introduction of new glosses in Signbank (see also Cormier et al., 2015). Basically, the phonological form of the sign has to be dissimilar from all existing forms in Signbank. However, also in case the form is similar to an existing form and the meaning of the two signs does not overlap, the sign is assigned a new ID-gloss.

of signs obtained in other ways. The first NGT dictionaries that were developed in the 1980s in the Netherlands had the explicit aim of evaluating regional variation (Schermer, Harder & Bos, 1988; KOMVA 1989). Information on the regional use came from elicitation studies (Schermer, 2012). The paper dictionaries that were published in the late 1980s and early 1990s all present information about the regions in which signs are used. While the Corpus NGT data form a broader empirical base for regional use (cf. the ideal scenario described by Schermer, Stroombergen &

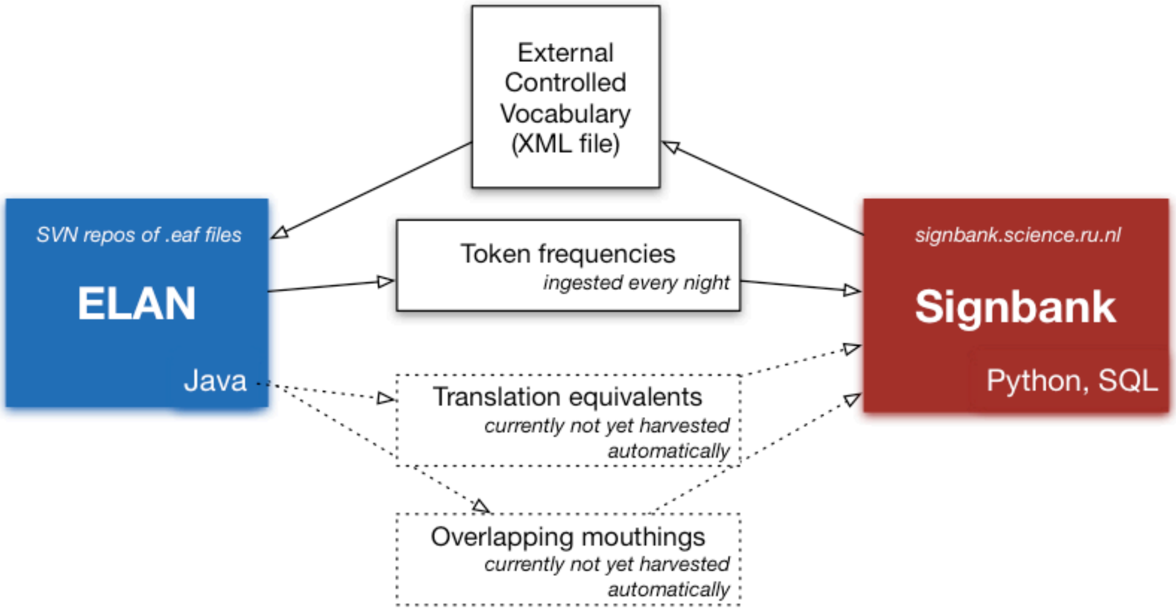


Figure 1: Overview of existing and foreseen relations between ELAN and Signbank.

3.2 Token Frequencies in the Lexicon

Two types of frequency data are automatically ingested in Signbank from the glosses in the Corpus NGT. First of all, there are token frequencies over the whole corpus and for each of the six regions distinguished in the metadata. These are the five traditional dialect regions in the Netherlands (Schermer, 2004), plus a rest category that includes signers with a mixed regional profile. Second, the number of signers that produce tokens of a sign is also calculated and ingested in Signbank, for the whole corpus and per region. This second type of information is particularly useful in determining how widespread the use of a sign is within a region: is it an idiosyncratic (perhaps older) form used by a single signer, or are there several people using the same sign? Together with the distinction in regions, this may help in selecting signs for inclusion in dictionaries, or even for research purposes: phonologists, for example, may not want to base their analysis on the phonology of forms that are only used by one or two signers, however frequently they may use these forms.

Token frequencies from the Corpus NGT form further empirical support for information on regional use Tervoort, 1984), they do face a similar limitation as the printed dictionaries of thirty years ago, in reflecting dialectal variation only at a specific moment in time. For the Corpus NGT, there was a short time window of the data recording between 2006 and 2008. Enriching token statistics by adding data from other annotated sources in the future, whether newly recorded or historic, can potentially provide a rich view on dialectal variation over time.

4. Data Set Interactions under Development

We presently foresee three types of data interactions that could be implemented fairly easily, and that would enrich Signbank on the basis of corpora.

4.1 Harvesting of Translation Equivalents from the Corpus

The meaning of signs in sign language dictionaries and lexical databases is typically represented in terms of a spoken language, by including translation equivalents and sometimes also translated sentences illustrating typical use of signs. NGT Signbank lists translation equivalents in Dutch, as was already mentioned in section 3.1. At present, these translation equivalents are added based on the knowledge of annotators and researchers. The latter will often overlap with the meaning of actual uses of those signs in the corpus, but mismatches in both directions are observed: Signbank also lists translation equivalents that are not observed in the corpus, and not all possible translations of signs in the corpus are (yet) present in Signbank. These translations can be specified with each ID-gloss on a separate tier named Meaning in the corpus. While this is not done systematically for every gloss at this moment, annotators are requested to specify the meaning of a sign in each case where it does not overlap with the meaning encoded in the ID-gloss. This workflow has been adopted more than a year ago, and there remains a large set of older files (with an estimated 100,000 glosses) for which meanings have only occasionally been specified.

By harvesting the meaning annotations that are specified for ID-glosses, translation equivalents can be generated in Signbank in a corpus-based way. By including frequencies of different translation equivalents, the quality of semantic information in Signbank can be enriched. At the same time, the corpus or corpora on which sign language lexical resources are based are restricted, due to the immense time effort that the manual annotation process takes up. It is therefore desirable to store both translation equivalents based on signer intuitions and corpus-based translation equivalents.

4.2 Harvesting of Mouthings from the Corpus

The ubiquitous use of mouthings and their presumed role in the interpretation of NGT (Bank, 2015) calls for its systematic annotation in sign language corpora. Crasborn & Bank (2014) propose an annotation scheme for its annotation. The study of mouth actions in relation to signs continues to raise many questions. The mechanisms behind the variation found in the use of either mouthings or mouth gestures with signs, for instance, is not yet fully understood. Inclusion of corpus-based information on mouthings in the lexical database can help us to better understand the relation between manual signs and mouthings, and as for translation equivalents, frequency information on mouthings can aid in the determination of the semantics of signs.

One of the biggest challenges in the automated harvesting of mouthings, however, is temporal alignment. Mouthings do not necessarily align with the signs they accompany: they can spread over adjacent signs, or a sign can co-occur with multiple mouthings, and all the variations inbetween. This is illustrated in Figure 2. Even when a stretch of connected signing co-occurs one-on-one with corresponding mouthings, annotation alignment is necessarily noisy, due to the complexity of the phonetic signal.

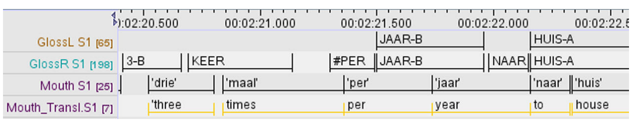


Figure 2: Partial alignment (JAAR with *jaar* 'year') and non-alignment (HUIS with *naar* 'to') of manual signs and mouthings.

The solution we aim for is to list for each sign all mouthings that co-occur with that sign, including those that only partly overlap. In addition, two distinct values may be calculated and stored in relation to overlapping mouthings. First of all and most importantly, for each mouthing type, it should be calculated how often it occurs with a sign, just as for the translation equivalents discussed in the previous section. Second, the average amount of overlap of a mouthing type with a sign could be computed. This provides a way of determining whether a mouthing should be seen as co-occurring with a sign (in case of large overlap) or as coincidental (in case of small overlap, see e.g. in Figure 2 the overlap of #PER with the mouthing '*maal*' that co-occurred with the previous sign but didn't really spread). Consequently, only mouthings with large overlap ratios (at least 60% overlap) should be included in the lexical database. The two numbers – frequency and overlap ratio – together provide a clear and concise measure of co-occurrence with sign types.

4.3 Use of Corpus Examples in the Lexical Database

A third possibility for enriching a corpus-based lexical database like Signbank would be to include information on the use of signs in their context. This can help in providing a richer view of the lexical semantics and pragmatics of signs, as well as form a solid basis for a learner dictionary in the long term (but see Hunston, 2009, on some of the complexities involved in presenting corpus data to learners).

Presenting information on signs in their context could take the form of collocations, highlighting the most frequent left and right collocates of signs. However, given that a lot of information in signed utterances is realised simultaneously, by means of features of facial expression and two-handed constructions to name but two prominent aspects, sequential collocates are less informative for sign languages than for many spoken languages (Crasborn et al. 2013). It may therefore be more informative to aim to include full sentences in a way that maximises the use of basic level annotations like the glosses of the two hands, mouth actions, and sentence level translations. A presentation like in the alignment view in ELAN (Crasborn & Sloetjes, 2014) could be ideal for this, yet may be difficult to implement. A basic list of sentence translations as used in some sign language dictionaries could be a simple start.

Both for scientific uses and for dictionary users, one may wonder, however, whether a full list of all occurrences is really desirable. Perhaps it would be betterto manually or automatically select typical examples. This is clearly an area that needs further exploration.

5. Where to Access the Corpus Metadata?

The metadata for the Corpus NGT have been described in the IMDI standard when the Corpus NGT was first archived in 2008. IMDI files in The Language Archive can be accessed by online searches in the archive; this makes it possible to first select a set of constraints (age, gender etc.) before searching the content of the EAFs.

To date, however, there is no straightforward way to refer to the available metadata while searching in ELAN. For instance, restricting searches to certain regions or age groups is not possible. Some metadata are stored in ELAN files 'natively', as part of the XML specification of EAF files: the participant property and the annotator property of tiers. This enables for instance searching for sign X overlapping with mouthing Y by participant Z. However, more options are desired for many types of searches that are now impossible.

One way to tackle this problem would be to include metadata in the EAF itself, on separate tiers. This is what we plan to do as long as search options in ELAN cannot refer to information in metadata files. We aim to include one extra tier per signer, containing metadata information like age or age group, gender, region and type of video content (e.g., fable, free discussion etc.). The tiers will contain a single annotation that covers the entire length of the EAF, to allow for searches on overlapping annotations (e.g. all tokens of AMSTERDAM by signers from Amsterdam). Each annotation will need to contain a string of metadata values that are preceded by a string, letter, or number identifying the metadata field.

6. Conclusion

In this paper we discussed several links between Corpus NGT annotations made in ELAN and the lexical database NGT Signbank. While the implementation of the links brings along some software development particular to the design of the two tools, the nature of the information is of a more general nature and has clear linguistic motivations. Information on lexical items stored in a lexical database is needed for a proper use of ID-glosses in the annotation of manual signs in sign language corpora. The frequency data, semantics and contextual information from corpora all form important additions to a lexical database. They can ultimately lead to corpus-based dictionaries (see also Hanke, 2006 for discussion).

The scenario we describe here is of course not the only one currently in use – but there are not too many alternatives. Hanke (2002), Konrad & Langer (2008),and Hanke & Storz (2009) describe the integrated iLex environment, where type and token data as well as metadata are integrated in a single database. This solution has also been adopted in Poland and Denmark, among other countries. One advantage of the use of ELAN over iLex is independent of the interaction with Signbank, namely the possibility of offline use of a corpus. This advantage is of course becoming less prominent as ELAN documents interface more directly with the online database Signbank. Together with the scenario described in this paper, these two seem to be the only solutions world-wide that have a substantial number of users, both in terms of the sign languages covered and the number of research groups working with them. For those corpus research groups that choose or chose ELAN as their annotation tool, the ECV/Signbank scheme described here provides a great enhancement in workflow control.

7. Acknowledgments

We thank Trevor Johnston and Steve Cassidy for letting us use the Auslan Signbank software as the basis for the NGT Signbank, and Wessel Stoop for his work in implementing NGT Signbank. We further acknowledge NWO grants 640.006.011 (Digging into Signs) and 360.70.500 (Form-Meaning Units).

8. References

Bank, R. (2015). *The ubiquity of mouthings in NGT. A corpus study*. Utrecht: LOT.

Crasborn, O., & Bank, R. (2014). An annotation scheme for the linguistic study of mouth actions in sign languages. In O. Crasborn, E. Efthimiou, S.-E. Fotinea, T. Hanke, J. Hochgesang, J.H. Kristoffersen, & J. Mesch (Eds.), *Beyond the manual channel. 6th Workshop on the Representation and Processing of Sign Languages* (pp. 23-28). Paris: ELRA.

Crasborn, O., Bank, R., & Cormier, K. (2015). Digging into Signs: Towards a gloss annotation standard for sign language corpora. Nijmegen: Radboud University& London: University College London. http://www.ru.nl/publish/pages/723853/dis_annotation_guidelines_4may2015.pdf

Crasborn, O., Bank, R., Zwitserlood, I., Van der Kooij, E., Ormel, E., Ros, J., … Vonk, M. (2014). NGT Signbank. Nijmegen: Radboud University, Centre for Language Studies.

Crasborn, O., Hulsbosch, M., & Sloetjes, H. (2012). Linking Corpus NGT annotations to a lexical database using open source tools ELAN and LEXUS. In O. Crasborn, E. Efthimiou, E. Fotinea, T. Hanke, J. Kristoffersen, & J. Mesch (Eds.), *Interactions between Corpus and Lexicon. Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages. Language Resources and Evaluation Conference (LREC)* (pp. 19–22).

Crasborn, Onno, Hulsbosch, Micha, Lampen, Lari, & Sloetjes, Han. (2013). *New Multilayer Concordance Functions in ELAN and TROVA*. Paper presented at TiGeR 2013: Tilburg Gesture Research Meeting, Tilburg, the Netherlands.

Crasborn, O., & Sloetjes, H. (2014). Improving the exploitation of linguistic annotations in ELAN. In ELRA (Ed.), *LREC 2014: 9th International Conference on Language Resources and Evaluation, Reykjavik, Iceland* (pp. 3604–3608). Paris: ELRA.

Crasborn, O., & Zwitserlood, I. (2008). Annotation of the video data in the Corpus NGT. Nijmegen:

Radboud University, Centre for Language Studies.http://www.ru.nl/publish/pages/527859/corpusngt_annotationconventions.pdf
Crasborn, O., Zwitserlood, I., & Ros, J. (2008). Corpus NGT. An open access digital corpus of movies with annotations of Sign Language of the Netherlands (Video corpus). Nijmegen: Radboud University, Centre for Language Studies
Crasborn, O., Zwitserlood, I., Van der Kooij, E., Sáfár, A., Ros, J., Nauta, E., … De Meijer, A. (2015). Corpus NGT gloss annotations, 3rd release. Nijmegen: Radboud University, Centre for Language Studies. http://www.ru.nl/corpusngtuk/methodology/annotation
Fenlon, J., Cormier, K., & Schembri, A. (2015). Building BSL Signbank□: The lemma dilemma revisited. *International Journal of Lexicography*, *28*(2), 169-–206.
Hanke, T. (2002). *iLex - A tool for sign language lexicography and corpus analysis.* Paper presented at the LREC 2002 conference, Las Palmas de Gran Canaria, Spain.
Hanke, T. (2006). Towards a Corpus-based Approach to Sign Language Dictionaries. In *Lexicographic matters and didactic scenarios: Second Workshop on the Representation and Processing of Sign Languages. (LREC2006)* (pp. 70–73).
Hanke, T., & Storz, J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In O. Crasborn, T. Hanke, E. D. Thoutenhoofd, I. Zwitserlood, & E. Efthimiou (Eds.), *Construction and exploitation of sign language corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages (LREC)* (pp. 64–67). Paris: ELRA.
Hunston, S. (2009). The usefulness of corpus-based descriptions of English for learners. The case of relative frequency. In Karin Aijmer (Ed.), *Corpora and language teaching* (Vol. 33, pp. 141-154). Amsterdam/Philadelphia: John Benjamin Publishing Company.
Johnston, T. (2001). The lexical database of Auslan (Australian Sign Language). *Sign Language & Linguistics*, *4*(1), 145–169.
Johnston, T. (2008). Corpus linguistics and signed languages: No lemmata, no corpus. In O. Crasborn, T. Hanke, E. D. Thoutenhoofd, I. Zwitserlood, & E. Efthimiou (Eds.), *Construction and exploitation of sign language corpora. Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages (LREC)* (pp. 82–87). Paris: ELRA.
Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, *15*(1), 106–131.
KOMVA. (1989). Basis-gebarenschat: een beknopt gebarenwoordenboek [Basic sign language vocabulary: a concise sign dictionary]. Amsterdam: NSDSK.
Konrad, R., & Langer, G. (2009). Synergies between transcription and lexical database building: The case of German Sign Language (DGS). In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.), *Proceedings of the Corpus Linguistics Conference, CL2009*. Liverpool, UK.
Schermer, T. (2004). Lexical variation in Sign Language of the Netherlands. In M. v. Herreweghe & M. Vermeerbergen (Eds.), *On the lexicon and beyond* (pp. 91-110). Washington D.C.: Gallaudet University Press.
Schermer, T. (2012). Sign Language Planning in the Netherlands between 1980 and 2010. *Sign Language Studies, 12*(4), 467-493.
Schermer, T., Stroombergen, M., & Tervoort, B. (1984). How to make a lexicon of sign? In F. Loncke, P. Boyes Braem, & Y. Lebrun (Eds.), *Recent research on European sign languages* (pp. 105-112). Lisse: Zwets & Zeitlinger.
Schermer, T., Harder, R., & Bos, H. (1988). *Handen uit de Mouwen: Gebaren uit de Nederlandse Gebarentaal in kaart gebracht* [Hands at work: Mapping signs from Sign Language of the Netherlands]. Amsterdam: NSDSK/Dovenraad.
Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)* (pp. 1556–1559).

Recognition of Sign Language Hand Shape Primitives With Leap Motion

Burçak Demircioğlu, Güllü Bülbül, Hatice Köse
Computer Engineering Department, Istanbul Technical University
Istanbul, Turkey
{demircioglubu, bulbulg, hatice.kose}@itu.edu.tr

Abstract

In this study, a rule based heuristic method is proposed to recognize the primitive hand shapes of Turkish Sign Language (TID) which are sensed by a *Leap Motion* device. The hand shape data set was also tested with selected machine learning method (*Random Forest*), and the results of two approaches were compared. The proposed system required less data than the machine learning method, and its success rate was higher.

Keywords: Sign Language, Sign Recognition, Leap Motion

1. Introduction

Sign recognition is vital for an efficient and robust human-machine interaction for hearing impaired people. To be able to use such a system in as many contexts as possible, the technology should be as small and as adaptable as possible. To achieve this goal, the hardware chosen for this study is the *Leap Motion* sensor[1], which is smaller than similar sensors, and can be integrated to any device with USB port, such as robots, computers or smart phones. This study is a part of an ongoing project on the recognition and generation of Turkish Sign Language (TID) gestural vocabulary by computer aided methods. For this purpose, depth sensors such as *Kinect* and *Leap Motion* are used to recognize a selected corpus from TID and this corpus is also generated via humanoid robots and virtual avatars, therefore a two way nonverbal communication on TID is achieved. Due to the modular structure of the project, it can be extended to any similarly structured sign language with minimum effort. The main purpose is to use this multi-modal platform to teach children sign language using interactive games (Köse et al., 2014), (Köse et al., 2015a), (Köse et al., 2015b).

This project consists of two phases: A real-time solution, and an offline solution for hand sign recognition. The real-time solution is based on the heuristic models we developed for every sign on *Leap Motion* based system after analyzing every hand sign for Turkish Sign Language. The data gathered to test this system is then used to train and test selected offline machine learning technique via the *Weka*[2] system to verify the success of the system. The second stage is offline. After all these stages, the same data is given to *Weka*, and offline working code, then the two results of these processes are compared.

There are some previous studies on sign recognition for Turkish Sign Language as well as for other sign languages. However, they differ from this project in their approaches and the technology used. For example, a system aiming at improving learning speed of children learning Turkish Sign language which uses machine learning techniques is presented in (Haberdar and Albayrak, 2005). Keskin et al. (2013) studied recognition of real-time data from selected TID signs (digits) by using *Kinect* technology. The key advantage of our approach is that it requires fewer sampling and therefore less preparation time. So, algorithm coding is chosen for this project which makes the project more efficient than others. To the best of our knowledge, this is the first study using a *Leap Motion* sensor for hand shape inventory recognition in Turkish Sign Language. This alphabet consists of 32 static hand shapes which are used to produce all signs in Turkish Sign Language (Kubuş, 2008), and different from the finger alphabet (for finger spelling). Finger alphabet greatly differs from this alphabet structurally, therefore the recognition systems for finger alphabet can not be employed for this study.

2. Theoretical Information

2.1. Sample Sign Data Set

The main purpose of the project is the recognition of the static hand shape inventory (not the alphabet) of the Turkish Sign Language; because static hand shapes are distinctively and categorically highest phonological features of sign languages as stated in (Kubuş, 2008). The static hand shape inventory of TID which is used in this paper is based on the same work which is the main linguistic source on this language feature. This reference work indicates that TID is a rich language in terms of morphology, phonology and classification. TID has significantly different linguistic properties than other sign languages, as in the case of spoken languages. Although several handshapes in TID are similar to other sign languages, most of them are different. For example; some ASL (8,E,K,M,N,T) and Taiwanese Sign Language (middle finger, ring finger) handshapes are absent in Turkish Sign Language. Therefore Turkish native signers can not differentiate some similar signs of ASL which indicates that TID has a unique handshape inventory (Kubuş, 2008).

Some of the handshapes in the sample data set are not included in this study as the *Leap Motion* sensor is inadequate to sense overlapping fingers. Also, the sensor is not precise enough to differentiate similar signs such as *ASL C-handshape* and *Hooked Flat Extended*. 18 signs from the sample data set are included in this project: 5 Handshape, V&2 Handshape, L Handshape, O Handshape, C

[1]URL: https://developer.leapmotion.com/. Last access date:16/01/16.

[2]URL: http://www.cs.waikato.ac.nz/ml/weka/index.html. Last access date:27/03/16.

Handshape, ASL A Bar, ASL 8 Handshape, ASL I&H Handshape, ASL I Handshape, ASL Y Handshape, ASL 3 Handshape, 4 Claw Handshape, 8 Handshape, 9 Handshape, ASL A Handshape, Baby O Handshape, Open 8 Handshape, and ASL Q Handshape (figure 1).

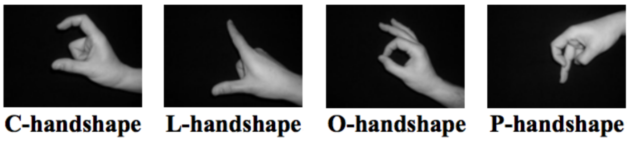


Figure 1: Sample data set from (Kubuş, 2008)

2.2. Related Works

There are several studies on the recognition of Sign Languages using depth sensors such as *Leap Motion* and *Kinect* which are summarized as follows:
The first study (Chuan et al., 2014) is based on the recognition of the 26 static single-hand finger alphabet signs of the ASL using *Leap Motion*. This study uses two machine learning methods which are *k-nearest Neighbor(NN)* and *Support Vector Machine(SVM)* based on the samples from human subjects with *Leap Motion*. The average success rate of this study is measured as k-NN=72.78% for four-fold cross validation with k=7 and as SVM=79.83% for four-fold cross validation using Gaussian radial basis function (RBF) kernel.
The second study (Mohandes et al., 2014) presents a project for the recognition of the 28 static single-hand finger alphabet signs in Arabic Sign Language using *Leap Motion*. Two machine learning methods, *Naïve Based Classifier(NBC)* and *Multilayer Perceptron (MLP)*, are used for the hand shapes classification. The success rate of these methods are NBC=98.3% with 76/2800 mis-classified samples and MLP=99.1% with 26/2800 mis-classified samples.
The last project (Keskin et al., 2013) is based on the hand gesture recognition using *Kinect*. Recognition of the ASL signs representing the 10 digits is presented in the project. The average success rates of real-time test achieved by the cross-validation tests is ANN (Artificial Neural Networks)=98.81% and SVM(Support Vector Machine)=99.90%. The average success rates for a synthetic data set test achieved by the cross-validation tests is ANN=99.89% and SVM=99.96%.

3. Analysis And Modeling

Leap Motion has two infrared sensors which are both directed towards the y-axis from the *Leap Motion* (axes of leap motion can be seen in Figure 2). These sensors detect the world as a funnel which extends as sensing reaches outward from the leap motion by maintaining end-to-end 150 degrees gap in three dimensions. The height of the sensing extends from 25 to 600 millimeters (Leap Motion, 2016). Because the sensing is done by the infrared sensors, when a finger is in front of another finger in the y axis, the finger positioned behind cannot be seen by the sensor and this situation lowers the precision of the sensing process. Also for precise sensing, the view must be clear in the sensing funnel and the light must be controlled to keep the high contrast stable, which is vital for infrared sensing.
139 different attributes, which are received from the sensor at a speed of approximately 127fps, are used to recognize the signs. By using these attributes, necessary angles and the distances between joints and bones (when angles are not sufficient enough) are calculated and the signs' limit values are defined. This solution is explained in details in the section 4.2.
In this study, the distances between the data points are not preferably used; because the distances may vary from person to person and from hand to hand. For example, distance between palm center to a finger tip may be considerably smaller for a child's hand compared to an adult's hand. For this reason, the angles rather than the distances are used for the calculations (when angles are sufficient enough).

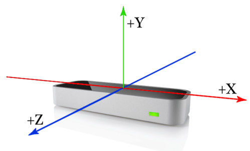


Figure 2: Axes of Leap Motion (Leap Motion, 2016)

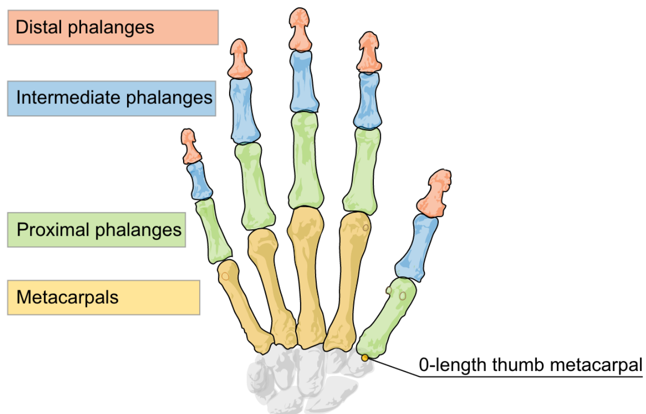


Figure 3: Bone types in hand (Villarreal, 2007)

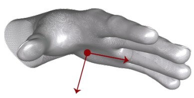


Figure 4: Palm Direction and Normal Vectors (Leap Motion, 2016)

4. Design, Implementation And Test

This project consists of two phases: sign language recognition with real-time data and offline recognition:

4.1. Real Time Recognition

The angles and distances between the defined attributes are calculated to be used in the recognition phase. The

distances are in use when the angles are not sufficient for recognizing the chosen sign. For example, to recognize the ”L” sign; the angle between the thumb’s tip direction and index finger’s tip direction is calculated, which is sufficient to recognize this sign. Here the directions are vectors in three dimensions. The sample pseudo code of the heuristic model of a sign is presented in Figure 5, all of the recognition rules can be seen in the Appendix section.
As a final step of the recognition, a stabilization phase is required for the output to overcome the noise in the real-time data. A *Sliding Window* based approach is used to discard the noise in the data. The window size is chosen as five frames for the method, and the most frequent data item in the window is selected as the output of the window, at each iteration.
In the test phase, real-time data from both hands of the three human subjects (two women and one man) are used as in the Figure 6 and Figure 7.

```
For 8-handshape:
    IF Extended Finger Count is 0
        IF Palm Direction is "Down"
        AND Distance Btw Ring and Pinky Finger Tip Positions<DISERRORMARGIN
        AND Distance Btw Thumb and Ring Finger Tip Positions<DISERRORMARGIN
        AND Distance Btw Thumb and Middle Finger Tip Positions>DISERRORMARGIN
        AND Angle Btw Index Finger's Proximal Bone's Direction and Hand's Direction Vectors > PI / 2
        AND Angle Btw Index Finger's Proximal and Intermediate Bones' Direction Vectors > PI / 2 - ERRORMARGIN
        AND Angle Btw Middle Finger's Proximal and Intermediate Bones' Direction Vectors > PI / 2 - ERRORMARGIN
            Sign is "8-handshape"
```

Figure 5: A sample from specified rules for a sign’s recognition

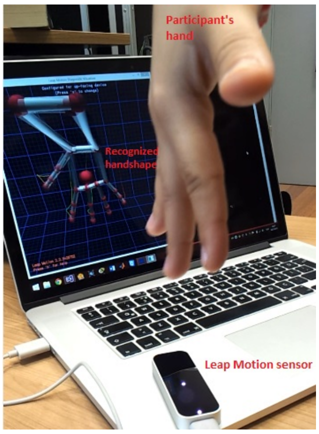


Figure 6: A Screenshot from the experiment 1

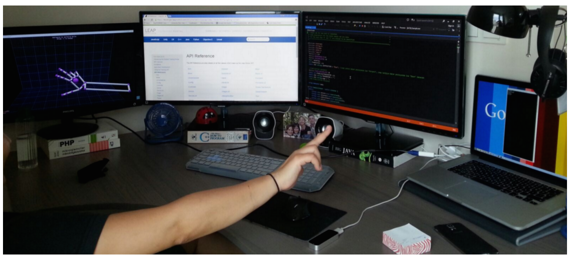


Figure 7: A Screenshot from the experiment 1

4.2. Offline Recognition

For the offline recognition phase, the data obtained from the real time recognition phase is recorded into the text files in the *arff* format. 30 frames are recorded for every trial. For every sign, 10 trials are recorded. The first five of these trials are used as training data and the remaining five are used as test data, and saved as separate *arff* files. In both files there are

$$30 * \frac{10}{2} = 150 \quad (1)$$

frames for every sign.
For the offline phase, the *Leap Motion API* cannot be used; due to the permission restrictions in the API, the objects’ data is not accessible, therefore it cannot be fed with the offline data. To overcome this problem, the necessary objects are listed, modeled, and created again within header files to be used in the offline phase’s code. The new objects are created with the same names as in the API to make the conversion simpler and error free. After the creation of new object structure, all of the data is read from the previously created *arff* files frame by frame and saved into the object variables in the code.
To adapt the code to the new object structure, some variables, and functions are modified, as well. In the *Leap Motion* API, the output data are obtained from the relevant functions defined in the objects. For example, direction vectors are obtained from the *direction()* function from objects such as *finger*, and the extended finger list is obtained from the hand object with fingers and extended functions such as *hand.fingers().extended()*. With the new object structure which is created for the offline use, all of the variables are directly accessible. Therefore the equivalents of these function calls in the offline code are; *finger.direction* and *hand.extendedFingerList*. In the same way, if just the x direction of direction vector is required, it can be obtained directly by the variable *finger.direction.x*.
Finally, a sample machine learning method (*Random Forest*) is implemeted via the *Weka* system and tested using the offline data. First of all, the method is trained using the *dataTrain.arff* file by 10-fold cross validation method. The trained method is then tested with the offline test data (*dataTest.arff*).

5. Experimental Results

5.1. Heuristic Method Results

The real time recognition phase results are summarized in the Figures 8 to 13. The success rates of the outputs are almost 100%, and the stabilization code avoids the flickers in the data, which makes the output more reliable and efficient in daily use of the program.
To compare the offline phase results with the machine learning method results, the stabilization part of the real time code is omitted and all frames are recognized one by one. Therefore if there are 150 frames per sign, there are exactly 150 recognition results for that particular sign. The sign specific results can be seen at the confusion matrix displayed in the Table 1. As seen in the table, all of the signs are recognized with 100% success rate, except the *Baby O Handshape* with 84% and *Open 8 Handshape* with 98.66%

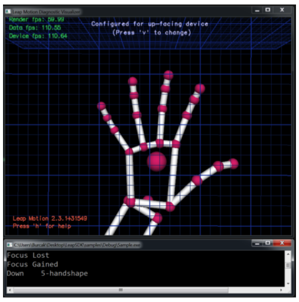


Figure 8:
5 Handshape

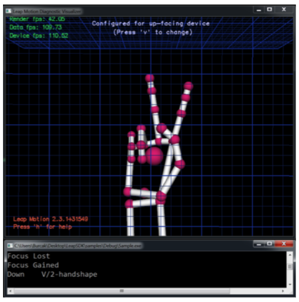


Figure 9:
V & 2 Handshape

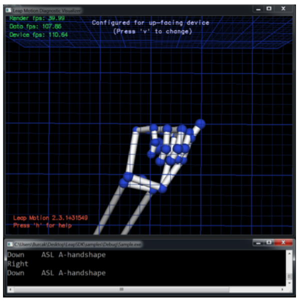


Figure 10:
ASL A Handshape

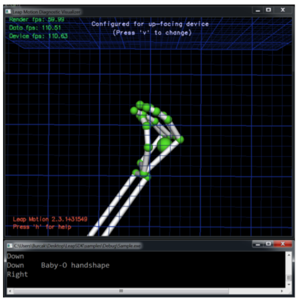


Figure 11:
Baby O Handshape

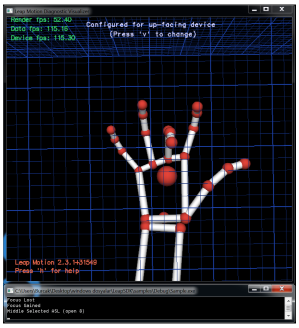


Figure 12:
Open 8 Handshape

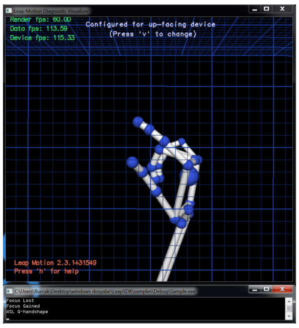


Figure 13:
ASL Q Handshape

success rate. Some of the *Baby O Handshape* sign samples are recognized as *ASL A Handshape* sign because these signs are very similar in the calculation point of view; both signs have no extended finger which is employed in the first rule of their recognition models. As the next step for the *ASL A Handshape* sign's recognition, the distance between the thumb's distal interphalangeal joint, and the index finger's proximal interphalangeal joint is calculated. If this distance is less than a threshold value then the system decides that the sign is *ASL A Handshape*. As the second step for *Baby O Handshape*'s recognition, the distance between the thumb and the middle finger's tip positions is calculated. If the distance is less than a threshold value, it is recognized as *Baby O Handshape*. Therefore when the actual sign is *Baby O Handshape*, sometimes the distance calculated for the recognition of the *ASL A Handshape* can be less than the threshold value because of the noise in the data. The *Open 8* sign is another sign which is occasionally misclassified. As it can be seen from the real time phase results, *Open 8 Handshape* is very similar to *5 Handshape*; just the middle finger's metacarpo-phalangeal joint angle is smaller in the *Open 8 Handshape*. In the recognition model, the middle finger's tip direction is compared to the palm's normal vector's direction. When the result of this comparison is less than a threshold value, the sign is recognized as *Open 8 Handshape*. Because of some slight detection errors of the middle finger's direction vector, a small percent (1.33%) of *Open 8 Handshape* sign samples are misclassified as *5 Handshape*.

5.2. Machine Learning Method Results

A Machine Learning method, namely *Random Forest* is also tested in the study. First, the model is trained using the *dataTrain.arff* file with 10 folds cross-validation. Then to test the trained model, the *dataTest.arff* file is used with the same method as the supplied test set. In the test there were 2700 instances and 139 attributes as explained before. There are no omitted attributes in the test. While using the *Random Forest* method, *Weka* constructed 100 trees by considering eight random features, and model building took 2.74 seconds. The test results of the *Random Forest* method are as follows: The correctly classified instances count is 2527 which is equivalent to 93.5926%, there are 173 misclassified instances which is equivalent to 6.4074%.

The sign-specific results can be seen in the confusion matrix in table 2. As displayed in the Table 2, some signs are recognized with 100% success rate. Those signs are: *V & 2 Handshape*, *L Handshape*, *O Handshape*, *C Handshape*, *ASL A Bar Handshape*, *ASL 8 Handshape*, *ASL I&H Handshape*, *ASL I Handshape*, *ASL Y Handshape*, *ASL 3 Handshape*, *4 Claw Handshape*, *8 Handshape*, *9 Handshape*, and *Open 8 Handshape*. Other signs have lower success rates, as the *5 Handshape* with 65.33%, the *ASL A Handshape* with 80%, the *Baby O Handshape* with 50.66%, and *ASL Q Handshape* signs with a 88.66% success rate.
The signs which had lower recognition rates are *5 Handshape*, *ASL A Handshape*, *Baby O Handshape*, and *ASL Q Handshape* is predicted to be *Open 8 Handshape*, *Baby O Handshape*, *8 Handshape*, *C Handshape* respectively.

5.3. Overall Results

In both methods, the success rate is lower in the same signs, mainly *5 Handshape* vs. *Open 8 Handshape* and *ASL A Handshape* vs. *Baby-O Handshape*. In the *Random Forest* method, *Open 8 Handshape* is also confused with *ASL I Handshape* and *8 Handshape*, and additionally *ASL Q Handshape* is confused with *C Handshape*. The success rate results of every hand shape for both methods can be seen in Table 3. The machine learning method's success rates for the signs are all lower than the heuristic method's success rates except for the *Open 8 Handshape*, which has a very small difference.
The results show that the proposed heuristic model is more robust and precise with 99.03% success rate as average result, while the machine learning method's average success rate remained at 93.59%. Also, the machine learning method requires big sampling sets to create the learning

model while the heuristic method needs none, which makes it easier to integrate more signs. Therefore, for the recognition of the hand shape primitives of Turkish Sign Language, the proposed heuristic method is precise and preferable, especially in real-time recognition systems.

Act.\Pred.	Handshapes																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 (5)	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2 (V & 2)	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 (L)	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 (O)	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 (C)	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0
6 (asl ABar)	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0
7 (ASL 8)	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0
8 (ASL IH)	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0
9 (ASL I)	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0
10 (ASL Y)	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0
11 (ASL 3)	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0
12 (4 Claw)	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0
13 (8)	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0
14 (9)	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0
15 (ASL A)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0
16 (Baby O)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	24	126	0	0
17 (Open 8)	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	148	0
18 (ASL Q)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150

Table 1: Confusion Matrix Table for Heuristic Results (Actual vs. Predicted)

Act.\Pred.	Handshapes																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 (5)	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	52	0
2 (V & 2)	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3 (L)	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4 (O)	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5 (C)	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0	0
6 (asl ABar)	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0	0
7 (ASL 8)	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0	0
8 (ASL IH)	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0	0
9 (ASL I)	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	0
10 (ASL Y)	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0
11 (ASL 3)	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0
12 (4 Claw)	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0
13 (8)	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0
14 (9)	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0
15 (ASL A)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	120	30	0	0
16 (Baby O)	0	0	0	0	0	0	0	0	30	0	0	0	39	0	5	76	0	0
17 (Open 8)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0
18 (ASL Q)	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	133

Table 2: Confusion Matrix Table for Machine Learning (Actual vs. Predicted)

Hand Shape	Heuristic	RF
5 Handshape	100%	65.33%
V & 2 Handshape	100%	100%
L Handshape	100%	100%
O Handshape	100%	100%
C Handshape	100%	100%
ASL ABar Handshape	100%	100%
ASL 8 Handshape	100%	100%
ASL I & H Handshape	100%	100%
ASL I Handshape	100%	100%
ASL Y Handshape	100%	100%
ASL 3 Handshape	100%	100%
4 Claw Handshape	100%	100%
8 Handshape	100%	100%
9 Handshape	100%	100%
ASL A Handshape	100%	80%
Baby O Handshape	84%	50.66%
Open 8 Handshape	98.66%	100%
ASL Q Handshape	100%	88.66%
Average	**99.03%**	**93.59%**

Table 3: Success Rate of Hand Shapes for Both Heuristic and Machine Learning Methods

6. Acknowledgements

This work is supported by "Avatar based Machine translation system for Turkish Sign Language" TUBITAK 1003 (FATIH) project with project number 114E263.

We wish to express our sincere gratitude to Researcher Yakup ÖZDEN for his technical assistance about machine learning.

7. Bibliographical References

Chuan, C.-H., Regina, E., and Guardino, C. (2014). American sign language recognition using leap motion sensor. In *Machine Learning and Applications (ICMLA), 2014 13th International Conference on*, pages 541–544.

Haberdar, H. and Albayrak, S. (2005). Real time isolated turkish sign language recognition from video using hidden markov models with global features. In Pınar Yolum, et al., editors, *Computer and Information Sciences - ISCIS 2005*, volume 3733 of *Lecture Notes in Computer Science*, pages 677–687. Springer Berlin Heidelberg.

Keskin, C., Kıraç, F., Kara, Y., and Akarun, L. (2013). Real time hand pose estimation using depth sensors. In Andrea Fossati, et al., editors, *Consumer Depth Cameras for Computer Vision*, Advances in Computer Vision and Pattern Recognition, pages 119–137. Springer London.

Kubuş, O. (2008). An analysis of turkish sign language (tid) phonology and morphology. Master's thesis, MIDDLE EAST TECHNICAL UNIVERSITY.

Köse, H., Akalin, N., and Uluer, P. (2014). Socially interactive robotic platforms as sign language tutors. *International Journal of Humanoid Robotics*, 11(01):1–22.

Köse, H., Uluer, P., and Akalin, N. (2015a). Sign language games in child-robot interaction. In *Proceedings of the XVII World Congress of the World Federation of the Deaf (WFD)*, pages 61–62.

Köse, H., Uluer, P., Akalin, N., Yorganci, R., Ozkul, A., and Ince, G. (2015b). The effect of embodiment in sign language tutoring with assistive humanoid robots. *International Journal of Social Robotics*, 7(4):537–548.

Leap Motion, I. (2016). Api overview.

Mohandes, M., Aliyu, S., and Deriche, M. (2014). Arabic sign language recognition using the leap motion controller. In *Industrial Electronics (ISIE), 2014 IEEE 23rd International Symposium on*, pages 960–965.

Villarreal, M. R. (2007). File:scheme human hand bones-en.svg - image is retouched by leap motion, inc.

8. Language Resource References

Kubuş, O. (2008). An analysis of turkish sign language (tid) phonology and morphology. Master's thesis, MIDDLE EAST TECHNICAL UNIVERSITY.

Appendix: Recognition Rules of the Heuristic Model

Rule for C Handshape

```
For C-handshape:
    IF Extended Finger Count is 1
        IF Sign is !(is 4-Claw)
        AND Sign is !(is ASL A-bar)
        AND Sign is !(is ASL I-handshape)
        AND 0th Extended Finger is "Thumb"
        AND Distance Btw Index and Middle Finger Tip Positions > DISERRORMARGIN
            IF Angle Btw Index Finger's Proximal bone's Direction and Palm Normal Vectors < PI / 2 + ERRORMARGIN
                IF ERRORMARGIN < Angle Btw Index Finger's Proximal and Intermediate Bones' Direction Vectors < PI / 2
                    IF Palm Direction is "Left" or "Right"
                        Sign is "C-handshape"
```

Rule for ASL I&H Handshape

```
For ASL I&H:
    IF Extended Finger Count is 2
        IF 0th Extended Finger is "Index"
        AND 1th Extended Finger is "Pinky"
            IF Palm Direction is "Up"
                Sign is "ASL I&H"
```

Rule for L Handshape

```
For L-handshape:
    IF Extended Finger Count is 2
        IF 0th Extended Finger is "Thumb"
        AND 1th Extended Finger is "Index"
            IF Palm Direction is "Down"
            AND Angle Btw Thumb and Index Finger Direction Vectors > LIMITANGLEFORL
            AND Angle Btw Index Finger's Proximal and Intermediate Bones' Direction Vectors < ERRORMARGIN
                Sign is "L-handshape"
```

Rule for Open 8 Handshape

```
For Open 8 Handshape:
    IF Extended Finger Count is 5
        IF Angle Btw Middle Finger Direction and Palm Normal Direction Vectors <= (PI*1/4)
            Sign is "Middle selected ASL (open 8)"
```

Rule for Q Handshape

```
For ASL Q-handshape:
    IF Extended Finger Count is 2
        IF 0th Extended Finger is "Thumb"
        AND 1th Extended Finger is "Index"
            IF Sign is !(is L-handshape)
            AND Palm Direction is "Left" or "Rigth"
            AND Angle Btw Thumb and Index Finger Direction Vectors < LIMITANGLEFORL
                Sign is "ASL Q-handshape"
```

Rule for ASL I Handshape

```
For ASL I-handshape:
    IF Extended Finger Count is 1
        IF Sign is !(is 4-Claw)
        AND Sign is !(is ASL A-bar)
        AND 0th Extended Finger is "Pinky"
            Sign is "ASL I-handshape"
```

Rule for 4 Claw Handshape

```
For 4-Claw:
    IF Extended Finger Count is 1
        IF 0th Extended Finger is "Thumb"
        AND Palm Direction is "Down"
        AND Angle Btw Index Finger's Proximal bone's Direction and Palm Normal Vectors < PI / 2
            Sign is "4-Claw"
```

Rule for V&2 Handshape

```
For V/2-handshape:
    IF Extended Finger Count is 2
        IF 1th Extended Finger is "Middle"
        AND 0th Extended Finger is "Index"
            IF Palm Direction is "Down"
            AND Angle Btw Index and Middle Finger Direction Vectors > DegreeToRadian(11)
                Sign is "V/2-handshape"
```

Rule for ASL 3 Handshape

```
For ASL 3-handshape:
    IF Extended Finger Count is 3
        IF 0th Extended Finger is "Thumb"
        AND 1st Extended Finger is "Index"
        AND 2th Extended Finger is "Middle"
            IF Palm Direction is "Down"
                Sign is "ASL 3-handshape"
```

Rule for ASL A Bar Handshape

```
For ASL A Bar:
    IF Extended Finger Count is 1
        IF Sign is !(is 4-Claw)
        AND 0th Extended Finger is "Thumb"
        AND Distance Btw Index and Middle Finger Tip Positions<DISERRORMARGIN
        AND Palm Direction is "Down"
            Sign is "ASL A-bar"
```

Rule for ASL 8 Handshape

```
For ASL 8-handshape:
    IF Extended Finger Count is 3
        IF Sign is !(is ASL 3-handshape)
        AND 0th Extended Finger is "Index"
        AND 1st Extended Finger is "Ring"
        AND 2nd Extended Finger == "Pinky"
            IF Distance Btw Thumb and Middle Finger Tip Positions < DISERRORMARGIN
            AND Distance Btw Thumb and Index Finger Tip Positions > DISERRORMARGIN
                Sign is "ASL 8-handshape"
```

Rule for O Handshape

```
For O-handshape:
    IF Extended Finger Count is 3
        IF Sign is !(is ASL 3-handshape)
        AND Sign is !(is ASL 8-handshape)
        AND 0th Extended Finger is "Middle"
        AND 1th Extended Finger is "Ring"
        AND 2th Extended Finger is "Pinky"
            IF Distance Btw Thumb and Index Finger Tip Positions < DISERRORMARGIN
            AND Distance Btw Thumb and Middle Finger Tip Positions > DISERRORMARGIN
                Sign is "O-handshape"
```


Rule for O Handshape

```
For O-handshape:
    IF Extended Finger Count is 3
        IF Sign is !(is ASL 3-handshape)
        AND Sign is !(is ASL 8-handshape)
        AND 0th Extended Finger is "Middle"
        AND 1th Extended Finger is "Ring"
        AND 2th Extended Finger is "Pinky"
            IF Distance Btw Thumb and Index Finger Tip Positions < DISERRORMARGIN
            AND Distance Btw Thumb and Middle Finger Tip Positions > DISERRORMARGIN
                Sign is "O-handshape"
```


Rule for 5 Handshape

```
For 5-handshape:
    IF Extended Finger Count is 5
        IF Sign is !(is ASL Open 8 handshape)
        AND Palm Direction is "Down"
        AND Distance Btw Index and Middle Finger Tip Positions > DISERRORMARGIN
        AND Distance Btw Ring and Pinky Finger Tip Positions > DISERRORMARGIN
            Sign is "5-handshape"
```

Rule for ASL Y Handshape

```
For ASL Y-handshape:
    IF Extended Finger Count is 2
        IF 0th Extended Finger is "Thumb"
        AND 1th Extended Finger is "Pinky"
            IF Palm Direction is "Down"
                Sign is "ASL Y-handshape"
```

Rule for 8 Handshape

```
For 8-handshape:
    IF Extended Finger Count is 0
        IF Palm Direction is "Down"
        AND Distance Btw Ring and Pinky Finger Tip Positions<DISERRORMARGIN
        AND Distance Btw Thumb and Ring Finger Tip Positions<DISERRORMARGIN
        AND Distance Btw Thumb and Middle Finger Tip Positions>DISERRORMARGIN
        AND Angle Btw Index Finger's Proximal Bone's Direction and Hand's Direction Vectors > PI / 2
        AND Angle Btw Index Finger's Proximal and Intermediate Bones' Direction Vectors > PI / 2 - ERRORMARGIN
        AND Angle Btw Middle Finger's Proximal and Intermediate Bones' Direction Vectors > PI / 2 - ERRORMARGIN
            Sign is "8-handshape"
```

Rule for 9 Handshape

```
For 9-handshape:
    IF Extended Finger Count is 0
        IF Sign is !(is 8-handshape)
        AND Distance Btw Ring and Pinky Finger Tip Positions<DISERRORMARGIN
        AND Distance Btw Thumb and Middle Finger Tip Positions<DISERRORMARGIN
        AND Distance Btw Index and Middle Finger Tip Positions > DISERRORMARGIN
        AND Angle Btw Index Finger's Proximal Bone's Direction and Hand's Direction Vectors >PI / 2
        AND Angle Btw Index Finger's Proximal and Intermediate Bones' Direction Vectors > PI / 2 - ERRORMARGIN
            Sign is "9-handshape"
```

Rule for Baby O Handshape

```
For Baby-O handshape:
    IF Extended Finger Count is 0
        IF Sign is !(is 8-handshape)
        AND Sign is !(is 9-handshape)
        AND Palm Direction is "Down"
        AND Distance Btw Index and Middle Finger Tip Positions < DISERRORMARGIN
        AND Distance Btw Ring and Pinky Finger Tip Positions < 40
        AND Distance Btw Thumb and Middle Finger Tip Positions < DISERRORMARGIN
        AND Angle Btw Thumb Finger's Distal and Intermediate Bones' Direction Vectors < DegreeToRadian(30)
            Sign is "Baby-O handshape"
```

Rule for ASL A Handshape

```
For ASL A-handshape:
    IF Extended Finger Count is 0
        IF Sign is !(is 8-handshape)
        AND Sign is !(is 9-handshape)
        AND Sign is !(is Baby-O handshape)
        AND Palm Direction is "Down" or "Up"
        AND Angle Btw Thumb Finger's Direction and Hand's Direction Vectors  < PI - ERRORMARGIN
            Sign is "ASL A-handshape"
```

The Importance of 3D Motion Trajectories for Computer-based Sign Recognition

Mark Dilsizian*, Zhiqiang Tang*, Dimitris Metaxas*, Matt Huenerfauth, and Carol Neidle*****

*Rutgers University , **Rochester Institute of Technology, ***Boston University,
110 Frelinghuysen Road, Piscataway, NJ 08854 ,
*Golisano College of Computing and Information Sciences, 152 Lomb Memorial Drive, Rochester, NY 14623
**Boston University Linguistics Program, 621 Commonwealth Ave., Boston, MA 02215
mdil@cs.rutgers.edu, zt53@cs.rutgers.edu, dnm@rutgers.edu, matt.huenerfauth@rit.edu, carol@bu.edu

Abstract

Computer-based sign language recognition from video is a challenging problem because of the spatiotemporal complexities inherent in sign production and the variations within and across signers. However, linguistic information can help constrain sign recognition to make it a more feasible classification problem. We have previously explored recognition of linguistically significant 3D hand configurations, as start and end handshapes represent one major component of signs; others include hand orientation, place of articulation in space, and movement. Thus, although recognition of handshapes (on one or both hands) at the start and end of a sign is essential for sign identification, it is not sufficient. Analysis of hand and arm movement trajectories can provide additional information critical for sign identification. In order to test the discriminative potential of the hand motion analysis, we performed sign recognition based exclusively on hand trajectories while holding the handshape constant. To facilitate this evaluation, we captured a collection of videos involving signs with a constant handshape produced by multiple subjects; and we automatically annotated the 3D motion trajectories. 3D hand locations are normalized in accordance with invariant properties of ASL movements. We trained time-series learning-based models for different signs of constant handshape in our dataset using the normalized 3D motion trajectories. Results show significant computer-based sign recognition accuracy across subjects and across a diverse set of signs. Our framework demonstrates the discriminative power and importance of 3D hand motion trajectories for sign recognition, given known handshapes.

Keywords: ASL, hand tracking, sign recognition, sign motion trajectory estimation

1. Introduction

Recognizing a large set of ASL signs is a difficult challenge when posed strictly as a computer vision classification problem. Classification would require vast amounts of training data representing a range of subject-specific signing variations. However, top-down linguistic knowledge imposed on the data analysis can help constrain the problem in order to make learning and sign recognition more feasible.

We have previously achieved high accuracy with respect to handshape recognition from video (Dilsizian et al., 2014). However, for frequently occurring combinations of start and end handshapes, there are large numbers of signs that have those handshapes in common. In the current study, the set of 3D hand configurations has been limited to a set of linguistically important ASL handshapes appropriate for sign recognition.

We demonstrate here that analysis of movement trajectories allows us to achieve high rates of accuracy in discriminating among signs, holding the start and end handshape constant. Thus we expect that combining the techniques reported here with our prior work on handshape recognition will allow us to achieve high accuracy in identification of specific signs.

2. Related Work

Sign recognition has been approached by Vogler et al. (Vogler and Metaxas, 1998; Vogler and Metaxas, 2003) as a time-series modeling problem using Hidden Markov Models (HMMs) over 3D hand models. However, this work is limited to a small vocabulary and laboratory conditions

Other works attempt to recognize signs from real world video. The work in (Ding and Martinez, 2007; Ding and Martinez, 2009) attempts to incorporate modeling of motion trajectories with face and hand configuration recognition. However these works are limited to 2D trajectories and fail to build a stochastic model of the sign that can leverage phonological constraints or inter-subject variation.

Cui et al. (Cui and Weng, 2000) monitor changes in hand observations over time in an attempt to capture *spatiotemporal events*; signs are then classified with respect to these events. In addition, (Buehler et al., 2009) recognize signs by matching windowed video sequences. Although, some success has been achieved, sign recognition research to date has failed to model different components of signs in order to fully leverage important linguistic information.

Some works focus entirely on handshape recognition as an intermediate step to sign recognition. Handshapes are recognized in 2D using nearest neighbor classification in (Potamias and Athitsos, 2008). (Thangali et al., 2011) achieve improvements in handshape recognition by modeling phonological constraints between start and end handshapes, but handshape estimation is limited to 2 dimensions. The handshape model is extended to 3 dimensions with significant improvement in handshape recognition accuracy in (Dilsizian et al., 2014). While these works show good recognition accuracy for handshape, this research has not yet been extended to full sign recognition/identification because of the existence of potentially large numbers of signs with the same start and end handshape pairs.

Although handshape-dependent upper body trajectories have not been previously explored in the literature, 3D human pose and upper-body estimation has been studied extensively. Several generative (Isard and Blake, 1998; Deutscher et al., 2000; Sigal et al., 2004; Bălan et al., 2007) as well as discriminative (Rosales and Sclaroff, 2001; Sminchisescu et al., 2007; Agarwal and Triggs, 2004; Sigal et al., 2007) methods exist for 3D human pose prediction. These works attempt to model multi-valuedness (ambiguities) in mappings from 2D images to 3D poses (Rosales and Sclaroff, 2001; Sminchisescu et al., 2007; Sigal et al., 2007) and employ coarser, global features (Agarwal and Triggs, 2004; Sminchisescu et al., 2007; Sigal et al., 2007) such as silhouette shapes, to generalize the trained models to different scenarios.

Alternatively, (Ferrari et al., 2008) proposed an algorithm to localize upper body parts in 2D images using a coarse-to-fine approach. Humans are coarsely detected using current human detectors. Foreground is extracted within the bounding box using grabcut. The work uses edge-based soft detectors (Yang and Ramanan, 2013) to first detect the torso and head and other parts. The appearance is learned from the detected parts and used to detect further parts using a MAP optimization. The method is extended to spatiotemporal parsing. Anthropometric priors have been extensively applied to constrain this problem.

However, both discriminative methods and the 2D-part based approaches are highly dependent on the use of training data. Because very little 3D upper body trajectory data of ASL signing exists, we are unable to sufficiently train state-of-the-art pose estimation methods.

3. 3D Hand Tracking Dataset

As is well known, along with handshapes, orientation, and place of articulation in space, movement trajectories are an essential component of signs, and thus computer-based recognition of motion patterns is essential for automatic sign recognition. In order to test the ability of a computer vision system to access this discriminative information, we recorded a dataset of 3D upper body motion trajectories across multiple signs and subjects, holding handshapes constant.

3.1. Data Collection

ASL signers

Five ASL signers were recruited on the campus of the Rochester Institute of Technology (home of the National Technical Institute of the Deaf) and from the surrounding community in Rochester, NY, using social media advertising. The participants included 2 men and 3 women, ages 21-32 (mean 24.2). Participants were recorded in a video studio space using a Kinect^{TM} v2 camera system and custom recording software developed at Rutgers University, as described below. A total of 3,627 sign productions were recorded (about 25 tokens each of 139 distinct signs). Because of time limitations, however, for this paper data from 2 signers were prioritized for processing and analysis. The entire set of subjects will be analyzed and discussed in the LREC presentation.

Recording of Motion Trajectories

The Microsoft Kinect^{TM} v2 provides a robust platform for recording 3D upper body joint configurations combined with calibrated 2D color video data. We developed a tool for recording and automatic annotation of joint locations for different ASL signs (see Figure 1).

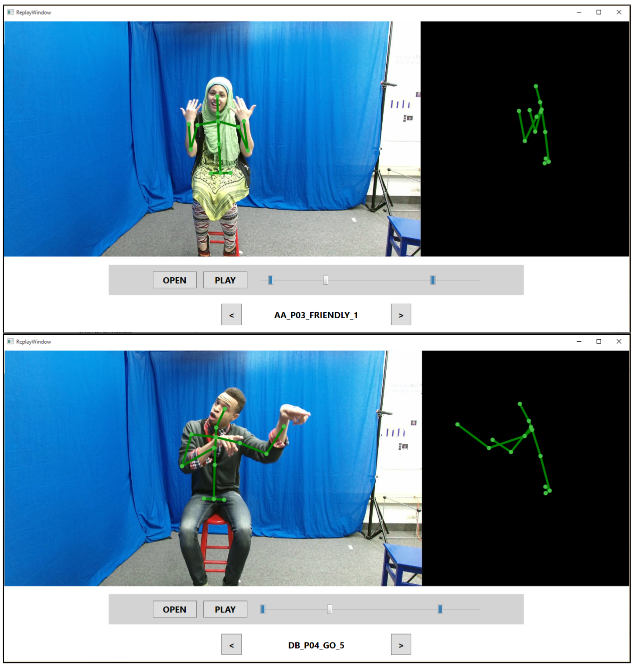


Figure 1: ASL trajectory recording software developed to capture a dataset of 3D ASL movements.

Stimuli

We considered the most common handshapes for 2-handed signs with the same handshapes on both hands throughout the sign production. The signers were recorded as they reproduced two-handed ASL signs shown to them in a video recording of signs from the ASLLVD data set (Neidle et al., 2012) (`http://secrets.rutgers.edu/dai/queryPages/`) with one of three common handshapes used at both the beginning and end of the sign (the B-L, 1, and 5) varying in their motion trajectories.[1] The B-L, 1, and 5 handshapes are illustrated in Figure 2.

3.2. 3D Tracking and Refinement

Because tracking from the Kinect™ v2 sensor is based on a trained discriminative model (Shotton et al., 2013), it is optimized for average case performance. In order to capture subtle discriminative cues in the motion, we refine the output of the camera by taking a cloud of neighboring depth points around each predicted joint location. We constrain each joint to lie near the center-of-mass of its neighborhood. We also smooth these predictions using a Kalman filter.

4. Sign Classification

In order to train a model for the trajectories of different signs, we must ensure that our modeling is invariant to several factors: (1) variation in sign production (*signing style*); (2) variations in body proportion between different subjects; and (3) noise in 3D tracking data.

4.1. Normalizing Motion Trajectories

Improved invariance to different anthropomorphic proportions and ranges of movement can be achieved by normalizing the 3D motion trajectories. First, trajectories are transformed to a common world coordinate system by computing joint locations as the relative distance from the *root* position (located approximately between the hips).

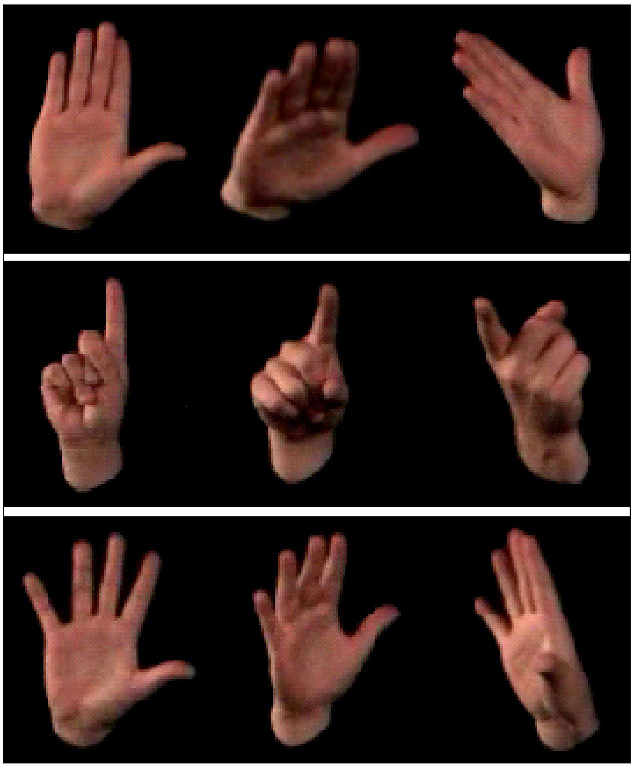


Figure 2: The B-L *(top)*, 1 *(middle)*, and 5 *(bottom)* handshapes.

Second, since it is also important that our model be invariant to differences in the ranges of movement across different subjects. Rather than normalizing according to the overall movement of each trajectory, we normalize over the average range of both left and right hands per subject; this ensures that we preserve the relative range of movement between the left and right hands. An example is shown in Figure 3 for the sign DARK. The bottom row shows the significant reduction of the variance between 2 subjects that results from use of our normalization methodology.

4.2. Training Sign Trajectory Models

In order to overcome noise in 3D hand tracking and variations in signing style, we must learn a robust model that avoids over-fitting to noise or insignificant variation.

The Hidden Markov Model (HMM) has been very popular in the machine learning community and widely applied to speech recognition, part-of-speech tagging, handwriting recognition, bioinformatics, and ASL recognition (Vogler and Metaxas, 1998). HMMs assume that a signing sequence is a Markov process describing how hand locations change through the sign production. A number of states are used to represent different parts of the signing action. These states are not directly visible. Instead, they are perceived

[1]Signs were generated by 5 subjects performing approximately 5 examples of each of the ASL signs glossed in the BU ASLLRP corpora (Neidle et al., 2012) (`http://secrets.rutgers.edu/dai/queryPages/`) as follows: (1)GO-STEADY++, (1)WHEELCHAIR, (1h)HAPPY, (5)WEATHER, (Vulcan)FILE, ABSTRACT+, AFTERNOON, AGREE, ALLERGY, ALL-RIGHT, ALSO, ANSWER, APPLAUSE, ARRIVE, AVERAGE, BALANCE, BEACH, BECOME, BELOW, BETWEEN/SHARE, BLOOD, BOAT, BOIL, BOTHER++, BOX_2, BREAK-DOWN-BUILDING, BRING-1p, BUT, CALM-DOWN, CHEAP, CHILD, CLOSE-WINDOW, COME, CONFLICT/INTERSECTION, COOKING, COOL, CORRECT, CRACK, CYCLE, DEAF-APPLAUSE, DEPEND, DIE, DISAGREE, DIVE, DONT, DURING/WHILE, EASY+, EMBARRASS, END, EVERY-MONTH/RENT, FALL-INTO-PLACE, FAT, FINALLY, FINGERS, FIRE, FOCUS/NARROW, FOOTSTEP, FRESHMAN_3, FRIENDLY, GENERAL, GENERATIONS-AGO, GLORY, GLOVES_2, GO, GRAY, HALL, HANDS, HARP, HERE+, HUMBLE, INSPIRE, JUNIOR_3, KNIFE, LAPTOP, LEAVE-THERE, LIFT, LOUDSPEAKER, MARCHING, MAYBE, MERGE/MAINSTREAM, MOOSE, MOTIVATE, MUSIC, NECKLACE, NEXT-TO, NOISE, OBSCURE, OFTEN+++, ONE-MONTH, OPPOSITE, PANCAKE, PARALLEL, PERSON, PIMPLES, PLEASE/(1h)ENJOY, POPE, PREGNANT, PROGRESS++, PSYCHOLOGY, PUSH, RAIN, REFLECT, REJECT, REQUEST, ROAD, SAD, SCARE, SENIOR, SIGN, SKYSCRAPER, SLOW, SMILE, SOCKS, SPANK, SPIN, STAR, STEEP, STOP, SUCCEED, SUNDAY_2, SWIM, TAP-DANCE, THING, THROAT-HURT, TORNADO, TRAFFIC, TRAVEL, VACATION, VARY, WAIST, WALK, WASH-DISH, WASH-HANDS, WATER-RISE, WEAVE, WHAT, WHEN, WIND, WRAP.

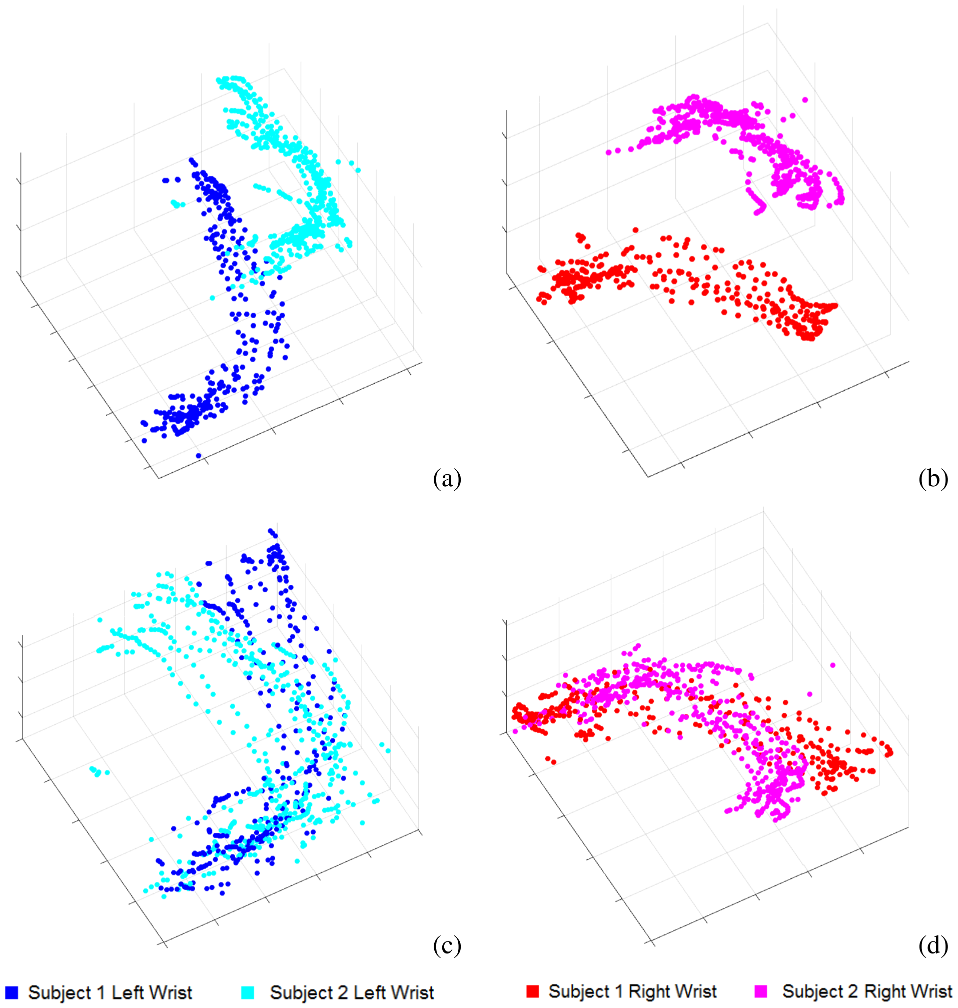


Figure 3: 3D wrist trajectories ({X,Y,Z} Euclidean locations) comparing multiple productions of the ASL sign glossed as DARK by each of two signers. The *top row*, (a) and (b), shows the original data space with evident variations between subjects with respect to sign production and anthropomorphic proportions. The *Bottom Row*, (c) and (d), shows the normalized data space which maximizes inter-subject overlap of trajectories. *Note: the PDF file of this paper contains interactive 3D content accessible by clicking on the figure.*

indirectly through depth image observations. An observation likelihood distribution models the relationship between the states and the observation.This likelihood distribution is represented by a mixture-of-Gaussian(MoG) density function, which is a combination of several Gaussian distribution components. Based on the previous state and the current observation, the HMM may switch from one state to another. During training, the number of states and the number of components in the mixture-of-Gaussian likelihood distribution are chosen using a model selection method known as the Bayesian Information Criterion (BIC). This BIC technique selects the optimal model that best describes the statistics of the training data while avoiding over-fitting. Therefore, using using the BIC technique allows for improved generalization to previously unseen test data.

In order to classify a given sign, we train a Support Vector Machine Hidden Markov Model (SVM-HMM) (Altun et al., 2003). The SVM-HMM is a discriminative sequence labeling model that combines the advantages of HMM and SVM by assuming Markov chain dependency structure between labels and using dynamic programming for optimal inference and learning. At the same time, the learning is based on a discriminative, maximum margin principle that can account for overlapping features. Moreover, unlike HMMs, it can learn the non-linear discriminant functions using kernel-based inputs. An SVM-HMM is trained for each sign which can best be discriminated from all other motion trajectories. This model implicitly captures properties of the motion that are invariant across different examples of the same sign.

5. Results

We train an SVM-HMM for each sign (with constant handshape) and use cross-validation and a two-tailed significance test to determine the parameters (states and Gaussian mixture components) of our SVM-HMMs. Sign labels are assigned to each test sequence according to the SVM-HMM that returns the minimum log-likelihood indicating that the sequence belongs to a trained sign trajectory.

Despite the fact that the sample included some signs with relatively similar motion patterns, we were able to discriminate among these signs with an average of 78.0% accuracy (with cross-validated 50/50 training/testing split). Accuracy by handshape is shown in Table 1.

Handshape	Signs Trained/Tested	Accuracy (%)
B-L	67	75.7
1	35	80.2
5	37	80.3

Table 1: Percent accuracy and number of signs trained and tested (5-10 examples per subject)

While initial results leave some room for improvement, the correct sign classification is located in the top 3 ranked estimations in 96.1% of test examples. We have only used data thus far from two of the subjects. As additional subjects are incorporated into the SVM-HMM model, more general and robust discrimination should be possible. Moreover, additional information from the upper body tracking (i.e. limb locations, body leaning, etc.) can be integrated to improve recognition rates. Overall, the trajectory classification results suggest that a complete sign language recognition framework is feasible when this approach is combined with previously demonstrated handshape recognition.

In order to test the robustness of our modeling, we also tested with different percentages of training and testing splits ($10-90\%$) using cross-validation on 30 common B-L signs. Results across different sized training sets are shown in Figure 4. The stability in sign recognition accuracy even for low percentages of training data suggests that our approach is scalable and can discriminate among signs even when trained on a small set of examples. This is a necessary and critical property to any framework that seeks to scale to a significantly large set of signs and variations.

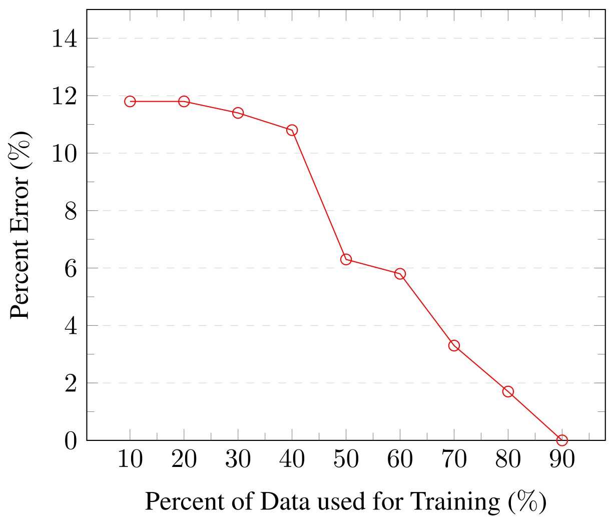


Figure 4: Sign recognition error rates across 30 signs (2-handed B-L handshape) and 2 subjects for different sized training and testing sets.

6. Conclusion

We show here that modeling movement trajectories of the hands provides important information that can be combined with previously demonstrated handshape recognition for purposes of discriminating among ASL signs. We chose a sample of 139 signs that have the most common combination of start and end handshape for 2-handed signs (i.e., signs that use the so-called B-L, 1, and 5 handshapes) throughout the articulation of the sign. We demonstrate a framework and methodology for classifying signs according to 3D motion trajectories.

The next step is to extend this method to construct a full system for sign recognition/identification from video based on a combination of the methods that we have developed for (1) handshape recognition and (2) analysis of motion trajectories. We plan to report on the extension of these preliminary results to larger sets of signs with varying handshapes and motion trajectories, and larger numbers of signers, in the LREC presentation.

7. Bibliographical References

Agarwal, A. and Triggs, B. (2004). 3D Human Pose from Silhouettes by Relevance Vector Regression. In *Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages II–882. IEEE.

Altun, Y., Tsochantaridis, I., Hofmann, T., et al. (2003). Hidden Markov Support Vector Machines. In *ICML*, volume 3, pages 3–10.

Bălan, A. O., Sigal, L., Black, M. J., Davis, J. E., and Haussecker, H. W. (2007). Detailed Human Shape and Pose from Images. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE.

Buehler, P., Zisserman, A., and Everingham, M. (2009). Learning Sign Language by Watching TV (using Weakly Aligned Subtitles). In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2961–2968. IEEE.

Cui, Y. and Weng, J. (2000). Appearance-based Hand Sign Recognition from Intensity Image Sequences. *Computer Vision and Image Understanding*, 78(2):157–176.

Deutscher, J., Blake, A., and Reid, I. (2000). Articulated Body Motion Capture by Annealed Particle Filtering. In *Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 126–133. IEEE.

Dilsizian, M., Yanovich, P., Wang, S., Neidle, C., and Metaxas, D. N. (2014). A New Framework for Sign

Language Recognition based on 3D Handshape Identification and Linguistic Modeling. In *LREC*, pages 1924–1929.

Ding, L. and Martinez, A. M. (2007). Recovering the Linguistic Components of the Manual Signs in American Sign Language. In *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*, pages 447–452. IEEE.

Ding, L. and Martinez, A. M. (2009). Modeling and Recognition of the Linguistic Components in American Sign Language. *Image and vision computing*, 27(12):1826–1844.

Ferrari, V., Marin-Jimenez, M., and Zisserman, A. (2008). Progressive Search Space Reduction for Human Pose Estimation. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE.

Isard, M. and Blake, A. (1998). Condensation - conditional Density Propagation for Visual Tracking. *International Journal of Computer Vision*, 29(1):5–28.

Neidle, C., Thangali, A., and Sclaroff, S. (2012). Challenges in Development of the American Sign Language Lexicon Video Dataset (ASLLVD) Corpus. In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC 2012, Istanbul, Turkey*.

Potamias, M. and Athitsos, V. (2008). Nearest Neighbor Search Methods for Handshape Recognition. In *Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments*, page 30. ACM.

Rosales, R. and Sclaroff, S. (2001). Learning Body Pose via Specialized Maps. In *Advances in Neural Information Processing Systems*, pages 1263–1270.

Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., and Moore, R. (2013). Real-time Human Pose Recognition in Parts from Single Depth Images. *Communications of the ACM*, 56(1):116–124.

Sigal, L., Bhatia, S., Roth, S., Black, M. J., and Isard, M. (2004). Tracking Loose-limbed People. In *Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages I–421. IEEE.

Sigal, L., Balan, A., and Black, M. J. (2007). Combined Discriminative and Generative Articulated Pose and Non-rigid Shape Estimation. In *Advances in Neural Information Processing Systems*, pages 1337–1344.

Sminchisescu, C., Kanaujia, A., and Metaxas, D. N. (2007). BM^3E: Discriminative Density Propagation for Visual Tracking. *Pattern Analysis and Machine Intelligence (PAMI)*, 29(11):2030–2044.

Thangali, A., Nash, J. P., Sclaroff, S., and Neidle, C. (2011). Exploiting Phonological Constraints for Handshape Inference in ASL Video. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 521–528. IEEE.

Vogler, C. and Metaxas, D. (1998). ASL Recognition Based on a Coupling between HMMs and 3D Motion Analysis. In *Computer Vision, 1998. Sixth International Conference on*, pages 363–369. IEEE.

Vogler, C. and Metaxas, D. (2003). Handshapes and Movements: Multiple-channel American Sign Language Recognition. In *Gesture workshop*, volume 2915, pages 247–258. Springer.

Yang, Y. and Ramanan, D. (2013). Articulated human detection with flexible mixtures of parts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(12):2878–2890.

Linking a Web Lexicon of DSGS Technical Signs to iLex

Sarah Ebling*, Penny Boyes Braem**
* University of Zurich, Zurich, Switzerland
E-mail: ebling@cl.uzh.ch
** Center for Sign Language Research, Basel, Switzerland
E-mail: boyesbraem@fzgresearch.org

Abstract

A website for a lexicon of Swiss German Sign Language equivalents of technical terms was developed several years ago using Flash technology. In the intervening years, the backend research database was migrated from FileMaker to iLex. Here, we report on the development of a web platform that provides access to the same technical signs by extracting the relevant information directly from iLex. This new platform has many advantages: New sets of signs for technical terms can be added or existing ones modified in iLex at any time, and changes are reflected in the web platform upon refreshing the browser. Just as importantly, the new platform can now also be accessed through all major mobile operating systems, as it does not rely on Flash. We describe how information on the glosses, keywords, videos of citation forms, status, and uses of the technical signs is represented in iLex and how the corresponding web platform was built.

Keywords: Swiss German Sign Language, technical signs, iLex, web platform

1. Introduction

For more than twenty years, work has been carried out on a lexicon of Swiss German Sign Language (DSGS) (Boyes Braem, 2001). While the initial form of the lexicon was a FileMaker database, the lexicon has recently been migrated to iLex, a sign language lexicon and corpus software (Hanke and Storz, 2008). It is stored on a university server and is accessible by researchers working on specific projects. This DSGS iLex lexicon currently contains approx. 9000 signs arranged into different sub-lexicons.

Two sub-lexicons of signs together contain 685 technical terms in the domains of nutrition and economy (Boyes Braem et al., 2012). Before the database was migrated to an iLex form, these signs had been made accessible on the web using Flash technology. Here, we report on the development of a web platform that provides access to the same technical signs by extracting the relevant information from iLex. This new platform has many advantages: New sets of signs for technical terms can be added or existing ones modified in iLex at any time, and changes are reflected in the web platform upon refreshing the browser. Just as importantly, the new platform can now also be accessed through all major mobile operating systems, as it does not rely on Flash.

2. iLex

iLex is a client-server application developed specifically to support the creation and exploitation of sign language lexicons and corpora. As such, it is somewhat similar to ELAN, a common tool for creating sign language and spoken language corpora (Wittenburg et al., 2006). In an ELAN transcript, the glosses a user assigns to a sign segment take the form of strings in free-text fields with no underlying consistency check. By contrast, iLex requires the use of glosses that are standardized for a particular sign language by demanding in a first step that the user choose from a set of previously introduced glosses. If an appropriate gloss is not available, a user can create a new one. All occurrences of a gloss in a transcript (gloss tokens) are linked back to their gloss type in the lexicon, and changes of the gloss type affect all gloss tokens in all transcripts. This is possible through an underlying relational database. The PostgreSQL database that is the backbone of iLex consists of 97 tables. Table 1 shows a selection of tables.

```
concepts
illustrations
illustration_instances
illustration_uses
meanings
metadata_descriptions
metadata_features
movies
projects
tags
terms
tiers
transcripts
types
```

Table 1: Selection of tables from iLex PostgreSQL database.

3. DSGS Signs for Technical Terms in the Previous Flash Web Platform

The DSGS signs for selected technical terms in the fields of nutrition (293 concepts) and economy (392 concepts) were analyzed and described in a two-year research project (Boyes Braem et al. 2012). The primary targeted users were DSGS signers, especially students in schools of higher education, who wanted more information in sign language about the meaning of specific technical terms. The website has also been useful to DSGS interpreters looking for sign equivalents for terms in these technical fields. The resulting web lexicon provided the following information about each sign:

- **Gloss**, e.g., ALKOHOLSTEUER_1 [*ALCOHOL-TAX_1*]
- **Keywords** (German): Keywords are included as glosses often reflect only one meaning of a sign. Examples of keywords corresponding to the gloss ALKOHOLSTEUER_1 are “Alkoholsteuer” [*alcohol tax*] and “Steuer” [*tax*].
- Video of the **citation form** of the sign and of up to three widely used form variants
- **Domain**: “nutrition” or “economy”
- **Status** of the sign: “used” by more than three signers, “known” to be used by at least two signers, or “new” (as developed by a group of experienced DSGS signers who had knowledge of the corresponding technical field)
- **Definition** of the term in the form of videotaped DSGS and a written German back-translation of the DSGS signed text. A few definitions come in the form of illustrations as opposed to videos of DSGS signing. Definition of ALKOHOLSTEUER_1 (German back-translation): “Steuer, die im Kaufpreis von Alkohol inbegriffen ist [*tax that is included in the price for alcohol*]”.
- One to three **examples** of the meaning of the concept in videotaped DSGS and back-translated written German. Example (German back-translation): “Beim Kauf von Spirituosen (Bier, Wein usw.) wird mit dem Kaufpreis eine Steuer bezahlt. [*When buying liquors (beer, wine, etc.), a tax is included in the price.*]” For some concepts denoting kitchen utensils (nutrition domain), illustrations or photos of the utensil (e.g., a whisk) are provided instead of text examples.

4. Representation of Technical Signs in iLex

When migrating the DSGS lexical database from FileMaker to iLex, we introduced a sign type in the iLex lexicon for each **gloss** entry in the previous FileMaker database. To link the **videos of the citation forms** to the sign types, we introduced transcripts for the videos that hold token tags of the respective types. Figure 1 (at end of article) shows a transcript containing a token tag of the sign type ALKOHOLSTEUER_1 in a tier “Type/Subtype”. The tag spans the entire duration of the video. Token occurrences of this kind are excluded from any corpus statistics in iLex, as they do not represent meaningful examples of signs in context.

Associated with the technical signs in iLex is metadata information on the **domain** of the sub-lexicon (nutrition or economy). Information on the **status** of a sign was also stored as sign type metadata, using a closed vocabulary that permits the assignment of only one of the three values “used”, “known”, and “new”. The **keywords** accompanying glosses were stored as concepts.

In iLex, technical terms are concepts assigned to specific projects. Terms may carry **definitions**. We stored the German back-translations of the DSGS definitions of our technical signs here. In addition, to link the German definitions with their DSGS video originals, we introduced a transcript for each DSGS video. In the transcript, the German definition is represented as a tag of a tier “Definition”, the length of the tag again being equal to the duration of the video. In addition, the transcript contains a second tier that records the sign type to which the definition is linked. Figure 2 (at end of article) shows the definition of the technical sign ALKOHOLSTEUER_1.

The same procedure was applied for the **examples** of usages of technical signs: Each example received its own transcript that holds the DSGS (video) and German (text) versions along with a link to the sign type.

It was noted in Section 3 that some definitions of technical signs come in the form of pictures as opposed to videos of DSGS signing. The same is true for examples, e.g., for concepts denoting kitchen utensils. iLex allows for such pictures to be linked to terms directly.

5. Web Platform

The new web platform that holds the technical signs is accessible at https://linguistik-signlang.uzh.ch/. Figure 3 (at end of article) shows the previously introduced entry ALKOHOLSTEUER_1 in the web interface.

Currently, users can consult the web lexicon of technical terms in four ways:

1. By searching for (sub-)strings of keywords
2. By restricting the search to one of the two domains nutrition and economy
3. By searching for signs with a given status, where possible values are “used”, “known”, and “new”
4. By searching for signs whose glosses start with a given letter

Combined searches, i.e., searches along more than one of the above parameters, are also possible.

The platform is implemented in PHP. Embedded in the PHP code are SQL statements that extract the information outlined in Sections 3 and 4 (gloss, keywords, video of citation form, status, definition, examples) from the iLex PostgreSQL database. More precisely, the information is queried from the 97 tables that make up the iLex database. To retrieve the information relevant for the web platform of technical signs, the 14 tables shown in Table 1 are consulted. Table 2 shows an SQL sub-query.

```
SELECT types.id, terms.name,
illustration_instances.file_path
FROM projects
JOIN terms ON projects.id=terms.project
JOIN concepts ON terms.concept=concepts.id
JOIN meanings ON concepts.id=meanings.concept
JOIN types ON types.id=meanings.type
LEFT JOIN illustration_uses ON
terms.id=illustration_uses.term
LEFT JOIN illustrations ON
illustration_uses.illustration=illustrations.id
LEFT JOIN illustration_instances ON
illustrations.id =
illustration_instances.illustration
```

Table 2: Sub-query to extract information from the iLex PostgreSQL database.

The database user issuing the statements has read-only rights to prevent possible SQL injection. Moreover, user input is handed over to the built-in PostgreSQL function `pg_escape_string` in PHP that masks a string.
The PHP code with the SQL queries embedded in it will be made available to interested parties upon request.

6. Conclusion

In this paper, we have reported on the development of a web platform that provides access to technical signs in the domains of nutrition and economy by extracting the relevant information from a DSGS instance of iLex. The signs had previously been stored in a FileMaker database.
We have described how information on the glosses, keywords, videos of citation forms, status, and uses of the technical signs was migrated from FileMaker to iLex and how the corresponding web platform was built.
Currently, we are extending the search functionality offered in the web platform. A medium-term goal is to change the design of the website, which is still the same as that of the original (Flash-based) platform.
One of the advantages of the new linking of the web lexicon of signs for technical terms to iLex is making it easier to not only correct or modify existing terms but also add new technical fields to the web lexicon. We are in the process of expanding the website to include name signs for places that are already in the larger iLex lexicon. In the near future, as funding becomes available, we plan to add signs in the fields of sign language linguistics, jurisprudence, medicine, natural hazards (such as earthquakes), language testing techniques, as well as name signs for well-known persons. The structure of the information presented on the website for these additional domains will in most cases be much the same as for the technical domains described above. As the information is based on iLex entries, however, this information will be backed up by the information and analysis necessary for all iLex entries, including metadata on the source of the information, more exact form descriptions (HamNoSys) as well as crosslinks to other lexical items in the larger lexicon and to examples in the annotated videos in the linked corpus. Finally, we expect all of the information from this website to be much more widely accessible as it is now available on tablets and mobile phones as well as on computers.

7. Acknowledgements

We gratefully acknowledge the work of Lorenz Nagele on the implementation of the PHP code. Johannes Graën’s help in formulating and optimizing the SQL queries is much appreciated. We would also like to thank Thomas Hanke for his ongoing support of our work with iLex.

8. Bibliographical References

Boyes Braem, P., Groeber, S., Stocker, H., and Tissi, K. (2012). Weblexikon für Fachbegriffe in Deutschschweizerischer Gebärdensprache (DSGS) und Deutsch. *eDITion: Fachzeitschrift für Terminologie*, 2, pp. 8-14.

Boyes Braem, P. (2001). A multimedia bilingual database for the lexicon of Swiss German Sign Language. *Sign Language & Linguistics*, 4(1/2), pp. 133-143.

Hanke, T., Storz, J. (2008). iLex: A database tool for integrating sign language corpus linguistics and sign language lexicography. In *Proceedings of the 6th Language Resources and Evaluation Conference (LREC)*, pp. 64-67, Marrakech, Morocco.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., and Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In *Proceedings of the 5th Language Resources and Evaluation Conference (LREC)*, pp. 1556-1559, Genoa, Italy.

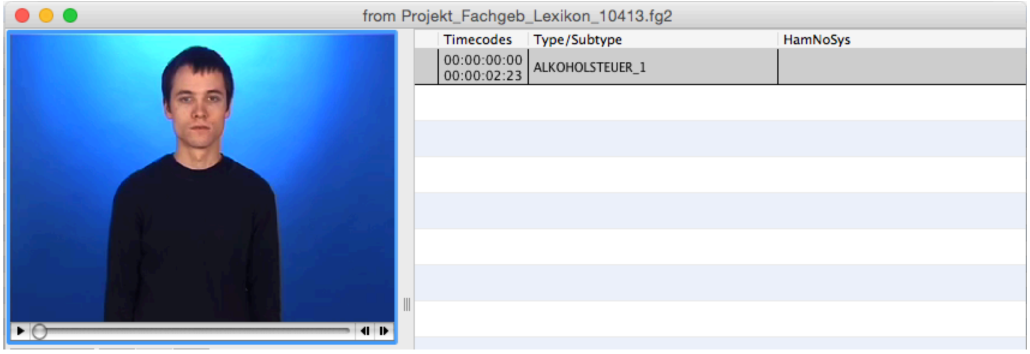


Figure 1: Link between videos of citation forms and sign types in the iLex lexicon: Sample transcript for ALKOHOLSTEUER_1 [*ALCOHOL-TAX_1*].

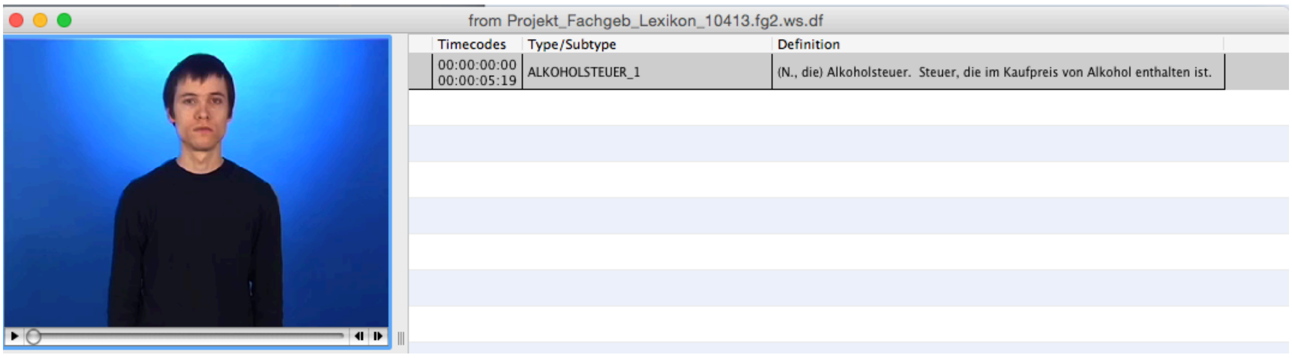


Figure 2: Link between videos of definitions and sign types in the iLex lexicon: Sample transcript for ALKOHOLSTEUER_1 [*ALCOHOL-TAX_1*].

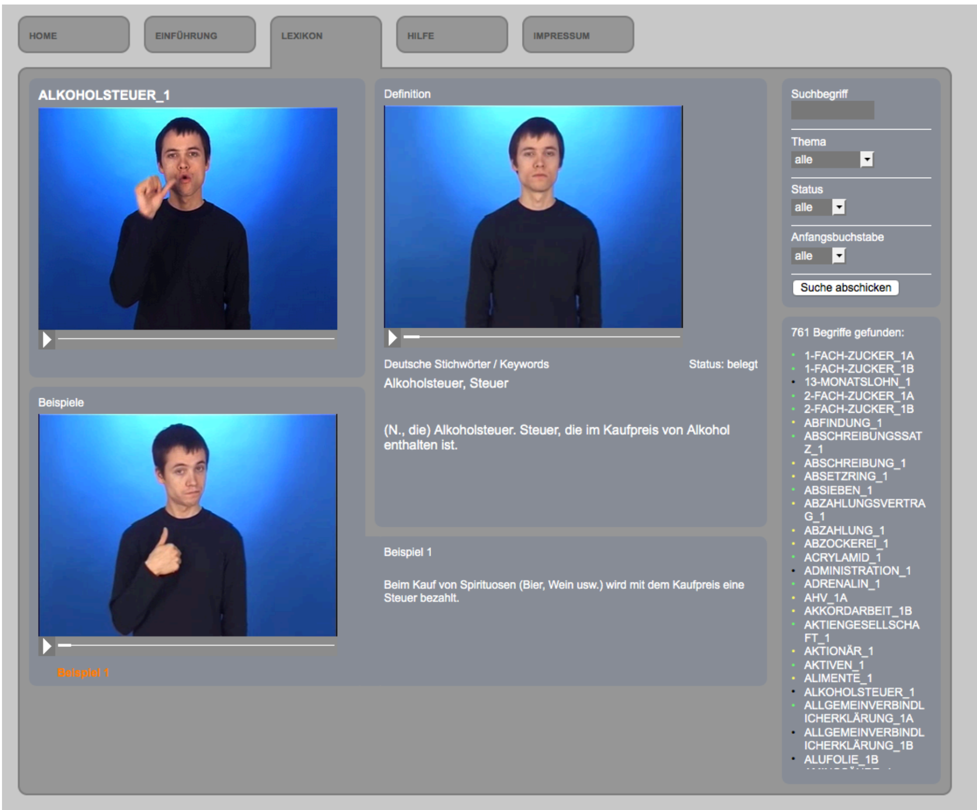


Figure 3: Screenshot of the web platform showing the entry ALKOHOLSTEUER_1 [*ALCOHOL-TAX_1*].

From a Sign Lexical Database to an SL Golden Corpus – the POLYTROPON SL Resource

Efthimiou, E.[1], Fotinea, S-E.[1], Dimou, A-L.[1], Goulas, T.[1], Karioris, P.[1], Vasilaki, K.[2], Vacalopoulou, A.[1], Pissaris, M.[1], Korakakis, D.[1],

[1]ILSP - R.C "Athena", [2]AUTH – Philology Department
[1]Artemidos 6 & Epidavrou, Maroussi, 15125 Athens, Greece
E-mail: {eleni_e, evita, ndimou, tgoulas, pkarior, avacalop}@ilsp.gr, kikivasilaki@yahoo.gr, pissarakia@gmail.gr, korakakis79@gmail.com

Abstract

The POLYTROPON lexicon resource is being created in an attempt i) to gather and recapture already available lexical resources of Greek Sign Language (GSL) in an up-to-date homogeneous manner, ii) to enrich these resources with new lemmas, and iii) to end up with a multipurpose-multiuse resource which can be equally exploited in end user oriented educational/communication services and in supporting various SL technologies. The database that hosts the newly acquired resource, incorporates various SL oriented fields of information, including information on compounding, GSL synonyms, classifier qualities, lemma related senses, semantic groupings etc, and also lemma coding for their manual and non-manual articulation activity. It also provides linking of GSL and Modern Greek equivalent(s) lemma pairs to serve bilingual use purposes. A by-product of considerable value is the parallel corpus which derived from the GSL examples of use accompanying each lemma entry in the dictionary and their translations into Modern Greek. The annotation of the corpus for the entailed signs and assignment of respective glosses in combination with data capturing by both HD and Kinect cameras in three repetitions, allowed for the creation of a golden parallel corpus available to the community of SL technologies for experimentation with various approaches to SL recognition, MT and information retrieval.

Keywords: SL data acquisition, SL lexicon resource, GSL-Greek bilingual dictionary, deaf accessibility services, SL technologies, SL-text parallel golden corpus

1. Introduction

In the framework of research activities undertaken within the POLYTROPON project[1], significant effort has been placed in maintaining and extending a Greek Sign Language (GSL) lexicon dataset which consisted of lemmas captured by means of diverse capturing devices, lemma list construction methodologies and approaches for verification of acceptance by the local deaf community, covering a time space of approximately fifteen years of acquisition phases.

In (Dimou et al., 2014), the rational and methodological principles for revisiting and recapturing the existing GSL lexicon resource have been justified, similarly to the goal of extending the already created lemma list and the adopted scheme of fields of information.

As regards the content, among the crucial issues that had to be faced was the verification of acceptance by the general deaf public of the sign content, as well as the presentation protocol regarding the different sign categories entailed in the lexicon (i.e. adaptation of the pronoun neutral predicate form, classification of classifier based lemmas according to the classifier generating them, and handling of compound sign lemmas as well as special expressions included in the lemma list). In parallel, a number of decisions on data acquisition methodology were related to the purpose of creating a multiuse resource.

To sum up, the POLYTROPON resource is being created in an attempt: i) to gather and recapture already available lexical resources of Greek Sign Language (GSL) in an up-to-date homogeneous manner, ii) to enrich these resources with new lemmas, and iii) to end up with a multipurpose-multiuse resource which is equally exploitable in end user oriented educational/ communication services and in supporting various SL technologies, including information extraction, Web accessibility tools, incorporation of lexical information in natural language processing (NLP) systems for sign language processing as in the case of machine translation (MT) from and into sign language, creation of training material for sign recognition technologies and input to sign synthesis tools enabling signing by virtual signers (avatars), along with simpler tasks which are depending on availability of language resources such as creation of bilingual dictionaries and glossaries.

We refer next to the various aspects of the POLYTROPON resource acquisition process and its current implementation.

2. Content Definition and Acquisition Methodology

The main sources for the POLYTROPON lexicon content are two pre-existing GSL lexicon databases: i) the lemma list of the bilingual (GSL-Modern Greek) multimedia dictionary NOEMA[2] (set to circulation in the form of DVD-ROM in 2001), and ii) the list deriving from the lemmatized GSL segment of the Dicta-Sign corpus[3].

However, both these lemma lists were not adopted in full in the new database. While on technical grounds, the rationale behind recapturing the data was based on the two following facts:

i) a considerable bulk of the available resources was

[1]http://www.ilsp.gr/el/infoprojects/meta?view=project&task=show&id=198

[2] http://www.ilsp.gr/en/services-products/products/item/item/2-noema

[3] http://www.sign-lang.uni-hamburg.de/dicta-sign/portal

captured in the late ’90s, thus being subject to capturing devices limitations of that time &

ii) the DICTA-SIGN lemmas bulk was extracted from a corpus annotation procedure as a result of glosses annotation, meaning that the extracted lemmas were articulated in the context of a phrase, thus being formed away from the typical lemma presentation scheme adopted in a dictionary

There have been recognized a number of serious issues related to language research and language data acquisition methodology, that have dictated the revision of the available lexicon content and the adoption of an acquisition methodology which would exclude any interference from the oral language environment.

2.1 Revision of pre-existing Lexicon Resources and Lemma List Enrichment

As already mentioned, the out-of-date video quality or the lemma in context articulation, were only partially reasons for recapturing the GSL lemmas. Planning of the new capturing procedure provided the opportunity for an in-depth evaluation of existing lemmas against SL linguistic criteria and the rethinking of lemma lists formation. These criteria included parameters such as whether the items already classified as lemmas were real lexical items or classifier constructions to express a concept imported from the environment spoken language, decisions about how sign lemmas have to be presented in the lexicon data base (i.e. the pronoun neutral predicate representation against signing the first person singular pronoun when forming a predicative sign, which is directly influenced from Modern Greek (MG) that lacks a morphology neutral form such as the Gerund form of English, or organizing treatment and presentation of compound lemmas in line with (Liddell & Johnson, 1986) and (Sandler & Lillo–Martin, 2006)), but also critical, whether all already captured lemmas provided formations connected with specific concepts and recognized as such by the (majority of) GSL signing community or were ad hoc formations improvised by informants of the early capturing.

Although such lemma formations were of limited number, this latter case of “mistakes” or “unknown” lemmas turned to be a source error factor in the early NOEMA dataset, which was only noticed through the actual use of the dictionary and active communication with the native GSL signers’ community. Serious consideration has been dedicated to the definition of what kind of new signed data should be included in the lemma list, the main issue being to guarantee that SL grammar principles are met and also the sign(s) representing a given concept are widely accepted by the GSL community. As a consequence of the above, in a first phase the 3.000 video lemmas of general language domain falling within the definition of basic lexicon content (Efthimiou & Katsoyannou, 2001), which formed the lemma list of the NOEMA dictionary, were thoroughly revisited in order to identify needed enhancements or corrections both in respect to content formal representation issues and wide acceptance of the video lemmas. As a result of this work, the NOEMA lemma list has been filtered in respect to GSL wise “peculiar” content and sign forms not widely accepted have either been removed or replaced by more appropriate ones. Similarly, all not appropriately performed entries in respect to morphological markers have been spotted and received a commentary to guide their proper acquisition during the new capturing sessions. In the same line of evaluation, the lemmas extracted from the GSL Dicta-Sign corpus have also been filtered for sense representation in the corpus and sense disambiguation, as well as against all check parameters holding for the NOEMA lemma list. The merge of the two originally available lists formed the initial content of the POLYTROPON lexicon, while GSL synonyms and antonyms of the entailed lemmas provided the first round of lemma list expansion.

Figure. 1: GSL synonyms linked with one sense and one corresponding lemma in Greek.

The structure of information accompanying sign lemmas in the database has partially adopted the scheme followed in the NOEMA database, enriched with information fields which serve the purposes of the POLYTROPON resource. Thus, each sign lemma is associated with one or more equivalents in MG, a gloss, its GSL synonym(s) if any (Fig. 1), and one semantic sense, where possibly different senses of a single form are disambiguated via linking with different entries in the lexicon database (Fig. 2) and different examples of use. The database incorporates a number of further fields for information which become visible only when a specific application needs to exploit this kind of information. Among the information inserted in the database is the category of “special” or “fixed expressions” in both GSL and MG (Fig. 3). Participation of individual lemmas in the formation of special expressions has become visible, in order to allow for the retrieval of such expressions as a result of searching by means of the lemmas involved in their construction. Here the main issue is that lemmas in principle lose their initial sense when they appear in a special expression, which usually expresses a metaphor.

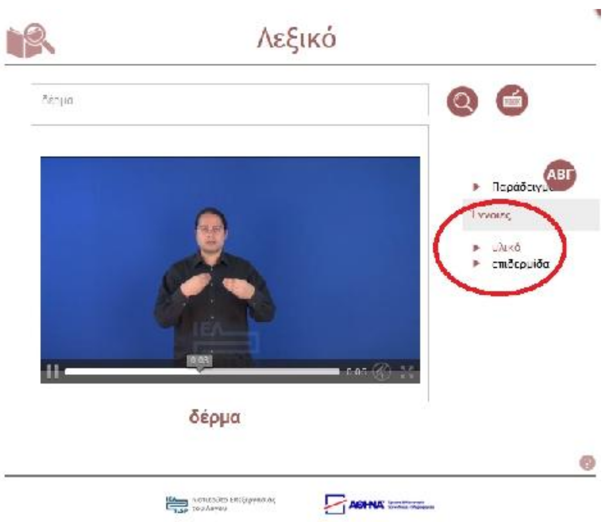


Figure 2: Multiple senses of a single form are visible in the lexicon database.

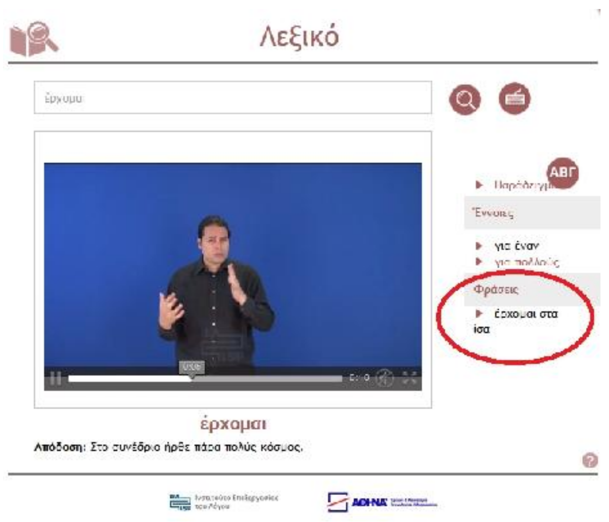


Figure 3: Special expressions –either of GSL or Greek origin– linked with entailed lemmas for easy search.

Furthermore, a number of information fields not visible to external users such as the HamNoSys[4] coding of sign lemmas, feature coding for the non-manual activity involved in sign formation (Pfau & Quer, 2010), GSL grammar markers for i.e. classifier constructions, plural formation and compounding are also available.

Especially in respect to classifier constructions, only those items which are identified by native GSL signers as lexicalized forms representing specific concepts without the need for associating their interpretation with information previously provided in their linguistic context, are treated as autonomous lemmas. Thus, in the adopted lexicon design, classifiers which have not been lexicalized are classified within their signed context and are treated in the lexicon either as bound morphemes adding semantic values or as semantic indicators with pronominal function.

However, the decision to include paradigms of use for each sign lemma (Fig. 4) has proved to be a turning point in the resource development. Since acceptance of the signs has been a priority issue, a significant criterion for the identification of generally accepted sign lemmas was their association with examples of use that would be indicative natural signers' productions clarifying the use of the examined lemma in context.

In order to best serve this goal, the addition of examples to lemma related information was organised as a corpus acquisition task following the good practice developed within the Dicta-Sign project for SL data acquisition (Matthes et al., 2010; 2012), taking measures to eliminate interference from the spoken or written forms of Modern Greek (MG) to the wider possible extend.

Fulfilment of this task was planned and executed by a group of experts where the major presupposition was the strict use of GSL in group discussions. The working group consisted of six signers, including native GSL signers, GSL codas and SL linguists. The native GSL signers of the team were the persons who had reviewed the NOEMA and the GSL Dicta-Sign lemma lists. Discussion of each lemma was based on meaning (=concept) representation, while the examples of use were decided after discussion among the group members while probing the best candidate phrases for each lemma. In the finalization phase and prior to each recording session, lemmas and their examples to be acquired next, were collected and archived in recording material dedicated sessions, where the content of each recording was fixed.

The new acquisition of lemmas and their examples of use was performed by means of one HD and one Kinect camera, while the whole of material was acquired in three repetitions of each item, in order to create a resource appropriate to be exploited in sign recognition.

The recorded example phrases are annotated for the included lemmas and translated into MG, while each lemma is assigned a gloss. This procedure allowed for checks and a remedy of possible inconsistencies with respect to the lemma list, thus ensuring that all signs used in the example of use phrases find an equivalent lemma in the lemma list of the lexicon. Through this procedure, the example phrases provided a further source for lemma enrichment, since a general convention has been that all signs used in the example phrases, need to be searchable and retrieved in the lemma list.

Given that the lexicon entries are constantly enriched, the POLYTROPON lexicon resource has become an expanded database, which is relatively difficult to check for possible mismatches and omissions. In order to facilitate cross-checks and also provide a tool for lexicon inspection by end users who are not necessarily familiar with database structure, a simple interface has provided visualization of various pieces of information related with each lemma as depicted in Figures 1, 2, 3 and 4.

In order to increase the visibility of the POLYTROPON resource by making it known to those interested in using parts of it in SL research and education, a bilingual dictionary for the language pair GSL-MG is already

[4] HamNoSys: the Hamburg Notation System (Hanke, 2004; Prillwitz et al., 1989).

extracted from the lexicon database while the resource is documented for its content and metadata within the *clarin:el* repository, the Greek sector of CLARIN[5], the European infrastructure for language resources and technology. The dictionary content has become available free of charge but subject to Creative Commons (CC) licensing[6].
For its identification in clarin:el, the POLYTROPON resource has received the persistent identifier (PID): http://hdl.grnet.gr/11500/ATHENA-0000-0000-42D5-5 (Fig. 7).

2.2 The POLYTROPN Parallel Corpus

The GSL phrases captured to serve as examples of use of lemmas in context along with their translations to MG constituted a parallel corpus of considerable length and richness, which is available in HD and Kinect for Linux captures in three repetitions for each signed utterance.

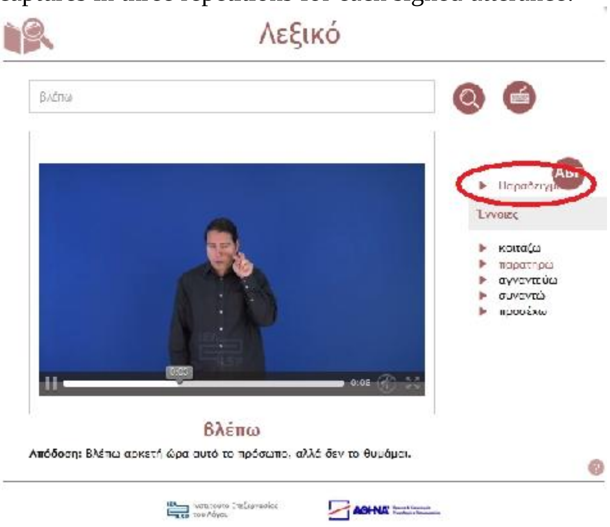


Figure 4: Example of use in GSL connected with its translation in Greek.

As data acquisition progressed, it became obvious that the lexicon database could be further exploited as an independent resource to serve technologies which crucially depend on a "golden" (parallel) corpus such as sign recognition, information retrieval directly from SL video and machine translation. To serve this goal, the POLYTROPON parallel corpus is being annotated in the iLex[7] (Hanke & Storz, 2008) environment. Annotation tiers provide further information on lemma timestamps, glosses, HamNoSys coding for manual activity, non manual features on the sign and utterance level to indicate morpho-phonological, semantic and/or syntactic functions, and also classifier related information. Upon completion of the annotation work, the corpus, which currently entails 3.400 signed utterances, will become available to the research community for experimentation, via the CLARIN infrastructure.

3. The POLYTROPON Database: a Resource for Sign Language Technologies

The parallel corpus created as reported above, is currently exploited in testing an approach to machine translation (MT), while the bilingual MG-GSL dictionary that has been produced as a by-product of the GSL lexicon database structure, -incorporating approximately 10,000 entries at the time of writing- has already been adopted in the official educational content platform of the Greek Ministry of Education to support accessibility of written content by Deaf end users, while a subset of it is incorporated in *e-class* the Greek platform for University level curriculum content, to support accessibility of educational information by deaf students.
Similarly, the HamNoSys coded information of the database is exploited in a simple interface for dynamic SL phrase formation to be used by both L1 and L2 users, while a Web based text accessibility tool is also supported by the POLYTROPON lexicon database.
These lexicon based technologies also exploit a suite of written language technologies including a lemmatizer and a morphological analyzer for MG, necessary to correctly identify and link the various tokens relating to a specific lemma in MG texts.
Such tools form the necessary background to allow successful bilingual connections and search retrieval results in the database underlying the interfaces to be presented next (Efthimiou et al., 2015).

3.1 The POLYTROPN Lexicon: a Resource for Synthetic Signing

Phonological coding of the POLYTROPON lemma list has enabled the development of a simple interface for dynamic synthetic signing, which can be equally used by GSL knowledgeable and non-knowledgeable end users in order to facilitate communication via GSL language productions.
A search box allows retrieval of constituents to compose the wished utterance. Phrase components may be reordered via drag-and-drop actions, while the users not familiar with GSL can advice template based instructions for structuring the phrases they create, according to GSL grammar. Special provision is taken for lexical items not included in the lexicon as well as for proper nouns to be visualized by means of fingerspelling.
The database copy to serve synthetic signing is hosted in the Cloud, while the related interface is currently attached to the set of deaf accessibility tools incorporated in the "Photodentro" platform that hosts the official educational content of primary and secondary levels of the Greek educational system (Fig. 5)[8].

[5] CLARIN: Common Language Resources and Technology Infrastructure (www.clarin.eu).
[6] https://creativecommons.org.
[7] www.sign-lang.uni-hamburg.de/ilex.

[8] The experimental implementation of all integrated tools to "fotodendro" can be reached for experimentation via: sign.ilsp.gr/jas/dev/demo.html.

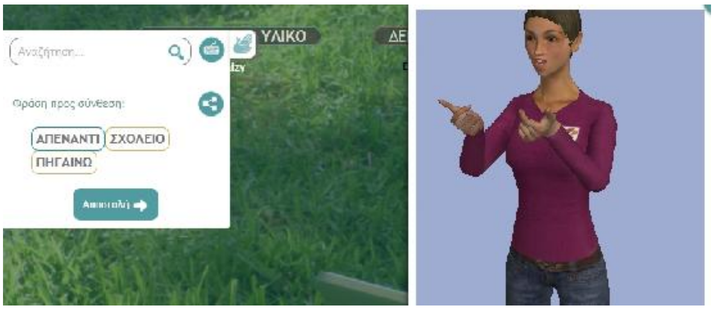


Figure 5: Input to synthetic signing.

Figure 6: Lexicon content linked with Web text accessibility tool.

3.2 The POLYTROPN Bilingual Dictionary Applications

Current on-line bilingual dictionaries based on material from the lexicon database are incorporated in two educational content platforms as deaf accessibility aids. They incorporate language technology tools which facilitate query entry and result retrieval, irrespective of the morphological complexity of the token form of MG used as the source for search (Fig. 8).

In both environments, deaf users may also prefer to insert search items by means of a virtual keyboard for fingerspelling as an alternative input device.

3.3 The POLYTROPN Lexicon as an Accessibility Tool for WEB Content

An especially well accepted application of the POLYTROPON resource in the educational context, is the direct linking of unknown words found in a text in the Web with their GSL equivalents. When activated, this option allows the user to view the GSL sign linked to a given word in a text by simply double clicking on the unknown item as depicted in Fig. 6.

Again language technology tools that run in the background enable retrieval of the proper pair in the lexicon resource irrespective of the morphological form of the search item in the text.

4. Conclusion

The POLYTROPON lexicon database has been created to mainly address SL processing needs in the framework of human language technologies applications and also in service of SL technologies with focus on sign recognition and synthetic signing. Given the scope of the resource and the range of usability cases it is intended to serve, design criteria which had to be satisfied extend from naming conventions of video-lemmas to coding of manual and non-manual elements of each sign for representation via synthetic signing and retrieval purposes. Within a time span of three years, the database has become the richest resource for GSL lexicographic data, its enrichment being steadily in progress. The so far acquired data are already exploited in a number of Web based applications supporting deaf education and communication needs. However, the collection of the resource has also triggered new challenges on technological and SL linguistic grounds. In this context, association of lemmas within an appropriate ontology scheme is required to enable more efficient bilingual associations between GSL and Modern Greek, which will significantly augment accessibility of written Greek texts by Deaf individuals in a variety of communication environments. Furthermore, the content of the resource allows for experimentation of new approaches in the framework of the standard and new SL technologies including SL recognition, dynamic synthetic signing, machine translation and information retrieval from video sources.

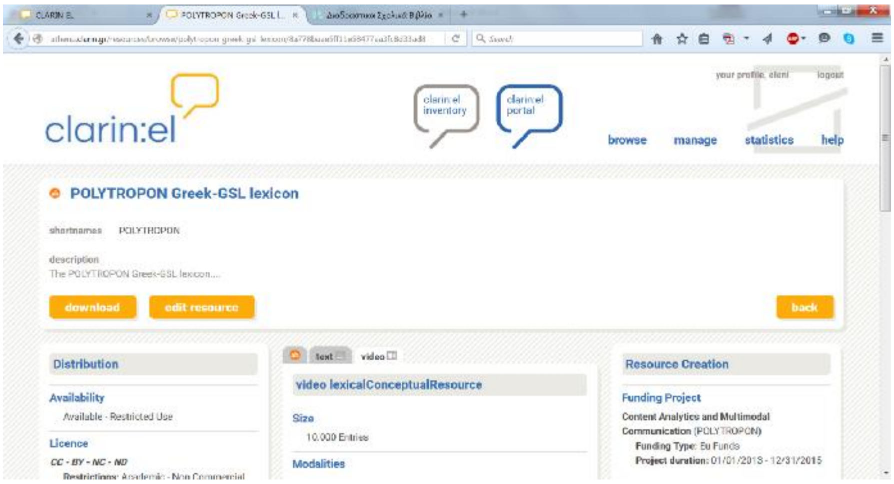


Figure 7: The POLYTROPON bilingual dictionary in the *clarin:el* repository.

In this framework, a new approach to corpus mining which is planned to be soon exploited on the basis of features relating to various parameters of sign articulation, classifier identification and features assigned to phrases as identifiers of sentence level properties, makes use of neural networks in combination to standard computer vision techniques already researched in the scope of SL technologies. Furthermore, the goal of the acquisition team is to provide the research community with a release of a golden corpus for machine learning in the areas of SL corpus mining and machine translation.

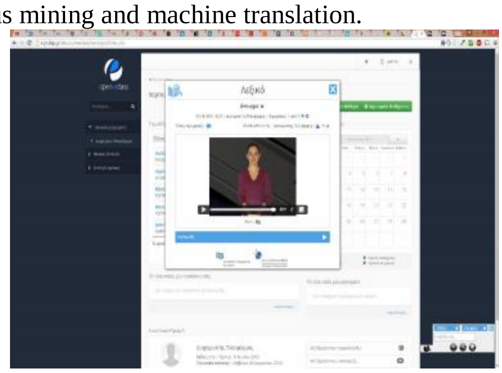


Figure 8: POLYTROPON resource use in *e-class* platform.

5. Acknowledgements

These research results have received funding from POLYTROPON project (KRIPIS-GSRT, MIS: 448306).

6. References

Dimou, A-L., Goulas, T., Efthimiou, E., Fotinea, S-E., Karioris, P., Pissaris, M., Korakakis, D., Vasilaki, K. (2014). Creation of a multipurpose sign language lexical resource: The GSL lexicon database. Proc. of *6th Workshop on the Representation and Processing of Sign Languages*, LREC 2014, Reykjavik, Iceland, pp. 37-42.

Efthimiou, E., Fotinea, S-E., Goulas, T., Kakoulidis, P. (2015). User friendly Interfaces for Sign Retrieval and Sign Synthesis. In M. Antona, M. & C. Stephanidis (Eds.). Proc. of *9th International Conference, UAHCI 2015, Part II*, held as Part of HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, LNCS vol. 9176, pp. 351-361, Springer, Heidelberg.

Efthimiou, E., Dimou, A-L., Fotinea, S-E., Goulas, T., Pissaris, M. (2014). SiS-builder: A Tool to Support Sign Synthesis. Proc. of *2nd Int'l Conference on the Use of New Technologies for Inclusive Learning*, pp. 26–36. York, UK.

Efthimiou, E. & Katsoyannou, M. (2001). Research issues on GSL: a study of vocabulary and lexicon creation. *Studies in Greek Linguistics*, Computational Linguistics 2:42-50 (in Greek).

Goulas, T., Fotinea, S-E., Efthimiou, E. and Pissaris, M. (2010). SiS-Builder: A Sign Synthesis Support Tool. In Dreuw, P. et al. (eds.), Proc. of *4th Workshop on Representation and Processing of Sign Languages*, LREC-2010, pp. 102-105.

Hanke,T. (2004). HamNoSys - representing sign language data in language resources and language processing contexts. Proc. of *1st Workshop on Representing and Processing of Sign Languages*, LREC-2004, pp. 1-6.

Hanke, T. and Storz, J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography.*Construction and Exploitation of Sign Language Corpora*. Proc. of *3rd Workshop on the Representation and Processing of Sign Languages*. ELRA, Paris, 64-67.

Klima, E., and Bellugi, U. (1979). The signs of language, Harvard University Press, USA, pp.205

Liddell, S. and Johnson, R. (1986). American Sign Language Compound Formation Processes and Phonological Remnants, In *Natural Language and Linguistic Theory,* nol.4, Reideil Publishing Co, pp.445-513.

Matthes S., Hanke T., Regen A., Storz J., Worseck S., Efthimiou E., Dimou A.-L., Braffort A., Glauert J. and Safar. E. (2012). Dicta-Sign – Building a Multilingual Sign Language Corpus. Proc. of *5th Workshop on the Representation and Processing of Sign Languages*, LREC-2012, Istanbul, Turkey.

Matthes S., Hanke T., Storz J., Efthimiou E., Dimou A-L, Karioris P., Braffort A., Choisier A., Pelhate J., Safar E. (2010). Elicitation tasks and materials designed for Dicta-Sign's multi-lingual corpus, LREC 2010, Valetta, Malta.

Phau, R., and Josep, Q., (2010), Nonmanuals: their grammatical and prosodic roles., Sign Languages, In D. Brentari (ed). 381-402. Cambridge: Cambridge University Press.

Prillwitz, S., Leven, R., Zienert, H., Hanke, T. and Henning, J. (1989). HamNoSys. Version 2.0. *Hamburg Notation System for Sign Language: An Introductory Guide*. Signum Verlag, Hamburg.

Sandler, W., Lillo – Martin, D. (2006). Sign Language and Linguistic Universals, Cambridge University Press, UK, pp.72.

Juxtaposition as a form feature – syntax captured and explained rather than assumed and modelled

Michael Filhol, Mohamed Nassime Hadjadj
CNRS–LIMSI, Université Paris-Saclay
Orsay, France
michael.filhol@limsi.fr, hadjadj@limsi.fr

Abstract

In this article, we report on a study conducted to further the design a formal grammar model (AZee), confronting it to the traditional notion of syntax along the way. The model was initiated to work as an unambiguous linguistic input for signing avatars, accounting for all simultaneous articulators while doing away with the generally assumed and separate levels of lexicon, syntax, etc. Specifically, the work presented here focused on juxtaposition in signed streams (a fundamental feature of syntax), which we propose to consider as a mere form feature, and use it as the starting point of data-driven searches for grammatical rules. The result is a tremendous progress in coverage of LSF grammar, and fairly strong evidence that our initial goal is attainable. We give concrete examples of rules, and a clear illustration of the recursive mechanics of the grammar producing LSF forms, and conclude with theoretical remarks on the AZee paradigm in terms of syntax, word/sign order and the like.

Keywords: Formal grammar, syntax, AZee

1. Production rules

As any language used productively within a community of users, a Sign Language (SL) is a linguistic system, allowing to express and interpret meaning through a set of underlying rules shared by the members, whether consciously or not. For decades now, SL researchers have taken paths looking to identify those rules, together to specify what can be called a Sign Language grammar. More than a few merely parallel schools of thought, the investigated paths are quite numerous and intertwined, forking and joining at milestone concepts. They include descriptive analyses and attempts to build predictive models, some transferring established features of general linguistic description, others rather choosing not to assume anything from other (often written) languages.

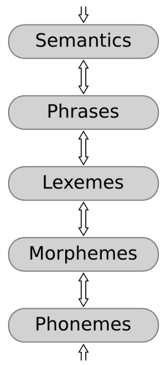


Figure 1: Stacked language construction layer scheme

1.1. AZee: purpose, principle, methodology

The general approach really took off half a century ago with Stokoe (1960). It takes on the stacked layer scheme of figure 1, where every level of language is built from an arrangement of pieces of its lower neighbour. The sensible argument to assume the validity of this scheme for Sign languages is that it was proven robust enough to be considered universal across all studied (written) languages, though it must be admitted that SLs had no fair part in the ones observed as it established.

Describing a language with this scheme implies the possible identification of the layers, each to be formalised with its own dedicated model. In earlier work, we have pointed out the general propensity to explain manual variations as syntactically driven modification to lexical units on the one hand, and to assign special roles to non-manual activity, if it is not simply overlooked, on the other hand.

Yet looking at corpus data with a global approach has revealed that a number of articulators often participate in grammatical functions jointly, whether the articulators are manual or not, and whether the grammatical function is regarded as lexical or not. Also, productive units such as those involving classifiers or complex iconic combinations are still borderline cases whose identification as lexical units versus higher-level productions—or a mix thereof—is a topic of debate, though they represent a significant proportion of the language: up to about half in the annotations of some studies (Garcia et al., 2010).

A few years ago, and initially to bypass this problem and enable synthesis without over- or ill-categorising the language objects, we proposed not to assume the layer stack and fall back on weaker linguistic hypotheses before approaching SL grammar with a formal model, namely:

- language productions are observable *forms* (states and movements of the language's articulators, e.g. "eyelids closed") carrying interpretable *functions* (interpreted purpose or meaning of the production, whether rhetoric, semantic, lexical or unidentified, e.g. "topic change" or "add pejorative judgement on person/object");
- any systematic link between the two is part of what specifies the language, and modelling it yields a rule of the grammatical system.

Such a rule is identified when either:

- an invariant, consistent form is found for many occurrences of an identified function—this raises a *production rule* that can be animated by SL synthesis software;
- a definite function can be interpreted for every occurrence of a certain form criterion—this creates an *interpretation rule*, to be triggered in SL recognition tasks.

Being mostly interested in Sign synthesis, our purpose has been to establish rules of the first kind. The methodology we use to establish those rules is a refining process, consisting in LSF corpus searches for occurrences, alternating between form and function criteria. Starting with a form or function criterion, we list the occurrences satisfying it in the data. If it is a form, give an interpretation (function) for each occurrence; if it is a function, list the form features observed in each case. In either way, group similarities in the new list and use the common features as the criterion for a new search. Alternatively, forms and functions are looked for in the data until one of the above occurs.
A few results have already been published following this approach, as well the descriptive formalism AZee used to describe rules and forms (Filhol et al., 2014). In this paper however, form descriptions will appear as box diagrams, more readable than source code.

1.2. Sequence in delevelled form descriptions

Since the beginning, we have been applying this methodology with the displayed goal of showing how much simultaneity could directly be accounted for when not considering SL streams primarily as sequences of lexical units or glosses. Looking at SL data and describing it with a delevelled approach to grammar and a holistic view of the body, we have shown that simultaneous body movements, head rotations, etc. can often be captured and formally made relevant more easily than, say, attempting to label non-lexical markers on separate levels. Given this intent, we had mostly focused on simultaneous gestures and identified production rules synchronising them.
However, the point was never to mean that sequence was no essential feature. In fact, many rules already described with AZee do produce sequences of signed parts. For example, the form description of our early-identified rule for function "open enumeration of non mutually exclusive items" contains a straight forward sequence of its argument items, only each is decorated with a head movement (see description below).

open-list a.k.a. "etc" (Filhol et al., 2010)

Function: non-exhaustive list of non mutually exclusive elements

Form: *items* signed in sequence; forward movement and retraction of the head near the end of each item of the argument list

Example: *item1* = "scissors"; *item2* = "folding knife"; full interpretation = "scissors, folding knife, and so on"

Incidentally, this rule is established without concern for the lexical or non-lexical status of its arguments. This delevelled parametrisation bears the advantage of allowing to choose them from a lexicon of one-stroke signs as well as to build complex items, which indeed have been observed in such enumerations just as well.
A second benefit of such approach, from which this whole study started, is that it allows to view the ordering of the argument items in the production (the fact that every next item in the argument list is signed *after* the previous) as a particular type of time synchronisation between them, comparable to that synchronising the head movement with each of them. In other words, production of argument forms in sequence is just another form feature available for description.
Our initial point above about simultaneity and the benefit of not reducing grammar to sequential syntax being made, it was time for us now to account for the numerous occurrences of sequences that were not captured by the searches initiated with simultaneous form features.

2. Experiment

In this paper, we propose that time precedence between two pieces of a signed production be regarded as a form feature like any other. Therefore, like one examines head nods or shoulder line rotations in search for their functional motives, one may push the delevelled/holistic approach mentioned above and apply the same methodology to examine sequences of signing chunks and describe their respective functional interpretations. As in the rest of this work, the hypothesis is that consistent appearances of observable forms are the result of intended linguistic functions to be determined, which applied to sequence might eventually account for sign and clause order in a general way.

2.1. Starting with form: juxtaposition

As our methodology goes, the starting point must be a criterion of either form or function, of which corpus occurrences must be listed. In this work, we started with the form criterion of **juxtaposition** of two distinct interpretable pieces (i.e. the beginning of the second occurring after the end of the first), and to describe the functional relationship, if any, that can be interpreted between them from that juxtaposition.
For example, the juxtaposition formed by the chunk meaning "tourist-appealing city" followed by the fingerspelt sequence D-A-H-A-B can be given the function of naming/identifying the former chunk with the latter. By contrast, the sequence of chunks "there were bombs/explosions" and "18 people died, among which several were foreigners" cannot be interpreted that way; the juxtaposition in this case is rather understood as a chronological order of events following that of the production, possibly implying causation in this case.
For each occurrence of that criterion, features of its interpreted function were then given, and sets of common features identified (to serve as the function criterion for the next iteration of the same process, only inverting function and form). After the first iteration, half of the ¡*item1*, *item2*¿ juxtapositions were interpreted as *item2* being a state—in

the most general sense of the term—for *item1*, whether its name, nature, quality or some other complex or detailed chunk of signing that would give information about it.
Following the methodology, this recurrent function feature "state/name/etc. of ... is ..." was made the pivot of the alternation and became a criterion for a new corpus search where forms would be specified in turn. Two patterns emerged involving chin/eyebrow positions and eye blinks, which triggered more iterations of function-to-form and form-to-function searches.

2.2. Results on juxtaposition

The exact statistics for every iteration observed along this corpus study are soon to be published elsewhere, but pulling the thread initiated with the "state" criterion led us to refine it into three stable function-to-form matches, which we summarise below:

category a.k.a. "cat"

- Function: *item2* is to be understood as the hyponym of *item1*
- Form (fig. 2.2.a): chin and/or eyebrow raise on the beginning of *item1*; minimal transition time (approx. 100 ms) between *item1* and *item2*
- Example: *item1* = "country"; *item2* = "Montenegro"; combined interpretation = "Montenegro"

add-info a.k.a. "()"

- Function: *item1* is given the additional side information *item2*
- Form (fig. 2.2.b): chin and/or eyebrow raise on the beginning of *item2*; minimal transition time
- Example: *item1* = "town/city"; *item2* = "the power is tourism"; combined interpretation = "tourist-appealing city"

info-about a.k.a. ":"

- Function: *item2* is the point being made about *item1*
- Form (fig. 2.2.c): longer transition time (˜340 ms); eye blink at the end of *item2*
- Example: *item1* = "power"; *item2* = "tourism"; combined interpretation = "the strength/power is tourism"

Aside from the "state" function leading to the rules above, other functions were interpreted from the occurrences of juxtaposition after the first search, of which we give a few recurrent examples below. However, those need to be refined with a similar process since no further iteration was carried out over them. In other words, while the ones above are consistently observed, the ones below are still hypothetical and need to be studied in more depth by means of

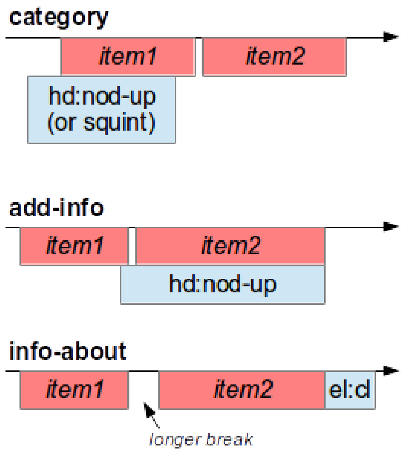


Figure 2: AZee box diagrams for parameterised forms of functions (a) *category*, (b) *add-info*, (c) *info-about*

more iterations of the methodology. The title functions may merge or be split according to the further observations that will be made.

context a.k.a. "ctxt"

- Candidate function: *item1* is the time or space where *item2* is true or taking place
- Example: *item1* = "today, May 3"; *item2* = "presidential elections"; full interpretation = "there are presidential elections today, May 3"
- NB: The functional distinction was difficult to make with structures where *item1* anchors an entity in signing space, subsequently referenced through their location anchor in *item2*. We therefore hypothesise that these two operations may be covered by the same function.

loc-from-ref a.k.a. "loc"

- Candidate function: *item2* is located with respect to *item1* (in time, space or some other projection)
- Example: *item1* = placement of Egypt; *item2* = placement of the Red Sea nearby; full interpretation = "the Red Sea, which is near Egypt"
- NB: Manual forms appear on the non-dominant hand, in addition to the juxtaposition taking place.

closed-list a.k.a. "and"

- Candidate function: exhaustive list of elements with equal part
- Example: *item1* = A; *item2* = B; *item3* = C; full interpretation = "A, B and C"

finger-spell a.k.a. "fs"

Function:	name a place, person or concept by spelling its written name letter by letter
Example:	*item1* = "D"; *item2* = "A"; *item3* = "H"; *item4* = "A"; *item5* = "B"; full interpretation = "Dahab"
NB:	This is a rather obvious example of sequence, but it does not come alone. Eye blink before start and eye gaze towards the spelling hand at start seem almost compulsory in our data.

2.3. Elaborating on results

As partly already illustrated in the instances given for add-info and info-about, we find that these rules seem to nest very well, both combining the functions into a correct interpretation and adding up the form features imposed by the rules involved.
For example, the forms we observe for the whole chunk "tourist-appealing city Dahab" are given in figure 3, and are equal to those that would have been combined programmatically from:

1. figure 2.2.c, functionally giving the information "tourism" about the topic "strength", which builds a chunk meaning "the strength is tourism";
2. figure 2.2.b, functionally using the chunk above "[its] strength is tourism" as side information to "city", which builds a new chunk meaning "tourist-appealing city";
3. figure 2.2.a, functionally making this new chunk a category for fingerspelt chunk identifying the city of Dahab.

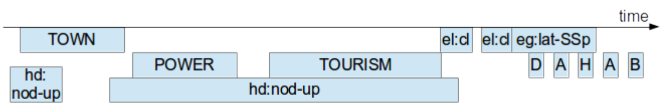


Figure 3: Time score of form features for "tourist-appealing city Dahab"

The combination of rules using recursively built chunks from rules of the same set can easily be represented in a tree diagram, where nodes are rule functions, children are rule arguments and child node order is that of the juxtaposed arguments named *item1*, *item2*... in the descriptions above. The tree corresponding to our example is given in figure 4. This observation about recursion in rule use is consistent with our earlier results on expressions of time sequences and durations, and supports the underlying AZee hypothesis of a recursive system of nestable rules.
Besides, more rules, already published from earlier work, had form descriptions with juxtaposition as the primary form feature, if not the only one documented for the moment. We have already mentioned open-list (§ 1.2.); here are two more examples:

chrono-sequence a.k.a. "seq" (Filhol et al., 2015)

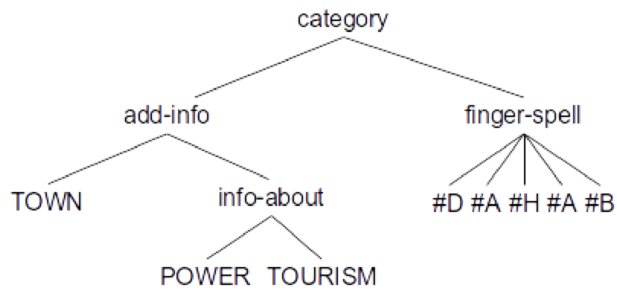


Figure 4: AZee function tree for "tourist-appealing city Dahab"

Function:	*item_i+1* took place after *item_i* chronologically
Form:	juxtaposition; more TBD
Example:	*item1* = "there were explosions"; *item2* = "people died"; full interpretation = "people died after the explosions"

option-list a.k.a. "either-or" (Filhol et al., 2010)

Function:	non-exhaustive list of non-mutually exclusive elements
Form:	items juxtaposed; head rest in a new location on each *item* of the argument list
Example:	*item1* = "alone"; *item2* = "friends"; *item3* = "family"; full interpretation = "either alone, with friends or with a family"

As the number of established rules grows and by conjoining them into a rule set and generally allowing nesting of one another, we are gradually forming a recursive grammar for LSF, without appealing to preconceived notions like syntax or lexicon.

3. Advantages of the forming grammar

Such delevelled grammar is difficult to situate among the traditional theories, but does provide a few advantages which we underline in this section. They will be useful to discuss the comparison with traditional syntax in the following section.

3.1. Coverage and productivity

Following our methodology, one can at best positively identify function-to-form rules (production rules) and appreciate them for their semantics and assess their coverage. But in view of building a full formal LSF grammar, one can only hope for the existence of a greater system, capturing the whole of the language and of which the identified rules would all be part. Until every corpus discourse becomes describable as a tree of functional nodes taken from an identified set of rules, the question remains of whether such a set exists, and if so what size it might grow into. Is it in reasonable reach at all?
At the beginning of our function-to-form rule search approach and after the first few rules had surfaced, we believed that in addition to those describing the more or less

fixed forms typically listed in dictionaries (the so-called "signs"), hundreds of production rules might be needed given the fine-grain semantic nature the functions we had identified: "open list of non mutually exclusive items", "event separation time exceeding a fortnight", etc.
However, using all identified rules from prior results and from this new study, we have tried to build the trees representing the recursive juxtapositions in 7 videos of 30 seconds each. We have found that out of 220 juxtapositions, only 21 remained unexplained, i.e. about 3 per 30-second discourse. Incidentally, many of them were cases of discourse continuation with hands retracted in a long pause and no semantically or rhetorically loaded function interpretable with consistency. Therefore with only half a dozen extra rules, the new state of our grammar accounts for most occurrences of juxtaposition in our corpus. In other words, it already captures most of what is otherwise called syntax.
The exact total number of production rules governing LSF sequence (if any such number exists) may fluctuate according to the various refinements still needed. But if we assume that the refinement searches to come will not break our tentative functions up into big numbers of finer rules ones, our study allows to hypothesise that it should hardly exceed 20 (this would already double the current count), plus dictionary signs.
With this study, our approach to grammar has taken a huge leap forward regarding coverage. At the same time, the resulting rule count ratio has turned from a somewhat alarming figure to an appealing one.

3.2. Expressiveness and precision

We see that the AZee function-to-form rule approach easily accounts for sequences generally considered syntactic. However, the system does not consider sequence—let alone lexical sequence—itself as the primary form of sentence articulation, nor are rules restricted to describe such sequences. It is *one of the many* describable forms, and quite importantly, rules will:

- not only constrain "before"/"after" but include any relevant timing indications;
- contain all other necessary markers directly in their form description.

For example, the add-info rule was observed:

- not only to have the 2nd argument signed "syntactically after" the first argument, but also to constrain the time duration in between (approx. ~100 ms, i.e. a quick transition);
- to impose a chin lift (or slight squint) starting immediately before and carrying over onto the 2nd argument.

Whereas the first feature is mostly regarded or discarded as "prosody", we observe enough consistency and bond to semantically categorisable functions to keep them in the grammatical descriptions where relevant. There is no reason here to discard transition lengths from the descriptions if they are as systematic as the other synchronised forms. For the same kind of reason, the second feature saves us from justifying grammatical markers on different levels.

This to us makes rules very expressive and brings a lot of precision to the model, for our original goal of computer synthesis but also in linguistic terms. When animating an avatar, if a form feature is consistently observed for a given function in the language data, it is expected to be rendered in the resulting animation, hence it is crucial to know about it and efficient to include it in the form description directly. Linguistically, considering transition durations as just another form allows to parse an input discourse with this information.
A significantly greater precision is thereby achieved when processing sign streams. From the traditional point of view of gloss sequences, utterances like:

> TOWN STRENGTH TOURISM

are ambiguous since nothing differentiates between meanings "strong town called [Tourism]" and "tourist-appealing city". With our holistic approach and an AZee grammar, they are directly distinguishable by the better-informed rules, hence the input to parsing process is not considered ambiguous. All so-called "non-manual" or "prosodic" markers will in fact play their equal part in the process directly, instead of being looked for afterwards as a means of resolving the ambiguity, which calls for yet a different system to model.

4. Discussion on "syntax"

Now that the philosophy of our approach has been clarified and a few benefits explained, this section addresses the tricky comparison with traditional syntax, defined as the paradigm governing word order or, for SL, sign sequences.

4.1. An alternative definition for syntax?

Starting from a simple search for function–form mappings, we end up with a set of production rules for every repeated synchronisation of a form feature set, possibly parametrised with arguments. By design, they are not tied to any labels in terms of the traditionally distinct levels of language construction (lexicon, syntax, etc.). However, the linguistic tradition induces a strong intuition of them and a tendency to assign one and one only to every rule. While this may not be needed as such for Sign synthesis from AZee, one might still have interest in such categorisation.
Since the levels were not assumed before building the rules, and as the rules are formally specified with an unambiguous description system, formal criteria can be expressed to characterise the levels formally from AZee. To do so, one can express binary conditions on the rules' form features to group those that intuitively pertain to the same level.
To characterise syntax for example, inspired by the original idea of it governing the order of components, we would suggest that be called syntactic:

> a rule that accepts at least two mandatory arguments, and whose arguments are juxtaposed in the form description.

This definition is interesting in the sense that it does not depend on the notion of lexical unit, which is itself can be problematic to define (see § 1.1.). This raises interesting prospects regarding known borderline cases such as "classifier predicates" (Cogill-Koez, 2000) or "partly-" or "non-lexical signs" (Johnston and Schembri, 1999). Providing AZee rules for such constructions and checking them against the proposed criteria should feed the discussions on their debated statuses on the basis of formal and purely data-driven arguments.

4.2. Syntax paradox

The last section investigated how AZee encompasses the traditional definitions above, and suggested that the effort would assist clarifying the lines between historical categories. The present section now takes a somewhat reversed point of view, and addresses the question of whether it is relevant at all in the AZee framework in return, specifically in the case of syntax.

A problem when categorising AZee rules as syntactic or non-syntactic on the basis of presence or absence of argument juxtaposition does not acknowledge the fact that no AZee rule exists in the first place if no meaningful interpretation can be made of it. This is quite opposite to the traditionally accepted profile of a syntactic rule. Syntax normally organises argument units according to their respective morphosyntactic categories and regardless of the semantic roles that they may take, and is what fills the gap between lexical units and the semantic relations between them. Contrarily, AZee rules with arguments will necessarily bear some semantic relationship between them directly, otherwise they will simply not exist. By construction, meaningless arrangements of of rules. So firstly, syntax cannot be defined as a category of rules kept clear of semantics by fear that it would be escaping its field. Such criterion would be paradoxical in AZee.

It is only if we accept that a rule might satisfy more than one level-defining criterion like that of section 4.1., that a category of syntactic rules can be identified. But such category does not lie as a necessary level bridging an open gap between two others, so the new question then is what the purpose of it would be. As noted in section 3.2., AZee rules can organise the form features in a variety of ways, including simultaneous features and differences in intervals even between non-simultaneous (juxtaposed) items. So juxtaposition taken on its own as a criterion for rule categorisation does not appear as a specifically salient feature. The relevance of a category of rules only interested in what is sequentially ordered (i.e. syntax) is therefore to be questioned.

In short, we have defined a grammar model without assuming a distinct syntactic organisation of the utterances, then given a criterion to create a syntactic category nonetheless, only by deriving it from the model rather than assuming it to build the model. We concluded that whereas the AZee criterion might have some relevance to clarify the traditional notions when applied to Sign Language, the category itself had little purpose in the AZee paradigm.

5. References

Cogill-Koez, D. (2000). Signed language classifier predicates: Linguistic structures or schematic visual representation? *Sign Language & Linguistics*, 3(2):153–207.

Filhol, M., Braffot, A., and Matthes, S. (2010). Dictasign deliverable d4.2: Sentence descriptions for bsl, dgs, gsl, lsf. DictaSign project, EU-FP7.

Filhol, M., Hadjadj, M. N., and Choisier, A. (2014). Non-manual features: the right to indifference. In *Representation and Processing of Sign Languages: Beyond the manual channel, Language resource and evaluation conference (LREC)*, Reykjavik, Iceland.

Filhol, M., Hadjadj, M. N., and Testu, B. (2015). A rule triggering system for automatic text-to-sign translation. *Universal access in the information society (UAIS), C. Stephanidis (Ed.), special issue on "Recent Advances in Sign Language Translation and Avatar Technology", R. Wolfe, E. Efthimiou, J. Glauert, T. Hanke, J. McDonald, J. Schnepp (guest eds.), Springer.*, 15/4.

Garcia, B., L'Huillier, M.-T., Frémeaux, A., and Makouke, D. (2010). Discursive and morphological processes in the formation and the stabilisation of lexematic units in french sign language. In *Theoretical Issues in Sign Language Research (TISLR)*, UCL, London.

Huenerfauth, M. (2006). *Generating American Sign Language classifier predicates for English-to-ASL machine translation*. Ph.D. thesis, University of Pennsylvania.

Johnston, T. and Ferrara, L. (2012). Lexicalization in signed languages: when is an idiom not an idiom? In *Proceedings the 3rd UK Cognitive Linguistics Conference, vol. 1*, University of Hertfordshire.

Johnston, T. and Schembri, A. (1999). On defining lexeme in a signed language. *Sign Language & Linguistics*, 2(2).

Liddell, S. (2003). *Grammar, gesture and meaning in American Sign Language*. Cambridge University Press.

Marshall, I. and Sáfár, E. (2004). Sign language generation in an ale hpsg. In *Proceedings of the HPSG04 conference*.

Stokoe, W. (1960). Sign language structure: an outline of the visual communication system of the american deaf. *Studies in linguistics, occasional papers*, 8.

Examining Variation in the Absence of a 'Main' ASL Corpus: The Case of the Philadelphia Signs Project

Jami N. Fisher[1], Julie A. Hochgesang[2], Meredith Tamminga[1]
University of Pennsylvania[1] and Gallaudet University[2]
E-mail: jami@sas.upenn.edu, julie.hochgesang@gallaudet.edu, tamminga@ling.upenn.edu

Abstract

The Philadelphia Signs Project emerged from the community's desire to document their local ASL variety, originating at the Pennsylvania School for the Deaf. This variety is anecdotally reported to be notably different from other ASL varieties. This project is founded upon the consistent observations of this marked difference. We aim to uncover what, if anything, makes the Philadelphia variety distinct from other varieties in the United States. Beyond some lexical items, it is unknown what linguistic features mark this variety as "different." Comparison to other ASL varieties is difficult given the absence of a main and representative ASL corpus. This paper describes our sociolinguistic data collection methods, annotation procedures, and archiving approach. We summarize several preliminary observations about potentially dialect-specific features beyond the lexicon, such as unusual phonological alternations and word orders. Finally, we outline our plans to test these features with surveys for non-Philadelphians using Philadelphia lexical items, extending to more abstract phonological and syntactic features. This line of inquiry supplements our current archiving practices, facilitating comparison with a main corpus in the future. We maintain that even without a main corpus for comparison, it is essential to document a language variety when the community wishes to preserve it.

Keywords: Language documentation and long-term accessibility for sign language data, experiences in building sign language corpora, sociolinguistics of signed languages, ASL, Philadelphia ASL, language variation

1. Introduction

This paper introduces the Philadelphia Signs Project, which emerges from the Philadelphia Deaf community's desire to document their local variety of ASL. Beyond some stereotypical lexical items, it is not known what linguistic features give rise to the frequent evaluation of this variety as "different." Comparing Philadelphia ASL to other ASL varieties may be difficult given the absence of a main ASL corpus that is representative of the overall language in North America but we maintain that it is still possible. This paper describes our sociolinguistic data collection methods, our annotation procedures, and our archiving approach. We summarize several preliminary observations about potentially dialect-specific features beyond the lexicon, such as unusual phonological alternations and word orders. Finally, we outline our plans for testing whether these features are actually unique to Philadelphia in order to facilitate comparison with a main corpus when it is ready. We maintain that even without a main corpus for comparison, it is essential to document a language variety when the community itself wishes to preserve it.

2. Background and Motivations

Philadelphia ASL emerged from the residential school context at Pennsylvania School for the Deaf, where deaf Philadelphians used the language brought by Laurent Clerc and his disciples from the first deaf school in Hartford, Connecticut. The PSD residential school campus was the major site for sign language transmission in the Philadelphia area until 1984, when it was closed and moved to a day-school setting in another part of the city. As at other deaf schools around the country, the signs at the Pennsylvania School for the Deaf began to look different over time. The variety that emerged, still currently in existence among older living signers, attracts attention in the Deaf community for being "different" and "strange." Beyond some stereotypical lexical items, it is not known what specific linguistic features give rise to such evaluations. The goal of the Philadelphia Signs Project is to uncover what, if anything, makes Philadelphia ASL distinct from other varieties in the United States. This goal is similar to that of the "Black ASL" project undertaken by McCaskill et al (2011). As described in Hill (2012), Deaf Americans could see a particular style of signing and be able to identify it as "Black ASL" but were not quite sure why. The "Black ASL" project strived to ascertain the linguistic features that marked this particular ASL variety. They came up with the following list: handedness, lowering, size of signing space, incorporation of African American English (AAE), use of repetition, use of role shifting, amount of mouthing, and lexical differences.

Since the 1984 closure of the PSD residential campus, younger speakers in the Philadelphia area have had significantly reduced exposure to native signers of Philadelphia ASL, presumably leading to a leveling of this variety toward a pan-regional variety influenced by Gallaudet. This leveling is actively reflected on by one of the participants, Colleen, an early-thirties female who has two Deaf parents and a Deaf brother. Colleen attended PSD day school until she was high-school age. She then transferred to Model Secondary School for the Deaf in Washington, D.C. and remained there to continue her postsecondary and graduate studies at Gallaudet University. Below we show a brief exchange between the interviewer and Colleen, demonstrating both her exposure to and awareness of the Philadelphian ASL variety as well as the influence of the Gallaudet, pan-regional variety on her own sign production.

Interviewer: *Tell me about your experience about transferring from PSD to MSSD and how that influenced your signing. Was it different? Did Gallaudet change your signing?*
Colleen: *Oh yes, there was definitely a difference.*
Interviewer: *Tell me about it.*
Colleen: *Well, for example, I still say "inch" like this* (see figure 1 which shows the ASL sign that is not recognized by other members of the ASL community at Gallaudet, only members of her local community in Philadelphia.).

Figure 1. Colleen signing "inch"

Colleen: *I sign it like this* (Figure 1). *But when I was at MSSD, if I tried signing it like this, others would not understand me. So I'd have to express the concept in a different way. They even told me their sign. I can't remember what it looks like. But it doesn't matter because I like my version of "inch". And I've stayed with the same version all this time. Although when I'm at Gallaudet, I'll just fingerspell "inch". But I still hold everything - all of the old signs. And when I return home, I use them… For example, for some months of the year, I'll produce our signs for them…. There are a lot of signs like that. Our sign for "eagle", for example… I catch myself changing. At home, I'll sign our version of "eagle"* (Figure 2, left) *but at MSSD, I'll sign their version of "eagle"* (Figure 2, right).

Figure 2. Philadelphia ASL variant for "eagle" (left); other ASL variant for "eagle" (right)

Colleen: *Why I change is because I think they won't understand me if I sign that way so I hold my "eagle" sign and remember that it's Philadelphian. I'll use the other ASL variant for "eagle" out there but when I go home I'll use the Philadelphian one. I go back and forth.*

2.1 Pressures towards Leveling

Since Philadelphia ASL is still in existence but on the verge of being lost, we have a time-sensitive opportunity to collect data documenting this example of regional variation within ASL. Doing so will allow us to identify features that are characteristic of Philadelphia ASL and to trace the loss of such features generationally under pressures toward leveling. We see evidence of these pressures in the Philadelphia community at large, within the interviews, and even in some of the self-reflective comments of the primary interviewer himself. One of the authors, a native Philadelphian and hearing native signer, has encountered many members of the Philadelphia Deaf community who lament the fact that their variety is dying out with the older members of the Deaf community. In turn, many have expressed appreciation of these efforts to document their variant.

There is also evidence that many older signers recognize diachronic change in younger generations, yet maintain their "old" Philadelphia signs despite seeing these changes around them; several of the interviewees comment--both in the interviews and in casual conversation before and after interviews--that they identify themselves as "using old (Philadelphia) signs," with the implication that their signing is markedly different from younger generations of Philadelphians. When one participant, Caroline, was asked by the interviewer if she "understood clearly" what one of the authors had signed to her in a conversation before the interview, she confirmed, but immediately distinguished her own signing from the author's by saying she, herself, does not use "new" signs but instead uses "old" ones and they are "hers."

One final example of the leveling pressure comes from the primary interviewer himself. Outside of the interviews, he has repeatedly referred to the noticeable decline of his own use of the Philadelphia variant, attributing the diminution of its use to the fact that he socializes with people from all over the country and travels to meet many Deaf people around the world. The distinctions between "old" and "new" Philadelphia variants are clearly noted by the community, as is their disappearance with time. Thus, documentation before complete disappearance is essential.

While it is likely that some of the perceived differences in the Philadelphia Deaf community are a result of age, we suggest that this variety is not only age-related. First, we see indications of intergenerational use of the Philadelphia variety. This is evidenced by some of the interviews we have done with families, as per one participant, Colleen's, reflections detailed earlier. In addition, anecdotes from older Deaf community members not from Philadelphia repeatedly point to the "strangeness" of the ASL sign productions of the

Philadelphia Deaf community as compared with their own ASL productions.

2.2 Beginning the Philadelphia Signs Project

This project has stakeholders of various motivations and has the potential to be applied to multiple ends. We started this project, in part, as a direct response to the Philadelphia Deaf community's longtime call to document the familiar language that they see changing and disappearing. As signed language researchers, we also recognize the significant void in publicly available ASL data in video form. We thus set out to video-record Philadelphia-native, Deaf signers conversing naturally in order to create a public, searchable, web-based corpus that will benefit Deaf community members and the research community alike. While this project has obvious benefactors in the field of signed language studies, we also anticipate that local organizations that serve the Deaf and Hard-of-Hearing communities will benefit from the existence of such a resource. For example, the documentation of the local variety will be of use to those who train interpreters in Philadelphia, while current and future generations of PSD students may find the narratives related by our participants to be of historical and cultural interest.

We have designed this project from the beginning with the understanding that these data will be public, and thus have taken great care to ensure that our participants are fully informed of the reality of such public accessibility. Furthermore, we would like to emphasize that we have been deliberate in attending to and including Deaf community members and organizations throughout this process.

3. Data Collection and Organization

The data collection procedure, in line with previous work in ASL sociolinguistics (Lucas, Bayley, & Valli, 2001), is modeled on the sociolinguistic interview methods laid out by Labov (1984). The interview questions are aimed at eliciting stories about the signer's lived experiences, with thematically similar questions grouped into modules. The interviewer, who is a native Philadelphia Deaf signer with strong ties to the local Deaf community and institutions, guides the conversation but allows the person being interviewed considerable conversational latitude in order to facilitate naturalistic conversational signing. At the end of the interview, the interviewer presents two more structured elicitation tasks. The first is the elicitation of known Philadelphia-specific lexical items using a picture-naming task. The second is a narrative description task in which the participant retells a story from a cartoon clip. The entire interview is video recorded from two perspectives: one capturing a frontal view of the participant, and the other capturing both the participant and the interviewer simultaneously.

The data have been organized in a cloud-based service that is accessible to the current research team following the basic archival format as described in Himmelmann (2006); that is, the data are consistently organized into bundles of "primary data" (the video sessions themselves) and "apparatus" (annotations, metadata, general access resources like the annotation conventions and project proposal). Currently the language documentation is temporarily archived on Google Drive but will eventually be hosted online in a way that the data can be searched and shared (e.g., using online language archives like other signed language documentation projects have used).

Currently, we have interview and elicitation data from about 25 Deaf Philadelphians. We have already started processing the data by annotating them in ELAN (Wittenburg et al, 2006). Current data processing efforts focus on partial annotation of our participants in each video session, specifically adding ID glosses for individual signs and free translation. Such efforts (ID glossing and translation) are considered to be the absolute minimum required to make primary data accessible (Himmelmann, 2006; Johnston, 2008). Figure 3 shows our current tier structure for our initial and minimal annotation.

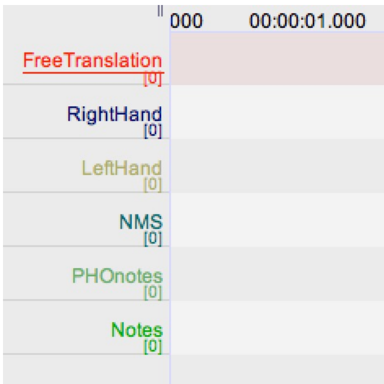


Figure 3. Tiers in our ELAN template

For annotation of ID glosses using the "RightHand", "LeftHand" and "NMS" tiers, we are using the SLAAASh annotation conventions (Hochgesang, 2015) as well as their ID gloss list to ensure consistency both within our data and potential comparison with other ASL documentations. "Free Translation" is used to give a loose English translations of the ASL utterances produced.

The "Notes" tier (a separate one is created for each annotator that adds content to the transcript) is used for general comments, questions about the signs produced on the video, feedback on the annotation itself, and so on. The "PHOnotes" tier is used by the primary annotator to provide observations on any interesting phonetic or phonological phenomena.

4. Examining Regional Variation in ASL

Regional variation has been documented across many signed languages, although such documentation is often limited to the observation of distinct lexical items (see Schembri & Johnston 2012 for an overview). The picture-naming task in the documentation we are creating is aimed at documenting uniquely Philadelphian ASL signs and providing empirical evidence on the extent to which such signs are attested across different generations. Because regional variability is the norm for signed languages, though, we hypothesize that the perception of the Philadelphia variety as unusual arises from deeper linguistic differences. Prior work on phonological variation in ASL has detected regional differences, but typically such differences have involved the quantitative preference for widely available phonological variants, rather than qualitative distinctions between varieties (Bayley, Lucas & Rose, 2002). Research on regional differences in syntax is even rarer; we are not aware of such work on ASL. Preliminary observations of the first interviews to be annotated, though, have presented a number of candidate examples of features in the ASL of the oldest Philadelphians that may be regionally unique.

As our data represent a diverse cross-section of the Philadelphia Deaf community including different generations of Deaf Philadelphians from the same families, we are potentially able to use sociolinguistic variables to examine the ways in which the variant demonstrates change over time through contact with signers of other variants including what we call a pan-regional ASL.

4.1 Potential Examples of Variation in Philadelphia ASL

As expected, there are several lexical variants that appear to be unique to Philadelphia ASL and have already been added to the ASL ID gloss list shared with other research teams. Figure 4 shows three ASL variants for "woman", two of which are used by the Philadelphian signers.

Figure 4. Three ASL variants for "woman"
(the top two by a Philadelphia signer;
the bottom is from the general ID gloss list)

As explained earlier, the "Notes" and "PHOnotes" tiers were opportunities for our annotators to record initial and casual observations about the language use and production. Many of those memos could be considered normal for signed language documentations (e.g., "assimilation occurring here"; "left hand is lower than right hand"; "subject has been dropped"). But there are also quite a few annotator observations that were unexpected and may point to potential variation specific to the Philadelphia variety.

<u>Annotator comments from "PHOnotes" or "notes" tiers</u>

- *PROBLEM is produced one-handed here when it's supposed to be two-handed and I wouldn't accept it as one-handed especially in this situation (the other hand is not occupied by something).*
- *HAPPEN is produced one-handed!*
- *LEARN. Morphologically interesting, may be that the movement towards the end marks aspect (perfective)*
- *INEPT. Different direction from ID gloss form*
- *USE. Interesting movement and orientation here.*
- *PEOPLE. One-handed!*
- *WORKSHOP. Note HC of second part (no "S" as expected at the end of the sign)*
- *FS(of). Wow, her pinky is extended before she even starts the sign. It's not unusual to see pinky extension because of the extension of all fingers in the LETTER-F but this is even before the fingerspelling starts. Whoa.*
- *FOURTEEN. Unusual repetition.*
- *HOME. Note how the second placement is very near to the first placement.*
- *HOME. Only one contacting hold. Usually two.*
- *FS(burrough). Something interesting about how she fingerspells - I can't quite put my finger on it. Age? Possible arthritis making joint movement stiff? Accent?*
- *MANY. Unusual orientation if this is indeed MANY.*
- *EMPTY. Signs this on top of a buoy!*
- *A lot of lip puckering during her responses or feedback*
- *ALL. I love this production. I'd say it looks "proper".*
- *THAT. I would have said that this is my "citation form" (the form I imagine when I think of this sign) but watching him produce it I realize I think this is a bit archaic, signing with the active hand contacting the palm of the weak hand. Seems more typical to do one-handed these days.*
- *TWO-YEARS. Interesting! This is numeral incorporation - number morpheme incorporated (is visible throughout the entire sign) with the rest of the sign (movement, location, etc). This is one possible grammatical variation. What number range OR word can take numeral incorporation seems to vary (geographically, etc).*
- *FS(then) Interesting production, it's almost signed with the same type of movement you'd expect in NEXT*
- *Unusual orientation in production of numbers. For example, the age "105" is usually produced with the hand oriented away from the signer for all of the numerals. However, this is "one hundred" with the hand facing forward and then turned inward to produce the number "five".*

Generally the informal observations as listed above can be categorized as referring to unexpected phonetic forms (handedness, repetition, path, orientation), phonetic alternations, and morphological processes. It has also been noted by some of the research team members that some of the older Deaf Philadelphian adults use a much larger signing space than younger signers and a few have unexpected syntactic constructions (particularly with ordering of constituents and pronoun dropping).

Because many of the signers being interviewed for this project are quite elderly, any distinct features we detect might either be unique to Philadelphia or be characteristic of older forms of ASL pan-regionally. While other collections of ASL video recordings exist, such as that of Lucas, Bayley & Valli (2001) which are currently archived at the Gallaudet Video Library (http://bit.ly/1PpjDaz, last accessed March 2016), none have been annotated (or annotations, if any, have not yet been made available) or set up to facilitate access to the primary data. This puts ASL in the position of lacking a representative corpus comparable to those of other national signed languages, such as German Sign Language ("DGS-Korpus," 2014), Australian Sign Language ("The Auslan Corpus," 2014), British Sign Language (Schembri, 2008), and Netherlands Sign Language (Crasborn, Zwitserlood, & Ros, 2014).

4.2 Current Proposed Methodology of Examining Variety in Philadelphia ASL

We propose to adopt the following four methodological approaches in order to both document Philadelphia ASL as a variety in its own right and also lay the groundwork for comparison to a pan-regional ASL corpus at some point in the future. First, we will continue to document explicit observations about Philadelphia ASL as we have been doing on our "Notes" and "PHOnotes" tiers and may potentially develop annotation codes similar to other signed language corpus projects. These observations include meta-linguistic commentary from members of the Philadelphia Deaf community as well as observations about how features of Philadelphia ASL differ from the expectations of the signing researchers involved with the project.

Second, we are collecting sociodemographic information in the form of a background questionnaire about the participants, allowing us to interpret differences between participants in light of known sociolinguistic generalizations (Lucas, 1989; Schembri & Lucas, 2015, Morris, 2016). For example, the language use of participants who have lived or studied outside of Philadelphia is expected to show greater assimilation to a pan-regional variety of ASL as demonstrated in the earlier exchange between our project interviewer and Colleen.

Third, we are preparing an experimental paradigm for eliciting acceptability and familiarity judgments on lexical items through an online platform that allows us to use both English and ASL in order to make sure it is accessible. A mock-up of one potential component of this survey is shown in Figure 4. The online survey will enable us to assess whether Philadelphia signs are recognized more widely in the US Deaf community; future work will

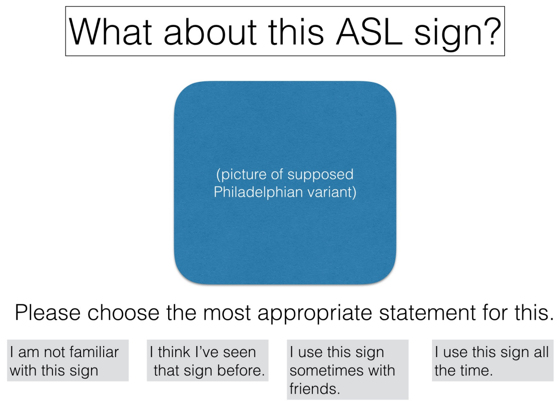


Figure 4. Potential design for online survey

extend the paradigm to acceptability judgments on phonetic, phonological, morphological, and syntactic features. We will also refer to similar work done by other signed language researchers for strategies that can be used to prevent scam contributions (e.g., the German Sign Language (DGS) Corpus Feedback Portal).

Finally, we emphasize the importance of transparent archiving methods (e.g., Bird and Simons 2003; Himmelmann 2006) and use of widely used, well-documented annotation conventions (i.e., the conventions listed in Hochgesang, 2015) which are currently used by a few research projects in the US and have been shared with and reviewed by other signed language researchers at the international workshop *Digging into Signs* (Crasborn, Bank & Cormier, 2015). This application of archiving and annotation best practices (both in general and specific to the study of signed languages) to our language documentation will facilitate eventual comparison to corpus data from other ASL varieties.

5. Long-Term Plans

The documentation of regional variation in ASL is an area of research that offers scientific progress in the study of language change in signed languages but also will directly benefit the Deaf community. The long-term aim of this project is to create a corpus of 100 interviews from the Philadelphia Deaf community, building on the preliminary work discussed here. We will survey a balanced sample of individuals with particular attention to capturing the full spectrum of age-related differences to document language change, as regularities in linguistic variability are detectable only with sufficiently large amounts of data from many different language users. The final product, including all annotations, will be made publicly available online; our hope is that it may serve as a model for natural signing in the local variety for training of translators and signed language education as well as preserve the linguistic heritage of the Philadelphia Deaf community for future generations.

6. Acknowledgements

This research was supported by a Research Opportunity Grant from the University of Pennsylvania School of Arts and Sciences.

We are grateful for the work of our research assistants: Heather Hamilton (Gallaudet University), Carmen Lugo (University of Pennsylvania), and Elana Chapman (University of Pennsylvania).

7. Main References

Bayley, R., Lucas, C., & Rose, M. (2002). Phonological variation in American Sign Language: The case of 1 handshape. *Language Variation and Change*, 14(01), 19-53.

Bird, S. & G. Simons. (2003). Seven dimensions of portability for language documentation and description. *Language*, 79(3). 557-582.

Crasborn, O., Bank, R., & Cormier, K. (2015). *Digging into Signs: Towards a Gloss Annotation Standard for Sign Language Corpora.* MS. Radboud University. Retrieved Mar 21, 2016, from http://bit.ly/1MkLOwM

Crasborn, O., Zwitserlood, I., & Ros, J. (2014). *Corpus NGT (Corpus)*. Retrieved November 3, 2013, from Creative Commons Attributions 3.0 Netherlands

DGS-Korpus. (2014). Digital corpus. Retrieved August 8, 2014, from http://bit.ly/1T5l70H

Hill, J. (2012). Black ASL. *ASLized! Journal of American Sign Language and Literature*, 2. Retrieved March 21, 2016 from http://aslized.org/journal/black-asl

Himmelmann, Nikolaus. (2006). Language documentation: What is it and what is it good for? In J. Gippert, N.P. Himmelmann, U. Mosel (Eds.), *Essentials of Language Documentation*. New York: Mouton de Gruyter, pp. 1--30

Hochgesang, J. (2015). *SLAAASh ID Glossing Principles and Annotation Conventions*. Ms., Gallaudet University and Haskins Laboratories. http://bit.ly/1NpOB1u

Labov, W. (1984). Field methods of the project on linguistic change and variation. In J. Baugh & J. Sherzer *Language in Use: Readings in Sociolinguistics.* Englewood Cliffs, NJ: Prentice-Hall, pp. 28--53.

Lucas, C., Bayley, R., & Valli, C. (2001). *Sociolinguistic Variation in American Sign Language*. Washington, D.C.: Gallaudet University Press.

Lucas, C., & Valli, C. (1989). *American Deaf Community. The Sociolinguistics of the Deaf Community*, 11. San Diego, CA: Academic Press.

McCaskill, C., C. Lucas, R. Bayley, and J. Hill. (2011). *The Hidden Treasure of Black ASL: Its History and Structure.* Washington, DC: Gallaudet University Press.

Morris, C.D. (2016). *Examining the Associations between Social Network Ties and Linguistic Production*. Dissertation. Washington, D.C.: Gallaudet University.

Schembri, A. (2008). British Sign Language corpus project: Open access archives and the observer's paradox. *Paper presented at the Sixth International Conference on Language Resources and Evaluation*.

Schembri, A. C., & Lucas, C. (Eds.). (2015). *Sociolinguistics and Deaf Communities.* Cambridge University Press.

The Auslan Corpus. (2014). Digital corpus. http://www.auslan.org.au/about/corpus/

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In: *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation.*

Slicing Your SL Data into Basic Discourse Units (BDUs)
Adapting the BDU Model (Syntax + Prosody) to Signed Discourse

Sílvia Gabarró-López, Laurence Meurant
F.R.S. - FNRS and University of Namur
61, rue de Bruxelles, B-5000 Namur, Belgium
silvia.gabarro@unamur.be, laurence.meurant@unamur.be

Abstract

This paper aims to propose a model for the segmentation of signed discourse by adapting the Basic Discourse Units (BDU) Model. This model was conceived for spoken data and allows the segmentation of both monologues and dialogues. It consists of three steps: delimiting syntactic units on the basis of the Dependency Grammar (DG), delimiting prosodic units on the basis of a set of acoustic cues, and finding the convergence point between syntactic and prosodic units in order to establish BDUs. A corpus containing data from French Belgian Sign Language (LSFB) will be firstly segmented according to the principles of the DG. After establishing a set of visual cues equivalent to the acoustic ones, a prosodic segmentation will be carried out independently. Finally, the convergence points between syntactic and prosodic units will give rise to BDUs. The ultimate goal of adapting the BDU Model to the signed modality is not only to allow the study of the position of discourse markers (DMs) as in the original model, but also to give an answer to a controversial issue in SL research such as the segmentation of SL corpus data, for which a satisfactory solution has not been found so far.

Keywords: discourse segmentation, Basic Discourse Units (BDU) Model, Dependency Grammar, pauses, sign holds, eye blinks

1. To Start with...

When one cooks, different steps need to be followed to elaborate a dish. Just like when one speaks or signs, different discourse segments are sequenced to produce an oral text. At present, there are still many unresolved theoretical issues concerning the study of orality in both spoken and signed languages, although research on the latter modality is still at its infancy when compared to the first modality. Discourse segmentation is an issue at stake for both modalities because it is at the basis of how discourse in the oral setting is structured. Still, spoken language (SpL) research is slightly in advance as scholars have already developed some consolidated models to segment spoken discourse. The reason why these models emerged is that the concept of sentence works well at the level of syntax, but it has proved to be insufficient to study the structure of oral productions because both monologues and dialogues contain other elements (interjections, discourse markers, etc.) that are not comprised within the traditional syntactic notion of a sentence.

To the best of our knowledge, six different models are used for discourse segmentation of SpLs in the oral setting:[1] the Geneva Model (Roulet et al., 1985), the Val.Es.Co. Model (Briz Gómez and Grupo Val.Es.Co., 2003), the Fribourg Model (Groupe de Fribourg, 2012), the Co-Enunciation Model (Morel and Danon-Boileau, 1998), the Prominence Demarcation Model (Lombardi Vallauri, 2009) and the Basic Discourse Units Model (Degand and Simon, 2005; Degand and Simon, 2009a; Degand and Simon, 2009b). These segmentation models define their discourse units on the basis of different approaches, namely pragmatic, prosodic or in a combination of syntax and prosody. Despite these three possibilities, prosody is almost always present to a greater or lesser extend because it is one of the main forces that organises spoken discourse (Morel and Danon-Boileau, 1998) and it provides objective and measurable criteria such as pauses, tone units and differences in f0.

On the one hand, the Co-Enunciation Model (CEM) and Prominence Demarcation Model (PDM) almost totally base their segmentation methodology on prosody by taking into account long pauses, tone units or a change in f0 to delimit discourse units. On the other hand, the Geneva Model (GM), the Fribourg Model (FM) and the Val.Es.Co. Model base their segmentation methodology in pragmatics. The GM completely discards prosody and their discourse units are delimited according to illocutive forces, to whether they elicit a response from the addressee or are an answer to the addressor's preceding talk. The FM consists of a micro-syntactic and a macro-syntactic segmentation, the first using rectional links (in the sense of *rection* in French, i.e. government) and the second using prosody to delimit discourse units. The Val.Es.Co. Model (VAM) segments according to a hierarchy of units (discourse, dialogue, exchange, turn, intervention, act and subact) where prosody is seen as an accessory that should only be used if needed. Eventually, the Basic Discourse Units Model (BDU) combines both syntax and prosody for the delimitation of their units. This is the model that we took as a basis for our research and that we adapted to the signed modality.

So far, sign languages (SLs) do not have a model for discourse segmentation, although the necessity and importance of having a consistent methodology to segment signed discourses has been widely acknowledged in the literature (Crasborn, 2007; Ormel and Crasborn, 2012; Börstell et al., 2014; Hodge, 2014), to name a few. Some of these authors have undertaken some initiatives to segment signed discourses into sentences (Börstell et al., 2014;

[1]There are other well-known methodologies for the segmentation of written productions such as the Basel Model (Ferrari, 2005; Ferrari et al., 2008) and the Rhetorical Structure Theory (Mann and Thompson, 1988). Since SL data is oral, these methodologies for the analysis of written texts are discarded for the purposes of this paper.

Fenlon et al., 2007; Jantunen, 2007; Hansen and Hessman, 2007) by taking into account visual cues. Their claim is that prosody reflects somehow syntactic constituency. To study the structure of discourse, however, the initiatives that aim to a syntactic segmentation into sentences face the same problem that SpLs, i.e. the notion of sentence is not enough to account for the different constituents of oral productions. A different approach is that of Hodge (2014), who proposes to segment into clause-like units (CLUs) defined as "units of analysis smaller than discourse level [...] that correspond with various types of communicative moves in face-to-face interaction" (p. 100). CLUs are identified according to content (semantic relations, image schemas) and perceived form (intonation contours of hand and body rhythms, facial movements and enactment). However, this annotation is largely inductive as the author herself admits.

The creation of a segmentation model that puts aside subjective interpretations and that allows the study of discourse structure is not anodyne as it requires a sound knowledge of the language and a certain amount of data to work on. Unfortunately, SL research is at its beginnings so we are in very preliminary stage of knowledge on how SLs are structured in different linguistic domains (syntax, discourse, etc.), and large amounts of SL data (i.e. corpora) are recently available. The adaptation of one existing SpL segmentation model could palliate these shortcomings. However, this is far from straightforward due to the specificities of SLs: the two hands are the main articulators and they produce simultaneous constructions, and nonmanuals also participate in the construction of meaning. Bearing in mind these specificities, this paper aims to propose a model for the segmentation of signed discourse whose ultimate goal is to allow the study of the structure of discourse, and particularly the position of discourse markers (DMs) through discourse, i.e. large sets of utterances.

After reviewing the different segmentation models for SpLs, the most suitable model for the segmentation of signed discourse seems to be the BDU Model. Its main advantage is that it is not only applicable to conversation, which was the main drawback for other potentially interesting models that have already been used for the study of DMs such as the VAM or the CEM, but it can also be used for monologic data. Due to the delay in SL research, our model needs to be as versatile as possible (i.e. applicable to as many discourse situations as possible) allowing the use of the 'same measures' to segment both monologues and dialogues, and therefore get comparable units in both settings.

This paper is structured as follows. Section 2 describes the data we selected for the adaptation of the BDU Model, i.e. a sample of the LSFB Corpus (Meurant, 2015), and the type of annotations available. Section 3 includes a short presentation of the BDU Model, the different adaptations made to segment signed data and some examples of the usages we have given to the model with our data. Section 4 concludes this work.

2. Method

This research is a crumb of a larger project that aims to describe cross-linguistically DMs in French Belgian Sign Language (LSFB) and Catalan Sign Language (LSC) across different genres. Our corpus for the adaptation of the BDU Model will exclusively be made up of LSFB data because LSC data are not openly available yet. 6 deaf native signers[2] were selected from the LSFB referential corpus (Meurant, 2015). This sample is balanced in terms of age (2 signers belonging to each of the following age groups: 18-29, 30-49 and 50-80) and gender (3 men and 3 women). Signers came in couples (both belonging to the same age group) to the studio based at the University of Namur, and their conversations (including argumentative, descriptive, explicative and narrative tasks) were guided by a moderator. For this paper, we chose an argumentation on deaf issues and a narration of a past memory. In total, the corpus lasts for 42'45". Table 1 describes the content of the sample including the genre, the task instruction and the duration of the task per couple of informants.

Our data, containing conversations in LSFB from different genres and different signers, constitutes a sound corpus because language bias are avoided. Indeed, our adaptation of the BDU Model is not constrained by the specificities of a genre or by the idiosyncrasies of a single signer.

All the data were previously annotated by deaf annotators with the multimodal ELAN software[3], which allows that tiers can always be added or hidden at any time of the annotation process. The resulting files contain a basic annotation (Johnston, 2015) consisting in ID-glosses for the left and right hands, and free translations. At the time of this research, the selected files had the manual activity fully annotated, but some translations into French were still lacking. The files neither include additional detailed annotation such as non-manual features nor the annotation of units larger than individual signs such as constructed action or constructed dialogue.

For the purposes of this work, we added three extra tiers: one for syntactic units (SyU), one for prosodic units (PrU) and another one for BDUs. Syntactic and prosodic segmentation were carried out independently. First, dialogues were segmented into syntactic units. Afterwards, the SyU tier was hidden in order to delimit prosodic units in the PrU tier. When this was done, both the SyU and the PrU tiers were displayed in order to delimit BDUs in the tier created to this end.

3. Using the BDU Model to Get Sliced SL Discourses

So far, the BDU Model recipe has been used with one type of cuisine: SpL data. As argued above, we would like to give it our touch in order to get sliced SL discourses. In this section, we first describe the BDU Model in a nutshell and afterwards we present the different adaptations made for the signed modality together with some possible usages.

[2]Researchers take different criteria to refer to the concept of 'deaf native signer'. In our case, we refer to individuals who have been born in deaf families and/or that have followed all their scholarship in a boarding school for the deaf.

[3]https://tla.mpi.nl/tools/tla-tools/elan/

Genre	Task	Duration	Age group
Argumentation	Explain the differences between deaf culture vs. hearing culture	04’53”	18-29
		07’12”	30-49
		08’46”	50-80
Narration	Explain a past memory	04’46”	18-29
		09’05”	30-49
		08’09”	50-80

Table 1: LSFB data

3.1. How Does the BDU Model Work?

The main idea behind the BDU Model (Degand and Simon, 2005; Degand and Simon, 2009a; Degand and Simon, 2009b) is that the only observable linguistic criteria that must be considered for the delimitation of units are syntax and prosody. The resulting basic discourse units (BDUs) are the minimal units the addressee uses when reconstructing what the speaker is saying, and they may be of different nature, i.e. they may regulate discourse, package information, express didactic focus, emphasis (Degand and Simon, 2009a). The segmentation consists of two independent analyses: a syntactic and a prosodic one that are carried out in Praat[4], a free open source software package to analyse speech.

On the one hand, syntactic mapping is based on the Dependency Grammar (DG) as conceived for spoken French by Blanche-Benveniste et al. (1984) and Blanche-Benveniste et al. (1990). In short, syntactic units are comprised of a ’nucleus’ (mostly verbs, but also nouns or adjectives) that governs its ’dependants’, i.e. actants (specific dependants belonging to the restricted valency of the verb) or circumstants (dependants of the verb but out of its valency). In addition, there are other elements called adjuncts that can be added to any construction in a less restrictive way. In example 1, borrowed from Degand and Simon (2005, p. 69) together with its translation, the clause contains a verb which is the nucleus, together with three elements: ’le permis’ is the actant (ACT) as it belongs to the valency of the verb, i.e. one needs to specify what is possessed. ’Pour le moment’ is the cirsumstant (CIRC) as it depends on the verb but it is out of its valency, i.e. if it is left out, the clause still makes sense. Finally, ’de toute manière’ is an adjunct (ADJ) because its role is to connect clauses.

(1) *<de toute manière>$_{ADJ}$ [j’ai pas le permis$_{ACT}$ pour le moment$_{CIRC}$]*
’<in any case>$_{ADJ}$ [I do not have the licence (driver’s)$_{ACT}$ for the moment$_{CIRC}$]’

The result of this syntactic mapping gives rise to three different types of dependency clauses (Tanguy et al., 2012): verbal dependency clauses (governed by a verb), averbal dependency clauses (governed by an element other than a verb), and elliptical dependency clauses (incomplete clauses that can be interpreted as verbal dependency units when referring to the context as in answers). Moreover, these units can either be interrupted (the clause lacks an obligatory complement and/or it has been started but not completed) or contain adjuncts such as DMs. The final step for the annotation of syntactic units consists in identifying the ’functional sequences’ that integrate each dependency clause. These functional sequences are the ”clausal constituents that occupy a main syntactic function like Verb, Subject, Object, etc.” (Degand et al., 2014, p. 248).

On the other hand, prosodic mapping is trickier because there is not a consensus on a prosodic model to be used for French. Prosodic segmentation is therefore performed on the basis of a semi-automatic annotation procedure developed by Mertens and Simon (2009) that allows the establishment of major, intermediate and minor prosodic boundaries. Neither minor nor intermediate boundaries are taken into consideration for the BDU Model, only major boundaries are used for segmentation. Major boundaries are marked by a silent pause, a lengthening of the syllable (three times longer than the syllables in context) or a sharp rise of f0 (intra-syllabic f0 superior to ten semi-tones). Four types of segmentation units result from this procedure: ”Continuation (rising f0 movement), Finality (falling or low f0), Focus (sharp falling from high to low contour) and Suspense (flat and lengthened contour)” (Degand et al., 2014, p. 249).

Once the syntactic and the prosodic units have been delimited, the frontiers of BDUs are established in the places where syntactic and major prosodic boundaries coincide. As a result, BDUs may be congruent (syntactic and prosodic boundaries coincide), syntax-bound (a syntactic unit contains several prosodic units), intonation-bound (a prosodic unit contains several syntactic units), regulatory (the unit is an adjunct or DM) or mixed (there are several syntactic and prosodic units within the BDU before the boundaries coincide). This segmentation including a syntactic and a prosodic mapping into BDUs allows the study of the position of different discourse elements.[5] The dependants of a nucleus can be found in initial or in final position with respect to the verb (SV), which means that they are syntactically dependent and prosodically integrated. Example 2 illustrates a clause with a CIRC in initial position.

(2) [*à la première manifestation la première journée de grève*$_{CIRC}$ *c’était*$_{SV}$ *le mouvement s’essouffle*]
’[at the first demonstration the first day of the strike$_{CIRC}$ it was$_{SV}$ the movement ran out of steam]’

[4] http://www.praat.org/

[5] The following examples of this section together with the translations are borrowed from Degand et al. (2014) and simplified. See note 7 for the symbols used in the examples.

Dependants can also be syntactically dependent but prosodically isolated, i.e. their position is either the prosodic left periphery (LP) or right periphery (RP). This means that, if we take the clause in example 2, there would be a prosodic break between the CIRC and the SV.
As for adjuncts (such as DMs, agents, etc.), they can be syntactically independent but prosodically integrated (syntactic LP or RP). Another possibility for them is to be syntactically independent and prosodically isolated (i.e. syntactico-prosodic LP or RP). In this last case, they constitute a regulatory BDU. Example 3 illustrates these two positions. There are two DMs and two BDUs. The first DM, i.e. 'bon', is situated at the syntactico-prosodic LP and makes up the first BDU which is regulatory. The second DM, i.e. 'mais', is at the syntactic LP as it is out of the dependency of the elliptical clause 'pas nous', but both the clause and the DM are within the second BDU.

(3) *<bon>$_{dm}$ / <mais>$_{dm}$ [pas nous]*
'<well>$_{dm}$ / <but>$_{dm}$ [not us]'

All these positions are summarised in table 2.

Position	Syntactically	Prosodically
Initial	Dependent	Integrated
Final	Dependent	Integrated
Syntactic LP	Independent	Integrated
Syntactic RP	Independent	Integrated
Prosodic LP	Dependent	Isolated
Prosodic RP	Dependent	Isolated
Syntactico-prosodic LP	Independent	Isolated
Syntactico-prosodic RP	Independent	Isolated

Table 2: Possible positions in the BDU

3.2. How Do I Use the BDU Model with My Signed Data?

We will answer this question by presenting how we adapted the recipe of the BDU Model to get sliced SL discourses. Afterwards, we also suggest some serving ideas, i.e. some possible applications of the model.

3.2.1. Adapting the Recipe

Take your videos and get prepared to slice them. The first adaptation is that we will use a different segmenting tool from the original BDU Model, i.e. we will be using ELAN instead of Praat. Anyway, our segmentation procedure will consist of three different steps: (i) delimiting syntactic units, (ii) delimiting prosodic units, and (iii) finding the convergence point between syntactic and prosodic units in order to establish BDUs. The first two steps are independent, which means that once the syntactic segmentation is finished, this tier will be hidden in order to carry out the prosodic segmentation independently.

Syntactic segmentation The BDU Model delimits syntactic units (i.e. clauses) using the DG for spoken French as conceived by Blanche-Benveniste et al. (1984) and Blanche-Benveniste et al. (1990), then annotates the type of clause (i.e. verbal, averbal, elliptical, interrupted or containing a nondependent element) and finally annotates the 'functional sequences' (i.e. verb, subject, object, etc.). For the moment, we will only delimit clauses and we will leave aside the annotation of the type of clause and the functional sequences. Since the ultimate goal of having a segmentation model of signed discourse for us is to study the position of DMs, this first step suffices.[6]
The DG establishes the verb as the nucleus that governs different dependants: actants (inside the valency of the verb) and circumstants (outside the strict valency of the verb). Adjuncts can be added to any construction in a less constrained way. The role of nucleus can also be fulfilled by other elements such as pronouns, nouns or adjectives. In what follows, we will give examples to illustrate these possibilities. We will gather these examples[7] under the three types of clauses that the BDU Model establishes.
Verbal dependency clauses. As its name reveals, the clause contains a verb that is the nucleus. In SLs, the verb can be either a fully-lexical or a partly-lexical sign. Fully-lexical signs are those tokens that can be dictionary entries of a sign language because they are "highly conventionalised signs in both form and meaning in the sense that both are relatively stable or consistent across contexts" (Johnston, 2015, p. 13). Partly-lexical signs cannot be listed in a dictionary as they are "combinations of conventional and non-conventional (highly contextual) elements" (ibid.). In example 4 (`http://www.corpus-lsfb.be`, session 2, task 4, 03:18-03:20), the verb is a fully-lexical sign with one actant (ACT) and one circumstant (CIRC).

(4) [PT:PRO1 BUY LAND$_{ACT}$ WITH HEARING$_{CIRC}$]
'[I bought a land$_{ACT}$ with a hearing person$_{CIRC}$]'

In example 5 (`http://www.corpus-lsfb.be`, session 21, task 4, 02:46-02:50), there is a fully-lexical (START) and a partly-lexical (DS:go-up) sign. The nucleus in this case is DS:go-up and START SCHOOL constitutes a clause that depends on the verb DS:go-up as it is expressing the moment in which the action happens. Therefore, START SCHOOL behaves as a circumstant of the verb DS:go-up. DS:go-up has an actant which is SCHOOL and a circumstant which is UNTIL TEN.

(5) [START SCHOOL$_{CIRC}$ DS:go-up SCHOOL$_{ACT}$ UNTIL TIME-TEN$_{CIRC}$]
'[when it was the time to start school$_{CIRC}$ we went upstairs (to the classroom)$_{ACT}$ until ten o'clock$_{CIRC}$]'

[6] We are aware of the interest of looking at the material inside the clause, specially if further research is to be undertaken with a focus on syntax. However, doing so would make the segmentation process too long and therefore not feasible if we want to take several productions (containing different genres and signers) into account.

[7] All the examples in LSFB are annotated using the following conventions: *PT:PRO1* stands for the first person pronoun, *DS:go-up* stands for a depicting sign and its description in context, *-I* stands for an interruption, [ ] delimit syntactic units (i.e. clauses), < > delimit adjuncts, / separates two different BDUs, and underlined text preceded by $\oint$ marks overlapping between signers.

Averbal dependency clauses. The verb is usually considered the nucleus of the clause. However, there are other elements in oral productions that constitute an averbal dependency clause in itself such as when somebody answers YES to a question (see example 8). On the other hand, SLs allow that another signs working as a pronouns, nouns or adjectives fulfil the role of a nucleus.

In example 6 (`http://www.corpus-lsfb.be`, session 2, task 4, 06:43-06:46), the copulative verb is not manually expressed in the clause, that's why we take the sign DIFFICULT as the nucleus.

(6) [BANK DIFFICULT MONEY RECEIVE]
'[it was difficult to get money from the bank]'

Elliptical dependency clauses. When a clause is incomplete but it is interpretable as a verbal dependency clause thanks to the context, it is called an elliptical dependency clause (Tanguy et al., 2012). Answers and interventions (among other possibilities) fall within this category. In example 7 (`http://www.corpus-lsfb.be`, session 21, task 3, 04:10-04:16), the signers are discussing about the football matches that used to take place between two boarding schools for the deaf in Brussels called IRSA and Woluwe. S045 says that both schools keep this tradition, but S044 answers that this is not as often as it used to be. The clause that S044 utters (overlapping with his partner) is interpretable as a verbal dependency clause when referring to what S045 says.

(7)
S045: [IRSA WOLUWE TEAM AGAIN PLAY AGAIN] ∮[FOOTBALL PLAY AGAIN] [LESS]
S044: ∮[PT:DET EVERYDAY] [LESS]
S045: '[the teams at IRSA and Woluwe still play] ∮[they still play football] [less (yes)]'
S044: '∮[not everyday][less often]'

Interrupted dependency clauses. This category encloses all those clauses (verbal, averbal or elliptical) that have been interrupted for a number of reasons such as a new thought that came to the signer's mind, an intervention from the addressee or because of any other contextual factor. In example 8 (`http://www.corpus-lsfb.be`, session 27, task 4, 00:32-00:36), there are two different interruptions. The first time S055 interrupts S056, who is word searching. The second time S055 interrupts herself: she begins a sentence but does not finish it because of the backchannel that S056 provides her.

(8)
S056: $[\text{FEEL}]^{-I}$
S055: [FEEL MORE DEAF] ∮<PALM-UP> $[\text{BECAUSE PERSON-BLOW}]^{-I}$ [YES] [THAT-S-IT]
S056: ∮[YES] <PALM-UP> [GIVE] [YES] <PALM-UP>
S056: '$[\text{I feel}]^{-I}$'
S055: '[you feel more deaf] ∮<don't you> $[\text{because I realised}]^{-I}$ [yes] [that's it]'
S056: '∮ [yes] <erm> [it makes me feel] <yeah>'

Clauses containing a nondependent element. This type of clauses include those cases in which adjuncts, i.e. elements that stay out of the dependency structure of the verb such as a DM, constitute a clause alone. Example 9 (`http://www.corpus-lsfb.be`, session 21, task 4, 02:46-02:53) retakes the clause in example 5 and the following one.

(9) [START SCHOOL_{CIRC} DS:go-up SCHOOL_{ACT} UNTIL TIME-TEN_{CIRC}] $\langle\text{AFTERWARDS}\rangle_{dm}$ [PLAY FIFTEEN MINUTES_{CIRC}]
'[when it was the time to start school_{CIRC} we went upstairs (to the $\text{classroom})_{ACT}$ until ten o'clock_{CIRC}] $\langle\text{afterwards}\rangle_{dm}$ [we played for fifteen minutes_{CIRC}]'

The two clauses express a sequence, i.e. one event happens after the other. This meaning is explicited by the sign AFTERWARDS that does not take part in the dependency structure of any verb, i.e. it is an adjunct. The nuclei of the two clauses are the depicting sign of movement (Johnston, 2015) DS:go-up and the fully-lexical sign PLAY respectively.

Prosodic segmentation Delimiting discourses into prosodic units is the step that follows syntactic segmentation in the BDU Model. As previously mentioned, both segmentation processes are independent, that's why the tier containing syntactic units has to be hidden before the delimitation of prosodic units starts.

Two types of adaptations are made in order to work with signed data: technical and modality-based adaptations. On the one hand, prosodic segmentation in the BDU Model for spoken data is semi-automatic (Mertens and Simon, 2009). This is far from being a reality with SL data and our prosodic segmentation will be completely manual. On the other hand, the differences between the spoken and the signed modality imply that instead of using a set of acoustic cues to segment spoken productions, we will need to establish a set of equivalent visual cues. The three acoustic cues used in the BDU Model are silent pauses, a lengthening of the syllable (three times longer than the syllables in context) or a sharp rise of f0 (intra-syllabic f0 superior to ten semi-tones). It is known that visual prosodic cues indicate either the duration of phrases (domain markers) or their end (boundary markers) (Wilbur, 2000; Wilbur, 2009). Boundary markers include eye blinks, short head nods, pauses and holds (Herrmann, 2012); and they segment discourse into rhythmic units as they are punctual in nature (i.e. they do not spread) (Pfau and Quer, 2010). Moreover, they coincide with some of the acoustic cues taken into account in the BDU Model for prosodic segmentation.

The first two acoustic cues have a straightforward equivalent in signed discourse as boundary markers: pauses and sign holds (or lengthened signs with respect to the context). *Pauses* are defined as periods of no signing at all in line with Fenlon (2010), and they include stops in which the hands are crossed, are left along the body or are relaxed and placed in the neutral space (Notarrigo and Meurant, 2014). The type of pause is not annotated because it does not seem to provide us with relevant information, and the length is neither measured.[8]

Sign holds and *lengthened signs* with respect to the context are the equivalents to lengthened syllables. A sign hold appears when the handshape of a sign is held for a longer duration, and a lenghtened sign implies that the movement of the sign is repeated, slowed or exaggerated. Although holds can appear at the beginning, in the middle or at the end of the sign (Notarrigo and Meurant, 2014), only holds occurring at the end are taken into account for segmentation. As happened with pauses, neither the type of hold nor the function of the sign hold or lengthened sign are annotated.

For the third acoustic cue used in the BDU Model, i.e. a sharp rise of f0, we did not find a straightforward equivalent within the group of visual boundary markers.[9] We propose to take *eye blinks* as a last visual cue because they are widely acknowledged a prosodic function of marking boundaries in the signed modality (Wilbur, 1994; Sze, 2008; Brentari and Crossley, 2002; Crasborn et al., 2004; Herrmann, 2010) and they segment the discourse into rhythmic units (Pfau and Quer, 2010; Herrmann, 2012) as raises in f0 do. However, not every eye blink is prosodically relevant as they may serve other linguistic purposes or be physiologically motivated (Wilbur, 1994; Sze, 2008; Herrmann, 2012).

Herrmann (2010) distinguishes between prosodic and non-prosodic blinks by analysing syntactic constituency and sentence structure together with intonational contours and other nonmanual features such as eyebrow movement, eye aperture, eye gaze, head movement, body movement, mouth gestures, and facial expressions (p. 22). We will not follow her syntactic criterion for the identification of prosodic blinks because it would contradict one of principles of the BDU Model, namely that prosodic segmentation has to be done independently from syntax. We will restrict the identification of prosodic blinks to those occurring with another nonmanual prosodic cue. Blinking layered with another prosodic cue was one of the three most common markers of discourse units' boundaries (after pauses and sign holds) in a previous study about the segmentation of LSFB data (Gabarró-López and Meurant, 2014). As with the other cues (i.e. pauses and sign holds or lengthened signs), we will not annotate the visual cues occurring at the prosodic units' boundaries because our purpose is not to do a prosodic analysis but to have a set of cues for prosodic segmentation.[10]

Delimitation of Basic Discourse Units The final step consists of establishing BDUs where syntactic and prosodic units coincide. Therefore, we will display both the syntactic and prosodic tiers and we will delimit BDUs in a separate tier as shown in Figure 1. The first two tiers in the figure are devoted to the annotation of the manual activity, the third is for syntactic units (SyU), the fourth is for prosodic units (PrU) and the fifth for BDUs. As for SpL data, different kinds of BDU arise after the segmentation of SL data depending on how syntactic and prosodic units align (congruent, syntax-bound, intonation-bound, regulatory and mixed). In Figure 1, there are two different BDUs, the first one is syntax-bound (there are two prosodic units within a syntactic unit) and the second one is intonation-bound (there are two syntactic units within a prosodic unit). So far, the type of BDU has not been annotated although we would like to do so in the future to see if a particular type of BDU is more likely to appear in one genre than in another.

3.2.2. Serving Ideas

Although our adaptation of the BDU Model recipe to get sliced SL discourses is time-consuming, the results have proved to be satisfactory. Undoubtedly, this revisited recipe can be taken as the basis for future research works that require the segmentation of SL discourse. Our serving idea for our sliced SL discourses is to study the position of DMs. As a matter of fact, we have seen that the position of the DM *AUSSI* (here translated as ALSO) in LSFB correlates with its function in a particular context. If we take two common functions of ALSO, i.e. addition (adding information to the same topic) and specification (introducing an example), we can see that each function displays a particular position with respect to the clause and the BDU. Addition is found at the left periphery (LP) of the clause and at the prosodic LP of the BDU as in example 10 (`http://www.corpus-lsfb.be`, session 21, task 04, 2:37-2:42). There are four clauses, two within each BDU. ALSO is out of the dependency structure of the verb GO (i.e. clausal LP), but it is prosodically integrated at the beginning of the BDU (i.e. syntactic LP).

(10) [HEARING PT:PRO1 GO BICYCLE LEARN] [BICYCLE THERE GO] / <ALSO> [GO HORSE] [PT:PRO1 GO HORSE]
'[the Hearing taught me how to cycle] [I went by bicycle] / <and> [I rode horses] [I went to ride horses]'

Specification is found at the LP of the clause and in the medial position of the BDU as in example 11 (`http://www.corpus-lsfb.be`, session 27, task 04, 2:29-2:33). In this case, there are three clauses within the same BDU. ALSO is out of the dependency structure of the verb REMEMBER (i.e. clausal LP), but it is prosodically integrated in the middle of the BDU (i.e. BDU medial position).

(11) [YES] <ALSO> [REMEMBER BEFORE LITTLE ALWAYS PT:PRO1] [TODAY SECOND MEMORY CHILD]
'[yes] <for instance> [I remember when I was young] [this is my second child memory today]'

[8]The length of a pause is what distinguishes an intermediate and a major boundary in the BDU Model. Not measuring the length of a pause may be seen as a shortcoming in our adaptation, but since we lack of a semi-automatic tool, we should do it manually. Doing so would make the segmentation procedure too long and therefore not feasible.

[9]Spontaneously, one could think of exaggerated signing as a possible equivalent. The main problem is that since there is not a semi-automatic tool that can measure exaggerated signing, this annotation is somehow arbitrary and depends on the annotator's perception.

[10]For a study of the most common boundary markers that can be found in LSFB, we refer the reader to Gabarró-López and Meurant (2014).

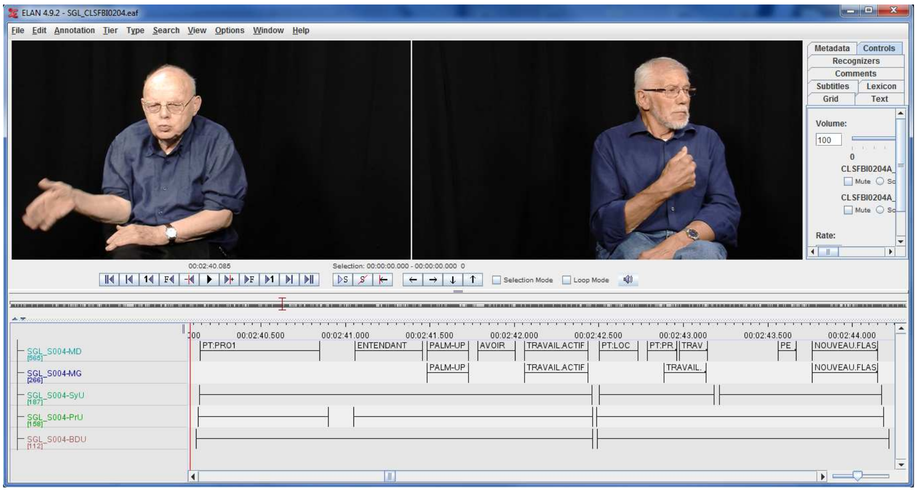


Figure 1: Screenshot of an ELAN file contaning syntactic units, prosodic units and BDUs.

This coupling of position and function of ALSO is regular across different examples of our corpus, which includes different signers and different genres. Therefore, the position can be used as a criterion to identify the function of a polysemous DM such as ALSO, whose annotation strongly depends on the annotator's interpretation.

4. To End with...

The recipe to get sliced SL data with the BDU Model is now ready, so make the most of it! At present, our adaptation has proved to be useful for the study of DMs. However, its applications are not restricted to this topic. Our proposal offers a solution for a controversial issue in SL research, i.e. the segmentation of SL corpus data, for which a satisfactory solution has not been found so far. Therefore, this methodology can lead us to explore many unknown issues and answer many questions related to how SLs work.

The segmentation of SL discourses into BDUs could cast light on some of the signers' cognitive processes such as the interpretation of discourses. It is said that a coherent interpretation of a discourse is the result of "integrating the discourse units into a mental representation" (Degand et al., 2014, p. 244). This information packaging takes place at the left periphery (LP) of discourse units as it is the place where the message is started and where it is connected with what was previously said. The study of the LP would give insight on the assumption that SLs prefer constructions of topicalization, or could reveal whether SLs prefer implicit discourse relations over explicit discourse relations.

Another possibility in line with this study of the LP from a broader perspective is to better understand the differences between genres and registers. For instance, we could get to know the discourse features that define a formal speech such as a conference vs. an informal speech such a joke; or the devices preferred in a monologue over a dialogue. Cross-linguistic SL studies on how discourse is structured would also be feasible because the model proposes a set of common criteria that avoid the annotators' subjective interpretations and therefore assure consistency across data.

Just try it and let us know!

5. Acknowledgements

This research was funded by a F.R.S.-FNRS Research Fellow Grant (1.A.697.14F) and the F.R.S.-FNRS Incentive Grant for Scientific Research (F.4505.12). We would like to show our gratitude to the informants of the LSFB Corpus project for their participation. We would also like to thank the anonymous reviewers and Gemma Barberà for their constructive comments. Any remaining mistakes are our own responsibility.

6. Bibliographical References

Blanche-Benveniste, C., Deulofeu, J., Stefanini, J., and van den Eynde, K. (1984). *Pronom et syntaxe. L'approche pronominale et son application au français*. SELAF, Paris.

Blanche-Benveniste, C., Bilger, M., Rouget, C., and van den Eynde, K. (1990). *Le français parlé: études grammaticales*. Éditions du CNRS, Paris.

Börstell, C., Mesch, J., and Wallin, L. (2014). Segmenting the Swedish Sign Language corpus: On the possibilities of using visual cues as a basis for syntactic segmentation. In Onno Crasborn, et al., editors, *Beyond the manual channel. 6th Workshop on the Representation and Processing of Sign Languages*, pages 7–10, Reykjavik. ELRA.

Brentari, D. and Crossley, L. (2002). Prosody on the hands and face. Evidence from American Sign Language. *Sign Language & Linguistics*, 5(2):105–130.

Briz Gómez, A. and Grupo Val.Es.Co. (2003). Un sistema de unidades para el estudio del lenguaje colloquial. *Oralia*, 6:7–61.

Crasborn, O., van der Kooij, E., and Emmerik, W. (2004). Prosodic features of focus in NGT.

Onno Crasborn, editor. (2007). *Identifying sentences in signed languages*, volume 10-2 of *Special issue of Sign Language and Linguistics*. John Benjamins Publishing Company, Amsterdam; Philadelphia.

Degand, L. and Simon, A. C. (2005). Minimal Discourse Units: Can we define them, and why should we? In *Proceedings of SEM-05. Connectors, discourse framing and* *discourse structure: from corpus-based and experimental analyses to discourse theories*, pages 65–74, Biarritz.

Degand, L. and Simon, A. C. (2009a). Mapping Prosody and Syntax as Discourse Strategies: How Basic Discourse Units vary across Genres. In Dagmar Barth-Weingarten, et al., editors, *Where Prosody meets Pragmatics*, volume 8 of *Studies in Pragmatics*, pages 75–105. Emerald, Bingley.

Degand, L. and Simon, A. C. (2009b). Minimal Discourse Units in Spoken French: On the Role of Syntactic and Prosodic Units in Discourse Segmentation. *Discours*, 4.

Degand, L., Simon, A. C., Tanguy, N., and van Damme, T. (2014). Initiating a discourse unit in spoken French. In Salvador Pons Bordería, editor, *Discourse Segmentation in Romance Languages*, volume 250 of *Pragmatics & Beyond New Series*, pages 243–273. John Benjamins Publishing Company, Amsterdam; Philadelphia.

Fenlon, J., Denmark, T., Campbell, R., and Woll, B. (2007). Seeing sentence boundaries. In Onno Crasborn, editor, *Special issue of Sign Language & Linguistics*, volume 10-2, pages 177–200. John Benjamins Publishing Company, Amsterdam; Philadelphia.

Fenlon, J. (2010). *Seeing sentence boundaries: the production and perception of visual markers signalling boundaries in signed languages*. Doctoral thesis, University College London, London.

Ferrari, A., Cignetti, L., De Cesare, A.-M., Lala, L., Mandelli, M., Ricci, C., and Roggia, E. (2008). *L'interfaccia lingua-testo. Natura e funzioni dell'articolazione informativa dell'enunciato*. Edizioni dell'Orso, Alessandria.

Angela Ferrari, editor. (2005). *Tipi di testo e tipi di gerarchie testuali, con particolare attenzione alla distinzione tra scritto e parlato*. Franco Cesati editore, Firenze.

Gabarró-López, S. and Meurant, L. (2014). When nonmanuals meet semantics and syntax: a practical guide for the segmentation of sign language discourse. In Onno Crasborn, et al., editors, *Beyond the manual channel. 6th Workshop on the Representation and Processing of Sign Languages*, pages 55–61, Reykjavik. ELRA.

Groupe de Fribourg. (2012). *Grammaire de la période*. Peter Lang, Berne.

Hansen, M. and Hessman, J. (2007). Matching propositional content and formal markers: Sentence boundaries in a DGS text. In Onno Crasborn, editor, *Identifying sentences in signed languages*, volume 10-2 of *Special issue of Sign Language and Linguistics*, pages 145–175. John Benjamins Publishing Company, Amsterdam; Philadelphia.

Herrmann, A. (2010). The interaction of eye blinks and other prosodic cues in German Sign Language. *Sign Language & Linguistics*, 13(1):3–39.

Herrmann, A. (2012). Prosody in German Sign Language. In Pilar Prieto et al., editors, *Prosody and Meaning*, pages 349–380. Mouton de Gruyter, Berlin.

Hodge, G. (2014). *Patterns from a signed language corpus: Clause-like units in Auslan (Australian sign language)*. Doctoral thesis, Macquaire University, Sydney.

Jantunen, T. (2007). The equative sentence in Finnish Sign Language. In Onno Crasborn, editor, *Identifying sentences in signed languages*, volume 10-2 of *Special issue of Sign Language and Linguistics*, pages 113–143. John Benjamins Publishing Company, Amsterdam; Philadelphia.

Johnston, T. (2015). Auslan Corpus Annotation Guidelines. Guidelines, Macquarie University; La Trobe University, Sydney; Melbourne.

Lombardi Vallauri, E. (2009). *La struttura informativa. Forma e funzione negli enunciati linguistici*. Carocci, Roma.

Mann, W. C. and Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

Mertens, P. and Simon, A. C. (2009). Automatic Detection of Prosodic Boundaries in Spoken French. Technical report, Ms. Katholieke Universiteit Leuven and Université catholique de Louvain, Leuven; Louvain-La-Neuve.

Morel, M.-A. and Danon-Boileau, L. (1998). *Grammaire de l'intonation. L'exemple du français*. Ophrys, Paris.

Notarrigo, I. and Meurant, L. (2014). Nonmanuals and markers of (dis)fluency in French Belgian Sign Language (LSFB). In Onno Crasborn, et al., editors, *Beyond the manual channel. 6th Workshop on the Representation and Processing of Sign Languages*, pages 135–142, Reykjavik. ELRA.

Ormel, E. and Crasborn, O. (2012). Prosodic correlates of sentences in signed languages: A Literature Review and Suggestions for New Types of Studies. *Sign Language Studies*, 12(2):109–145.

Pfau, R. and Quer, J. (2010). Nonmanuals: their prosodic and grammatical roles. In Diane Brentari, editor, *Sign Languages*, pages 381–402. Cambridge University Press, Cambridge.

Roulet, E., Auchlin, A., Moeschler, J., and Schelling, M. (1985). *L'articulation du discours en français contemporain*. Peter Lang, Berne.

Sze, F. (2008). Blinks and intonational phrasing in Hong Kong Sign Language. In Josep Quer, editor, *Signs of the time: selected papers from TISLR 2004*, pages 86–107. Signum, Hamburg.

Tanguy, N., van Damme, T., Degand, L., and Simon, A. C. (2012). Projet FRFC "Périphérie gauche des unités de discours" - Protocole de codage syntaxique. Technical report.

Wilbur, R. (1994). Eyeblinks and ASL phrase structure. *Sign Language Studies*, (84):221–240.

Wilbur, R. (2000). Phonological and prosodic layering of nonmanuals in American Sign Language. In Karen Emmorey et al., editors, *The signs of language revisited. An Anthology to Honor Urusla Bellugi and Edward Klima.*, pages 213–244. Lawrence Erlbaum Associates, Mahwah, N.J; London.

Wilbur, R. (2009). Effects of varying rate of signing on ASL manual signs and nonmanual markers. *Language and Speech*, 52(2/3):245–285.

7. Language Resource References

Meurant, L. (2015). Corpus LSFB. Corpus informatisé en libre accès de vidéo et d'annotation de langue des signes de Belgique francophone.

Towards a Visual Sign Language Corpus Linguistics

Thomas Hanke
Institute of German Sign Language and Communication of the Deaf, University of Hamburg, Germany
E-mail: thomas.hanke@sign-lang.uni-hamburg.de

Abstract

Visualisations have a long tradition in linguistics, as in many fields dealing with complex structure. New forms of representations have been introduced to Visual Linguistics in the recent past, e.g. to help the researcher find the needle in a haystack, i.e. corpus. Here we present visualisation services available in iLex making a combined corpus and lexical database visually accessible. While many approaches suggested for textual languages transfer to sign language data as well, others explore sign-specific structure, such as multi-dimensional concordances not being restricted to sequentiality. Experimental combinations of animated visualisation and image processing might support the researcher to compensate for incomplete high-quality (=manual) annotation. In the long run, we see the potential that visualisation and data manipulation go hand in hand, allowing future user interfaces that are less text-heavy than today's sign language annotation environments.

Keywords: annotation, lexical database, iLex, visualisation, SQL, charting, mapping, geospatial data, R, graphviz, D3js

1. Introduction

Even though sign linguistics works on a visual language, it is not visual itself, or not more than linguistics on any other language. As Visual Linguistics, by no means being a new field, but having received increased attention over the last years, often operates on the levels of types and their relations to token, informants, or other types, many of the visualisation ideas transfer one-to-one to sign language corpora. With many interesting approaches having appeared in the last years, there is a lot to gain from.

Whether the researcher tries to find the needle in a haystack (like interesting structure worth a closer look or potential encoding errors maybe showing as either clusters or outliers in a visual representation) or to get an intuition what hypotheses to formulate and test, visualisation techniques should be readily available and well integrated into the sign language corpus linguist's workflow.

Visual representations are also most useful when illustrating complex relations to others, be it colleagues or students. Depending on the audience, the same representations that are used by the researcher to get an overview or detect new facts may be used, or more sophisticated graphics need to be produced, often abstracting further away from the actual data.

In addition to the degree of sophistication (from quick & dirty to fine-tuned graphics for lectures, for example), another independent dimension has come up in the past years: The degree of interactiveness. For visual representations that go into traditional print publications, non-interactive graphics are enough. For slides, more and more researchers make use of interactive visual representations. Animation not only makes visuals more attractive, but also allows the presenter to direct the viewer's (or even user's) attention to specific aspects. But interactive graphics also make sense for the researcher him-/herself when it opens the possibility to sort, zoom, or focus the attention or move back and forth on a timeline. With modern libraries such as D3js making this kind of display easier to implement, more and more researchers want to explore the potential of such displays.

2. Data Visualisation in iLex

As iLex is a corpus and lexical database (cf. Hanke, 2002, and Hanke & Storz, 2008), providing the data is "simply" a question of selection. We use SQL queries to provide the data to be rendered since SQL is a very powerful way of searching, selecting, grouping and ordering the data, spanning annotation and lexical database. The obvious disadvantage of this approach is that the user needs a good command of SQL to produce the tabular data s/he is interested in. To partially overcome this problem, iLex allows the user to store "chart" definitions, i.e. the underlying SQL query as well as the chart style. That way, the user him-/herself or any other user can execute the same chart at a later point of time, either on the same data or on other data points of the same category. iLex charts are either global or take data points of a certain category as input, like the types selected in a list of types. Thereby, it is easy to produce graphics specific for a set of types (or concepts etc.) the user is interested in – without having to read or even understand the SQL. This holds true for all kinds of charts implemented in iLex:

2.1. Business Charts in iLex (figs. 1-4)

iLex can convert tabular data into pie charts, bar charts or scatter plots most users are familiar with from popular spreadsheet applications. While the customisation options fall short compared to specialised application, the user can select in iLex which data points to create the graphics for, there is no need to copy the data elsewhere, and most importantly the user can double-click on a bar or pie segment or scatter point to open the related data point, or, in case of aggregation, a list showing all data points belonging to the selected aggregate.

Typical uses include token counts for selected types grouped by informant metadata such as sex and age group but also statistical data on annotation progress.

2.2. Graphs, Nets and Lattices in iLex (fig. 5)

Types and concepts quite naturally form complex nets that can be visualised inside iLex by virtue of the Graphviz library[1] integrated (cf. Gansner & North, 2000). Graphviz implements several algorithms to layout complex graphs with minimal edge overlaps. Double-clicking nodes or edges may open relevant detail.

2.3. Maps in iLex (figs. 6-7)

The combination of corpus data and related informant metadata allows for most interesting geolinguistic queries, such as the regions where users of a particular sign are from when trying to make up one's mind about the regional distribution of signs. Such data obviously is best displayed in maps.

iLex makes use of sophisticated geospatial R scripts [2] to plot the maps (Perpiñán Lamigueiro, 2014). For this to work, some data is needed in the background that relates geographical regions of interest to regions on a map. This data needs to be preloaded into iLex to match the regional distribution of target countries the database contents is related to. Again, the chart definition determines what happens when the user double-clicks on a map tile.

In our database, we offer geolinguistic queries on different levels of granularity (states, counties and data collection regions of the DGS Corpus[3] project[4]). While the infrastructure would also allow maps showing the exact living places of informants using a specific sign, such queries are generally blocked in this database for data privacy reasons as with a rather small set of informants from a regionally distributed minority re-identification often is possible from the living place alone.

2.4. Interactive Graphs in iLex (figs. 8-10)

The most recent addition to iLex's charting capabilities is the integration of D3js[5], a JavaScript library to design graphics that have more interactive functions than the aforementioned chart types (cf. Murray, 2013). There is a plethora of business chart and graph styles available building on D3js[6], but any real application requires tweeking the JavaScript code so that some JavaScript programming skills are needed to integrate new styles into iLex.

The advantage of these interactive graphs is that one can program them in a way to display a node's children when double-clicked or just grab a node and move it to another part of the window e.g. to sort by individual criteria. The logical next step would be to use these graphs not only for visualising data, but also for manipulating them. D3js has all the needed capabilities and easily connects with the iLex database. So in the long run, we expect such graphs to replace the text-heavy tabular data displays used all over the place. We hope that over time D3js will develop in a way to allow a clear separation of display and manipulation code so that security measures can apply. For the time being, we do not allow data manipulation SQL code inside D3js code, but only queries.

2.5. Exporting Charts from iLex

All charts created in iLex are in Scalable Vector Graphics format (svg) and thus can directly be integrated into web pages. For other programs not capable of importing svg, iLex allows printing the chart to PDF. Unfortunately, only the final view of animated visualisations shows in the PDF. So for exporting into slides etc., there also exists the option to export to a movie file.

3. Relations Explorable via Charts

A lexical database has a rich inventory of explicit relations between entities that can be visualised in a task-specific way. In addition, any distance measure defined between types implicitly establishes additional relations between them. We here explore similarity of HamNoSys descriptions; ASL-LEX (Sevcikova Sehyr et al., 2016) demonstrates that the same is possible and extremely insightful, based on phonological properties.

Combing corpus and lexical data in the database, there are both traditional and sign-specific approaches: With sign languages being able to articulate two (one-handed) signs at the same time, concordances become multi-dimensional. For our purposes, a concordance graph that color-codes the dimensions seems to be a good solution.

With the corpus data providing durations of tokens in a reliable way, it becomes possible to observe a signer's signing speed over the course of conversations in different elicitation settings.

While it is possible to combine various chart types into one window using the R and D3js renderers, e.g. to have pie charts for each region shown in a map, iLex offers another easy and flexible way of combining charts: The hyperlink determining what happens when the user double-clicks a chart segment, a graph node, or a map tile can also refer to another chart. That way, cascades of charts can be built with minimal effort. For example, the user can select from types visually grouped by phonetic features in order to see their regional distribution on a map.

[1] http://www.graphviz.org ; last access: March 26, 2016
[2] http://www.r-project.org ; last access: March 26, 2016
[3] http://dgs-korpus.de ; last access: March 26, 2016
[4] The maps have been produced from data provided by a German public body responsible for geodata, cf. http://www.geodatenzentrum.de/docpdf/vg1000.pdf ; last access: March 26, 2016
[5] http://www.d3js.org ; last access: March 26, 2016
[6] http://bl.ocks.org ; last access: March 26, 2016

4. Animation Overlays to Videos (fig. 11)

While not technically being a "chart" type in iLex, graphics video overlays look like an animation when the video is played back, thereby establishing an interesting visualization per se: One type of annotation that iLex offers is coordinates (of a point or rectangle, measured in percentages of the video resolution width and height). This is typically used to make the results from automatic 2D face and hand tracking available to the annotator as these points and rectangles tags can be superimposed to the video. Transparency ramp functions are one possibility to achieve a ghosting effect to the animated rectangles and points. When combined with a grid of positions to memorize, this results in "temporal heat map" that at least visually comes close to Dalle's idea to model signing space (cf. Braffort & Dalle, 2007).

5. Future Developments

While we expect many more chart styles being used for sign language corpus work with the technology available, it remains a goal for us to make their definition easier, by providing a query language or a query builder tool that is closer to linguistics than SQL: For written languages, there are a number of impressive examples such as ANNIS (cf. Krause & Zeldes, 2014).

Video overlays and the underlying annotation are currently restricted to two-dimensional video coordinates. This means that annotations do not transfer from one camera perspective to another which is rather annoying for the annotator. Our plan is to make three-dimensional structure annotation available and feed 3D tracking data into it. Only then it will become to use this approach to verify manual annotation by also visualising the sign trace derived from the HamNoSys notation.

We are well aware that visualisations produced so far with the tools integrated into iLex are limited to linguistic categorisations of the signed texts that make up the content of sign language corpora. To explore the content itself in various humanities research dimensions, completely different approaches what to visualize may be needed (cf. Uboldi & Caviglia 2015).

6. Acknowledgements

This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.

7. References

Braffort, A., Dalle, P. (2007). Sign Language Applications: Preliminary Modeling. *Universal Access in the Information Society*, 6(4), pp. 393--404.

Gansner, E., North, S. (2000). An open graph visualization system and its applications to software engineering. In *Software – Practice and Experi*ence, (30)11: 1203—1233.

Hanke, T. (2002). iLex. A tool for Sign Language Lexicography and Corpus Analysis. In M. González Rodriguez, & 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, pp. 923--926. [Online resource; URL: http://www.lrec-conf.org/proceedings/lrec2002/pdf/330.pdf ; last access: March 22, 2016]

Hanke, T., Storz, J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood, & E. Thoutenhoofd (Eds.), *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages.* Paris: ELRA, pp. 64--67.

Krause, T., Zeldes, A. (2014): ANNIS3: A new architecture for generic corpus query and visualization. In *Digital Scholarship in the Humanities 2014*.

Murray, S. (2013). *Interactive Data Visualization for the Web*. O'Reilly: Sebastopol CA.

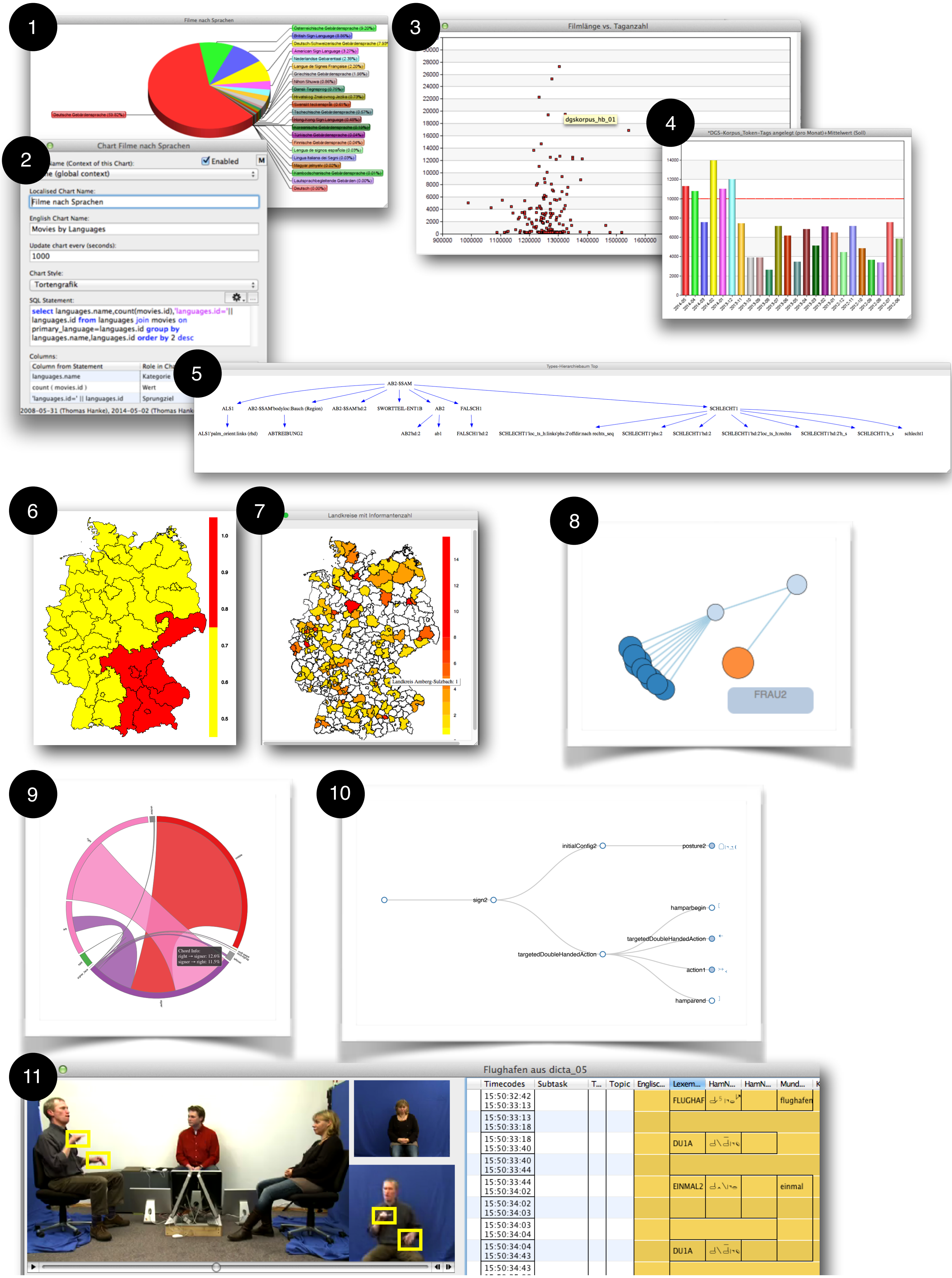
Perpiñán Lamigueiro, O. (2014). *Displaying Time Series, Spatial, and Space-Time Data with R*. Taylor & Francis. Hoboken.

Sevcikova Sehyr, Z., Caselli, N., Cohen-Goldberg, A., Emmorey, K. (2016). ASL-Lex, a Lexical Database for American Sign Language. Poster presented at the Theoretical Issues on Sign Language Research (TISLR12) Conference in Melbourne, Australia, 4-7 Jan 2016. [Online resource; URL: http://slhs.sdsu.edu/llcn/files/2016/01/TISLR2016_ASLLEX_FINAL.pdf ; last access: March 26, 2016]

Uboldi, G., Caviglia, G. (2015). Information Visualizations and Interfaces in the Humanities. In Bihanic, D. (Ed.), *New Challenges for Data Design*. Springer: London, pp. 207—218.

8. Figures Legend

Fig. 1: Pie chart, fig. 2 is the corresponding chart definition: Distribution of movies by language. Figs. 3 and 4: Scatter chart and bar chart on progress monitoring. Fig. 5: Type hierarchy. Fig. 6: Data collection (sub-) regions with informants using FRAU2, one of several signs meaning woman. Fig. 7: Regional distribution of informants in the DGS corpus project. Fig. 8 Force-directed graph showing a segment of the type hierarchy around FRAU2. Fig. 9 Chord graph showing the distribution of source and goal in directed verb by token counts. Fig. 10: Excerpt of a syntax diagram for the HamNoSys notation for the sign AB1A (away). Fig. 11: Rectangle annotation overlayed to video.


1
2
3
4
5
6
7
8
9
10
11
Filme nach Sprachen
Deutsche Gebärdensprache (59.82%)
Chart Filme nach Sprachen
Enabled
Localised Chart Name:
Filme nach Sprachen
English Chart Name:
Movies by Languages
Update chart every (seconds):
1000
Chart Style:
Tortengrafik
SQL Statement:
select languages.name,count(movies.id),'languages.id='||
languages.id from languages join movies on
primary_language=languages.id group by
languages.name,languages.id order by 2 desc
Columns:
Column from Statement
languages.name
count (movies.id)
'languages.id=' || languages.id
Kategorie
Wert
Sprungziel
Filmlänge vs. Taganzahl
dgskorpus_hb_01
Types-Hierarchiebaum Top
AB2-$SAM
ALS1
SCHLECHT1
FALSCH1
Landkreise mit Informantenzahl
Landkreis Amberg-Sulzbach: 1
FRAU2
initialConfig2
posture2
sign2
hamparbegin
targetedDoubleHandedAction
action1
hamparend
Flughafen aus dicta_05
Timecodes
Subtask
Topic
FLUGHAF
flughafen
DU1A
EINMAL2
einmal

Annotated Video Corpus of FinSL with *Kinect* and Computer-Vision Data

Tommi Jantunen[1], Outi Pippuri[1], Tuija Wainio[1], Anna Puupponen[1], Jorma Laaksonen[2]

[1]University of Jyväskylä (JyU), Sign Language Centre, P.O. Box 35, FI-40014 JyU, Finland
[2]Aalto University, Department of Computer Science, P.O. Box 15400, FI-00076 Aalto, Finland
E-mail: {tommi.j.jantunen, outi.pippuri, tuija.wainio, anna.puupponen}@jyu.fi, jorma.laaksonen@aalto.fi

Abstract

This paper presents an annotated video corpus of Finnish Sign Language (FinSL) to which has been appended *Kinect* and computer-vision data. The video material consists of signed retellings of the stories *Snowman* and *Frog, where are you?*, elicited from 12 native FinSL signers in a dialogue setting. The recordings were carried out with 6 cameras directed toward the signers from different angles, and 6 signers were also recorded with one *Kinect* motion and depth sensing input device. All the material has been annotated in *ELAN* for signs, translations, grammar and prosody. To further facilitate research into FinSL prosody, computer-vision data describing the head movements and the aperture changes of the eyes and mouth of all the signers has been added to the corpus. The total duration of the material is 45 minutes and that part of it that is permitted by research consents is available for research purposes via the LAT online service of the Language Bank of Finland. The paper briefly demonstrates the linguistic use of the corpus.

Keywords: Finnish Sign Language, corpus, annotation, grammar, prosody, *Kinect*, computer-vision

1. Introduction

This paper presents a completed set of Finnish Sign Language (FinSL) material that has been collected in the *CFINSL* project (Corpus project of Finland's sign languages)[1] and processed in the *ProGram* project (a research project that focuses on the grammatical and prosodic investigation of FinSL).[2] The material consists of signed retellings of the stories *Snowman* and *Frog, where are you?*, elicited with the help of text-less picture books from 12 native FinSL signers (8 female, 4 male; ages between 20 and 60 years) and used also in other sign language corpus projects (e.g. Johnston, 2010; Mesch, 2015). The recordings were carried out so that the signers worked in pairs in a dialogue setting in which the recording set-up consisted of 6 Full HD cameras (1920x1080, 25-50 fps) directed toward the signers from different angles (see Figure 1); 6 signers (i.e. one from each pair) were also recorded with one *Kinect* motion and depth sensing input device (see Puupponen et al., 2014). The main video material is available in H.264 compressed MP4 containers. The *Kinect* data is stored and distributed in *OpenNi*[3] and CSV formats.

All the material has been annotated in *ELAN*[4] (Crasborn & Sloetjes, 2008) for signs (see Pippuri et al., 2015), sentence-level translations (see Pippuri, 2015), clauses and their semantic–syntactic structure (including constructed action) as well as for head and body movements (see Figure 2). The annotation work has been carried out by altogether three researchers, all of whom have native competence in FinSL. All the annotations have been checked several times in order to ensure that the work is done to the highest possible standard. To further support the investigation of FinSL prosody, CSV files containing computer-vision data describing the head movements and the aperture changes of the eyes and mouth of all the signers have been linked to the material (Luzardo et al., 2014). The computer-vision data has been produced with the help of *SLMotion*[5] software, a tool specifically developed for the (semi-)automatic analysis of sign language and gestures (Karppa et al., 2014). The total duration of the material is 45 minutes. That part of it that is permitted by research consents is available for research purposes via the LAT – Language Archive Tools online service[6] of the Language Bank of Finland (Finnish *Kielipankki*).

Figure 1: The camera angles of the video material. The signers are labelled *1* and *2*. Signer 1 is always on the left and signer 2 on the right in Cam 1 view.

The details of how the video material was recorded have been presented earlier, in Puupponen et al. (2014). This paper now describes the annotation conventions of the material and the basic characteristics of the additional *Kinect* and computer-vision data. The paper also demonstrates how the present material can be used for linguistic research.

2. Basic Annotation

The basic annotation consists of meaning-based annotations for word (signs) and sentence-level units (translations). Signs are treated as relatively long units (Jantu-

[1] *https://www.jyu.fi/hum/laitokset/kielet/en/research*
[2] *http://users.jyu.fi/~tojantun/ProGram*
[3] *http://www.openni.org*
[4] *https://tla.mpi.nl/tools/tla-tools/elan/*
[5] *http://research.ics.aalto.fi/cbir/software/slmotion/*
[6] *https://lat.csc.fi/*

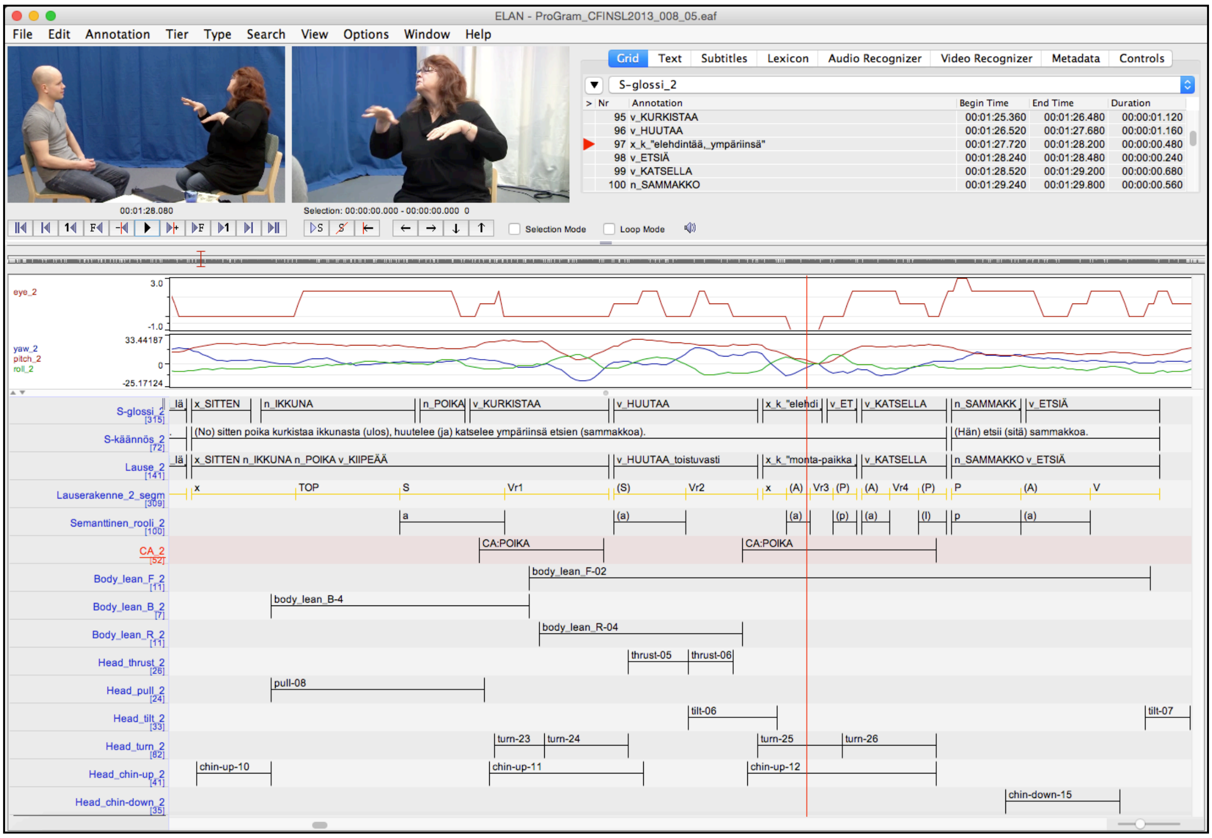


Figure 2: *ELAN* screenshot showing the computer-vision based descriptors for eye aperture changes and head movements (the time series panels), and annotations (the tiers).

nen, 2015a) and defined as distributionally free combinations of form and meaning. They are annotated on the tier *S-glossi* 'S-gloss' (belonging to the linguistic type *Gloss*) in such a way that the same sign is always given the same gloss (i.e. the annotation follows the principle of ID-glossing proposed by Johnston 2008). The glosses on the tier S-glossi (in which the capital S stands for 'systematized') fall into two main categories: *Meaning glosses* (n=3356), i.e. glosses used primarily in the annotation of lexicalized signs and described in a separate Excel-based lexicon (539 lexemes), and *Description glosses* (n=953), i.e. glosses used in the annotation of various depicting signs and gestures. Our syntax of writing both types of glosses is shown in Table 1.

Meaning glosses	prefix(es)_GLOSS(structural info to help distinguish synonyms):meaning-specifier
Examples	n_KENGÄT 'shoes' n_y_VANHEMMAT 'parents' v_JUOSTA(BB) 'run' v_EI-HALUTA 'does not want' v_ANTAA:minulle 'give to me' x_SITTEN 'then'
Description glosses	prefix(es)_k_"description"
Examples	v_k_"poika-liikkuu" 'boy moves' n_k_"puunrunko" 'shape of a trunk'

Table 1: Our syntax of writing glosses.

Glosses contain prefixed information about the major word-class of the sign (i.e. whether the sign is a nominal *n* or a verbal *v*, or unspecified *x*; see Section 7). In our meaning-based annotation (cf. form-based annotation, e.g. Johnston, 2016), the category is decided on the basis of the linguistic context (see Pippuri et al., 2015). Consequently, we distinguish, for example, between TYÖ 'work' (a nominal) and TEHDÄ 'to work' (a verbal), regardless of their fairly similar forms in FinSL.

If two or more signs are used in combination to refer to a single concept, they are analyzed as compounds (*yhdiste* in Finnish) and annotated with a single gloss, with the prefix *y* following the word-class prefix. Synonymous signs are distinguished by adding structural information (based on Rissanen, 1985) about the handshape(s), location or movement in parenthesis after the main gloss. Semantically negative signs are indicated with the Finnish verb *ei* 'no' at the beginning of the main gloss.

Pointing signs (glossed as OS 'pt', abbreviated from Finnish *osoitus* 'pointing') and the palm-up gesture (PALM-UP) are both treated as semantically and formally independent units and annotated without any information prefixed to them. However, the glosses of pointing signs may contain suffixed information about the semantics or form of the pointing (e.g. OS:minä 'me', OS:tuo 'that', or OS(B) 'pointing with a B-handshape').

Finnish translations are annotated on the level of sentences on the tier *S-käännös* 'S-translation' (belonging to the linguistic type *Translation*). As has been described in

Jantunen (2009), FinSL sentences cannot be defined comprehensively by any formal criteria. Consequently, the identification of translatable sentences has relied a lot on the intuitions and semantic insights of the annotators.

The guiding principle of the translation process has been as far as possible to maintain structural correspondence between the original signed sentence and its Finnish translation. In order to follow this principle to the maximum, the translations include additional information about what elements have been elided from the signing as well as what elements present in the Finnish translation are not expressed in the signing in the first place (both are indicated with parentheses). In addition to this, the translations also show what elements in the signing are expressed via constructed action and with other (often nonmanual) mimical behavior (this is indicated with square brackets). For example, the Finnish translation *(Poika) [menee takaisin ulos] (ja jatkaa lumiukon) tekemistä* 'the boy goes back out and continues to make the snowman' shows that the theme *poika* 'boy' is not expressed lexically, and that there is no lexical material in the original signed version that expresses the meaning 'and continues [to make] the snowman' either. The translation also shows that expression of the meaning 'going back out' relies on the use of constructed action. A detailed description of the translation procedure and conventions is presented in Pippuri (2015).

3. Syntactic and Semantic Annotation

Signs are grouped into clauses on the tier *Lause* 'clause' by following the conception of the clause summarized in Jantunen (2013, 2016). A selection of clauses that are formed around a verbal predicate are then further analyzed for their syntactic and semantic structure. The syntactic structure (for the productions of 10 signers; n=1077 clauses) is annotated on the single tier *Lauserakenne* 'clause structure'. In the standard case, clauses are analyzed into predicates (V) and their core arguments: S, the single core argument of an intransitive clause; A and P, the primary and secondary core argument of a transitive clause, respectively; and E, the third core argument of a ditransitive clause. The analysis is done also in cases where the core argument is not expressed overtly. In such cases – that is, when the semantics of the predicate requires a core argument to be present but it is not expressed – the symbol of the core argument is written in parentheses and linearized by routinely following the SV, AVP or AVPE scheme.

An exception to the standard case in the annotation of clause structure is the annotation of clauses that have a Type 3 verbal as their predicate (Type 3 verbals resemble Liddell's 2003 depicting verbs and they are annotated typically with description glosses; see Section 7). Very often such clauses are composed only of the verbal sign. Analytically, the (classifier) handshape(s) of the verbal can be treated as the core argument(s) of the clause (cf. head-marking; Nichols, 1986; Jantunen, 2008), and the core argument analysis can be extended to cover the layered nonmanual behavior, too (Ferrara & Johnston, 2014). As such core arguments are not free lexical units but rather nominal morphemes fused to the verbal head, or nonmanual gestural expressions occurring simultaneously with the Type 3 verbal, they are indicated in the annotation with lower-case letters connected to the predicate symbol (e.g. sV, aVp). A summary of the annotation symbols used in the annotation of clause-internal core elements (i.e. the predicate and its core arguments) is given in Table 2.

Symbol	Description
V	The verbal predicate of the clause.
S, A	The primary core argument of an intransitive or a transitive clause, respectively.
P	The secondary core argument of a transitive clause.
E	The third core argument of a ditransitive clause.
sV, aVp	A predicate that is a Type 3 verbal. The verbal is a well-formed clause on its own. The classifier handshape(s) of the verbal and the layered nonmanual behavior are analyzed as the core argument(s) of the clause.
( )	Parenthesis indicates that the core argument has been omitted.

Table 2: The annotation symbols of the clause-internal core elements.

In addition to the main elements described above, clauses may also contain other types of elements. The symbols used in their annotation are summarized in Table 3.

Symbol	Description
a	Typically a clause-final pointing that is co-referential with the A argument.
v	An auxiliary-like secondary predicate.
x	A syntactically peripheral element or constituent whose function/internal structure is left unanalyzed (e.g. a question sign, a conjunction sign, an adjunct).
N	A clause-internal nominal constituent that often complements the meaning of the classifier handshape(s) of Type 3 verbals.
TOP	Typically a left-detached clause-external topic constituent that sets an interpretative framework for the following clause.
e	An error or a false start.

Table 3: Additional symbols used in the annotation of clause-internal elements.

Clausal coordination (i.e. the linking of two or more clauses of the same rank) as well as subordinated complement clauses (i.e. full clauses that function typically as P arguments) have also been indicated in the annotation (for a discussion, see Dixon, 2006; Haspelmath, 2007; Jantunen, 2016). In practice, this has been done indirectly by adding extra symbols to the end of the main predicate symbol V. Clausal coordination is indicated with the lower-case letter *r* (from Finnish *rinnastus* 'coordination') and this is followed by the number of the clause in the coordinated sequence (e.g. Vr1, Vr2). The subordination of complement clauses, on the other hand, is marked in the predicates of both the matrix clause and the complement clause: in the matrix clause the predicate symbol is appended with the matrix marking letter *m* (i.e. Vm) whereas in the complement clause the added letter is the complement (Finnish *komplementti*) marking *k* (i.e. Vk; the complement clause may also be nominally headed, in which case it is annotated holistically as Nk). The symbols of coordination and subordination may be combined (e.g. Vr2m, which indicates the predicate of the second coordinate clause, also taking a clausal complement).

Two other types of coordination have also been taken into account in the syntactic annotation. These are the coordination of predicates (cf. serial verb constructions; Velupillai, 2012) and the simultaneous coordination of two Type 3 verbals/clauses, each one expressed with a different hand. The coordination of predicates is indicated simply with a number following the predicate symbol (e.g. V1, V2). The primary indicator of simultaneous coordination is the *plus* sign (e.g. Vr1+sVr2). Table 4 summarizes the symbols used in the annotation of coordination and subordinated complement clauses.

Symbol	Description
Vr	The predicate of a coordinated clause.
Vm	The predicate of a matrix clause.
Vk	The predicate of a complement clause.
Nk	A nominally headed complement clause.
V1,V2	Coordination on the level of verbal predicates is indicated with a number directly attached to the predicate symbol.
+	Two-handed simultaneous coordination of two clauses manifested as Type 3 verbals is indicated with a plus sign.

Table 4: The annotation symbols of coordination and subordinated complement clauses.

In some cases it is necessary to simply indicate that a certain sequence is a certain type of clause. The symbols used for this purpose are presented in Table 5.

Symbol	Description
[ ]	A sequence that forms a clause.
advl	An adverbial clause.
rell	A relative clause.
upol	An embedded clause.
ketl	An undefined chain of clauses.

Table 5: The annotation symbols for clause-level units.

The semantic structure of clauses has been annotated (for the productions of 6 signers) in terms of the basic semantic roles of the core arguments. The annotation is done on the tier *Semanttinen_rooli* 'semantic role', which follows the symbolic subdivision of the annotations on the tier *Lauserakenne_segm*, created on the basis of the tier *Lauserakenne* (i.e. the tiers Lauserakenne and Lauserakenne_segm contain essentially the same information, the only difference being the way the information on the tiers is structured). A summary of the semantic roles used and their symbols is given in Table 6; the roles are based on Givon (2001), Ojutkangas et al. (2009) and Velupillai (2012).

Symbol	Description
a	Agent
p	Patient
r	Recipient
e	Experiencer
b	Benefactor
i	Instrument
l	Location
s	Source
g	Goal
t	Theme

Table 6: The annotation symbols for semantic roles.

Periods of constructed action and constructed dialogue (see Ferrara & Johnston, 2014; Hodge & Ferrara, 2014) have also been annotated in the material (for the 6 retellings of the story *Frog, where are you?*); these annotations are written on the tiers CA and CD, respectively. The annotation of constructed action and constructed dialogue follows the conventions established for Australian Sign Language (Johnston, 2016). In practice, this means that the notations CA and CD are suffixed with information about whose actions or dialogue is being enacted or reported (e.g., CA:POIKA 'the actions of the boy' or CD:LUMIUKKO 'the speech of the snowman').

The linguistic type of all the tiers used in the syntactic and semantic annotation is *Research*. The only exception to this rule is the tier Lauserakenne_segm which, due to its symbolic subdivision, belongs to the linguistic type *Structure*.

4. Nonmanual Annotation

In order to facilitate research into FinSL prosody, the material has been annotated on a low level for various types of head and body movements. The annotation of head movements is roughly based on the categorization presented in Puupponen et al. (2015) and makes a distinction between 10 types of head movements (e.g. *nods, tilts*). The number of distinctively annotated body movement types in the material is 9 (e.g. *body tilts, shoulder shrug*). Table 7 presents a summary of the tiers used in the annotation of head and body movements; the linguistic type of all the tiers is *Nonmanual*.

Head movement tiers	Body movement tiers
Head_nod	Body_lean_F[orward]
Head_nodding	Body_lean_B[ack]
Head_thrust	Body_lean_R[ight]
Head_pull	Body_lean_L[eft]
Head_tilt	Body_turn_R
Head_shake	Body_turn_L
Head_turn	Body_tilting
Head_tilting	Body_shoulders_up
Head_chin-up	Body_shoulders_shrug
Head_chin-down	

Table 7: The tiers representing the low-level types of annotated head and body movements.

The nonmanual annotation has been done with the help of all the 6 camera angles (i.e. a particular movement may be visible only from one angle). That the annotations are low level means that they are not organized hierarchically into more abstract classes.

5. *Kinect* Data

One of the informants in each pair has been recorded with a *Kinect* sensor. As described in Puupponen et al. (2014), the purpose of recording *Kinect* data has been to complement the video with quantitative information about depth, a dimension not inherently present in traditional video recordings (recording done with a ceiling camera being, of course, an exception). In practice, the *Kinect* data consists of a low-quality RGB video, augmented with a 16 Hz infrared video, and a skeleton model of the signer (Figure 3). Of these, the infrared video allows one to investigate the signer's activity in the dimension of depth to the precision of one millimeter. The skeleton data, on the other hand, adds further value to the analysis of the signer's movements as it provides data analogous to that collected with motion capture equipment (see Jantunen et al., 2012).

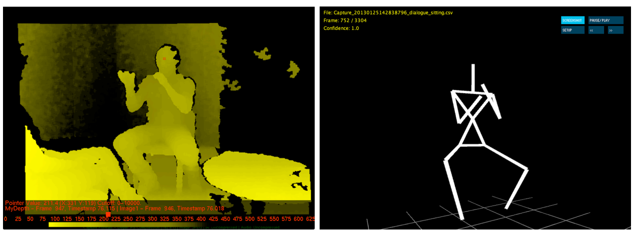


Figure 3: Screenshots from the *Kinect* data showing the infrared video (left) and a skeleton model of the signer (right) (Puupponen et al. 2014).

The *Kinect* data is not currently linked to the main data in *ELAN*. The infrared recordings can be investigated with specially coded *NiRecorder* software, based on *OpenNi* technology. The numerical skeleton model data is stored in Comma Separated Value (CSV) files, which can be easily imported into common mathematical software, such as *Matlab*, for further analysis.

6. Computer-Vision Data

A novel feature of the completed material is that the videos of each signer recorded from the near frontal angle (Cams 4 and 5, see Figure 1) have been automatically processed with computer-vision technology implemented in the *SLMotion* software specifically developed for the motion analysis of sign languages (Karppa et al 2014). With *SLMotion*, we have been able to estimate the movement of the signer's head in three dimensions as well as the relative degree of openness of the signer's eyes and mouth (Luzardo et al. 2014). This quantitative information, contained in *SLMotion* produced CSV files, has been linked into *ELAN*, where it can be visually inspected in the time series panels together with the annotations (see Figure 2). The computer-vision data can be used directly in the analysis of the interplay between prosody and syntax and the visualizations are also helpful in detecting potentially interesting sequences within the video material.

The movement of the head is estimated in three dimensions: *yaw*, *pitch*, and *roll*, which associate with turning-like movements, nodding-like movements, and tilting-like movements, respectively (Figure 4). The estimate is based on a combination of techniques which result in the detection of the signer's face and, for example, the corners of their eyes and the mouth (Luzardo et al., 2014). On the basis of this information, *SLMotion* is able to calculate the geometrical angle of the head (in the three dimensions) for each video frame (Figure 5).

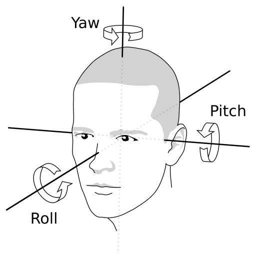


Figure 4: The three dimensions of human head motion (Jantunen et al. 2016b).

The estimation of the openness of the eyes and mouth is a classification task based on techniques similar to those used for the estimation of head movement (Luzardo et al., 2014). For the classification of eye aperture, *SLMotion* uses four classes: eyes *shut* (e.g. in blinks), eyes *squinted*, eyes *neutral*, and eyes *wide open*. The classification of mouth aperture is estimated separately in horizontal and vertical dimensions. For the horizontal dimension, the classes are *narrow*, *relaxed* and *wide*. For the vertical dimension, the classes are *closed*, *open* and *wide*.

In addition to the numerical data on the head movements and the classes of aperture changes of the eyes and mouth of the signers, *SLMotion* also indicates the exact method it has used in making the estimate. This information is valuable for research as it can be used to assess the reliability of the automatically generated estimates.

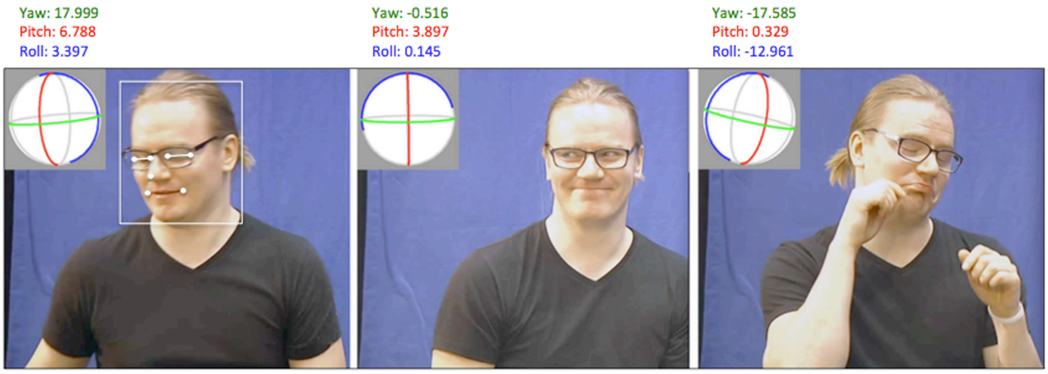


Figure 5: A visualization of the calculation of the head pose angle in yaw, pitch and roll dimensions for three different video frames (Jantunen et al. 2016a).

7. Exploitation of the Material

The material has been prepared so that it can be used in answering many types of research questions, both grammatical and prosodic. To begin with, a lot of information can already be derived from the annotations themselves. For example, as each sign is tagged for word-class, a few simple regular expression searches targeting the tier S-glossi can be used to collect information about the frequency of nominal and verbal signs in the material. The results of such a search are demonstrated in Figure 6, which shows both the overall percentual share of nominal and verbal signs in the whole material (signing from 12 signers containing altogether 4309 signs) as well as the internal composition of the classes nominal and verbal in the data. Note that in Figure 6 all the pointing signs are grouped into the class of nominals (Jantunen, 2010). On the other hand, all PALM-UP gestures are treated as unspecified in terms of category.

In research into FinSL (e.g. Jantunen, 2008, 2010, 2013, 2016), the categories of nominal and verbal have been defined by semantic and grammatical criteria such as reference (nominals refer to entities, verbals to temporally manifested dimensions of events), the marking of aspect (the markers of aspect and Aktionsart distinctions attach only to verbals), and distribution in clauses (the position of verbals in clauses is more constrained than that of nominals). Both categories can be further divided into subclasses, of which the three subclasses of verbal signs – Type 1, 2 and 3 verbals (resembling plain, indicating and depictive verbs of Liddell, 2003, respectively; see Jantunen, 2010, for a full discussion of the differences) – are the most researched ones (see Section 3).

In the future, the syntactic, semantic and nonmanual annotations will be exploited extensively in the investigation of, for example, word-order, ellipsis, syntactic functions (i.e. subject and object) and the interplay between clausal structure and constructed action in FinSL.

Thanks to the additional *Kinect* and computer-vision data, the present corpus can also be used in research into grammar and prosody in more novel ways. First, the added data can be exploited as supporting material in qualitative investigation of the interplay between syntactic structure and prosody. An example of such an investigation was the research into clausal coordination in FinSL by Jantunen (2015, 2016; also Jantunen & De Weerdt, 2016). This work explored primarily the purely grammatical properties of conjunctive ('and'), adversative ('but') and disjunctive ('or') clause linkage. However, in addition to this, the study also exploited the computer-vision data on head movements together with human-made annotations to discover systematic patterns of head motion in conjunctively coordinated complex sentences. Through a visual observation of the head movement descriptors for yaw, pitch and roll in *ELAN* (see the view in Figure 2), a recurring pattern of a back and forth head movement in the roll dimension was identified in the study. A functional analysis of this tilting-like movement revealed that signers used it to increase the prosodic cohesion of the clauses involved in the process of conjunctive coordination.

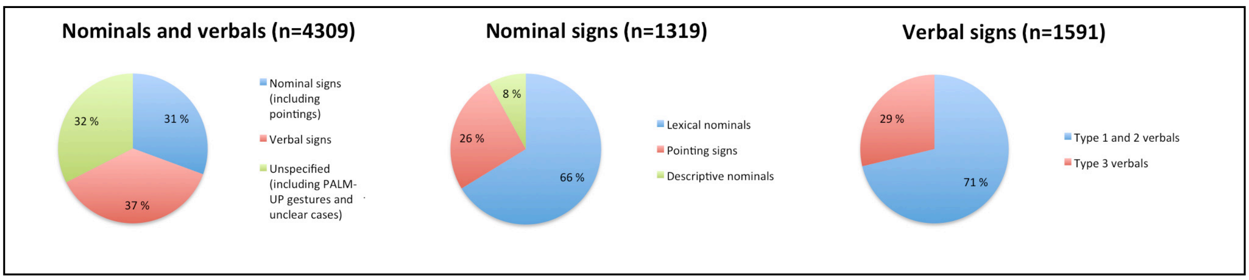


Figure 6: The distribution of nominals and verbals in the material.

The additional *Kinect* and computer-vision data can also be exploited directly in the quantitative investigation of prosody. An example of this type of study is the work by Jantunen et al. (2016ab) which investigated, with similarly collected and processed data from Swedish Sign Language (SSL), the rhythm of head movements in a small sample of semantically and structurally comparable FinSL and SSL sentences. In the study, altogether 8 FinSL and 8 SSL sentences, all extracted from retellings of the story *Snowman*, were first divided automatically into three sequences in *ELAN*. After this, the numerical signer-specific computer-vision data was used to calculate a language-specific range value for yaw, pitch and roll in each sequence. The information that was found was then used to investigate similarities and differences in the amplitude (cf. smallness, bigness) of the head movement in the sentences of the two languages.

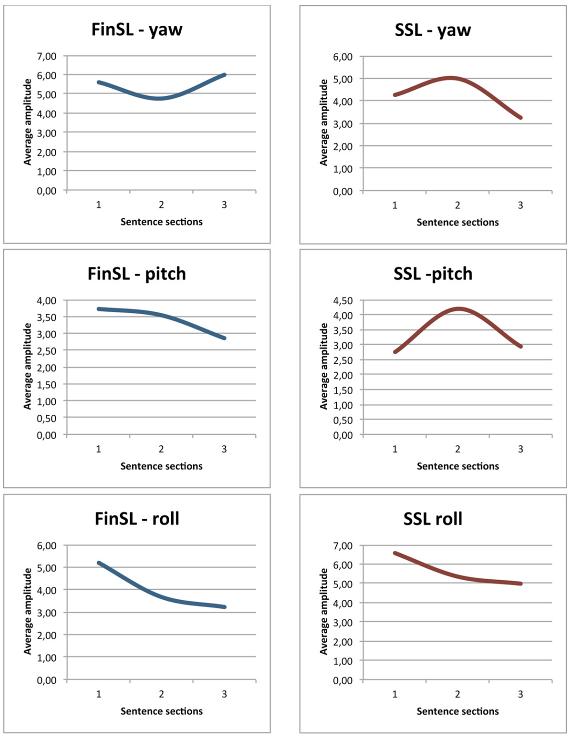


Figure 7: The average amplitude (in degrees) of the head movement in yaw, pitch, and roll dimensions for the three sections of 8+8 comparable FinSL and SSL sentences (Jantunen et al. 2016ab).

The main results of Jantunen et al.'s (2016ab) study are summarized in Figure 7. While the language-specific amplitude curves were different in both yaw and pitch dimensions (correlation co-efficient r=-0.95 and r=0.19, respectively), the curves were identical in the roll dimension (r=1.0; i.e. a perfect positive correlation). This identicality tells us that, in both languages, the movement of the head in the roll dimension was larger in the early parts of the analyzed sentences than in the final parts, and that there is a rhythmic similarity between the two languages concerning the way the head moves in the roll dimension in these sentences.

8. Conclusion

This paper has presented an annotated video corpus of FinSL to which has been added *Kinect* and computer-vision data. We have described how the material has been annotated for signs, sentence-level translations, syntactic and semantic structure, and for nonmanual activity. Moreover, we have outlined the basic characteristics of the additional *Kinect* and computer-vision data and given examples of how this material can be exploited for linguistic purposes. In the future, more layers and features will be added to the material. We are convinced that, eventually, this type of multidimensionally processed material will help us deepen our understanding of FinSL grammar and prosody by making it possible to ask completely new kinds of research questions.

9. Acknowledgements

The authors wish to thank the Academy of Finland for its financial support under grants 269089 & 273408 for *ProGram*, 140245 for *CoBaSiL*, and 251170 for *COIN*.

10. Bibliographical References

Crasborn, O. & Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, Paris: ELRA, pp. 39–43.

Dixon, R. M. W. (2006). Complement clauses and complementation strategies in typological perspective. In R. M. W. Dixon & A. Y. Aikhenvald (Eds.), *Complementation: A cross-linguistic typology*. Oxford: Oxford University Press, pp. 1–48.

Ferrara, L. & Johnston, T. (2014). Elaborating who's what: A study of constructed action and clause structure in Auslan (Australian Sign Language). *Australian Journal of Linguistics* 34(2), pp. 193–215.

Givón, T. (2001). *Syntax. Volume 1*. Amsterdam: John Benjamins.

Haspelmath, M. (2007). Coordination. In T. Shopen (Ed.), *Language typology and syntactic description: Complex constructions. Volume 2*. Second edition. Cambridge: Cambridge University Press, pp. 1–51.

Hodge, G. & Ferrara, L. (2014). Showing the story: Enactment as performance in Auslan narratives. In *Proceedings of the 44th conference of the Australian Linguistic Society – 2013*. Melbourne: University of Melbourne, pp. 372–397.

Jantunen, T. (2008). Fixed and free: Order of the verbal predicate and its core arguments in declarative transitive clauses in Finnish Sign Language. *SKY Journal of Linguistics* 21(2008), pp. 83–123.

Jantunen, T. (2009). *Tavu ja lause: tutkimuksia kahden sekventiaalisen perusyksikön olemuksesta suomalaisessa viittomakielessä* [Syllable and sentence: Studies on the nature of two sequential basic units in FinSL]. Jyväskylä Studies in Humanities 117. Jyväskylä: University of Jyväskylä.

Jantunen, T. (2010). Suomalaisen viittomakielen pääsanaluokat [The main word-classes in FinSL]. In T. Jantunen (Ed.), *Näkökulmia viittomaan ja viittomistoon* [Perspectives on sign and lexicon]. Theory and practice in applied linguistics 5. Jyväskylä: University of Jyväskylä, pp. 57–78.

Jantunen, T. (2013). Ellipsis in Finnish Sign Language. *Nordic Journal of Linguistics* 36(3), pp. 303–332.

Jantunen, T. (2015a). How long is the sign? *Linguistics* 53(1), pp. 93–124.

Jantunen, T. (2015b). Rinnastuksen prosodiaa suomalaisessa viittomakielessä [The prosody of clausal coordination in FinSL]. In M. Lehtinen & U. K. Laine (Eds.), *XXIX Fonetiikan päivät, Espoo 20.-21.3.2015, Julkaisut – Papers* [Proceedings of the XXIX Phonetics Symposium]. Tiede+teknologia [Science+technology] 7/2015. Helsinki: Aalto University, pp. 15–21.

Jantunen, T. (2016). Clausal coordination in Finnish Sign Language. *Studies in Language* 40(1), pp. 204–234.

Jantunen, T.; B. Burger; D. De Weerdt; I. Seilola & T. Wainio (2012). Experiences from collecting motion capture data on continuous signing. In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions Between Corpus and Lexicon*. Paris: ELRA, pp. 75–82.

Jantunen, T. & De Weerdt (2016). The coordination of clauses in Finnish Sign Language. Poster presented at the *12th Conference on Theoretical Issues in Sign Language Research* (TISLR 12), Melbourne, Australia, 4 January, 2016.

Jantunen, T.; J. Mesch & A. Puupponen (2016a). Aspects of the rhythm in FinSL and SSL. Paper presented at the *12th Conference on Theoretical Issues in Sign Language Research* (TISLR 12), Melbourne, Australia, 6 January, 2016.

Jantunen, T.; J. Mesch; A. Puupponen & J. Laaksonen (2016b). On the rhythm of head movements in Finnish and Swedish Sign Language sentences. To appear in *Proceedings of the Speech Prosody 2016*, to be organized in Boston University, May 31–June 3, 2016.

Johnston, T. (2008). Corpus linguistics and signed languages: No lemmata, no corpus. In the *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, Paris: ELRA, pp. 82–87.

Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics 15*(1), pp. 106–131.

Johnston, T. (2016). *Auslan Corpus annotation guidelines*. Macquarie University (Sydney) & La Trobe University (Melbourne). February 2016 version.

Karppa, M.; V. Viitaniemi; M. Luzardo; J. Laaksonen & T. Jantunen (2014). SLMotion: An extensible sign language oriented video analysis tool. In *Proceedings of the 9th International Conference on Language Resources and Evaluation* (LREC'14), Paris: ELRA, pp. 1886–1891.

Liddell, S. K. (2003). *Grammar, gesture and meaning in American Sign Language*. Cambridge: Cambridge University Press.

Luzardo, M.; V. Viitaniemi; M. Karppa; J. Laaksonen & T. Jantunen (2014). Estimating head pose and state of facial elements for sign language video. In *Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel*, Paris: ELRA, pp. 105–112.

Mesch, J. (2015). Svensk teckenspråkskorpus – dess tillkomst och uppbyggnad [Building the corpus for Swedish Sign Language]. In *Forskning om teckenspråk XXIV* [Research on sign language XXIV]. Stockholm: Stockholm University, Department of Linguistics, pp. 1–25.

Nichols, J. (1986). Head-marking and dependent -marking grammar. *Language* 62(1), pp. 56–119.

Ojutkangas, K.; M. Larjavaara; M. Miestamo & J. Ylikoski (2009). *Johdatus kielitieteeseen* [Introduction to linguistics]. WSOY: Helsinki.

Pippuri, O. (2015). *Annotointi ja käännöstyö ProGram-projektissa* [Annotation and translation process in ProGram project]. Thesis in the degree programme in sign language interpreting, Humak University of Applied Sciences.

Pippuri, O.; J. Salonen; R. Takkinen; A. Puupponen & T. Jantunen (2015). Annotation conventions in ProGram and CFINSL projects. Poster presented at the *Digging into Signs workshop*, University College London, March 30–31, 2015.

Puupponen, A.; T. Jantunen; R. Takkinen; T. Wainio & O. Pippuri (2014). Taking non-manuality into account in collecting and analyzing FinSL video data. In *Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel*, Paris: ELRA, pp. 143–148.

Puupponen, A.; T. Wainio; B. Burger & T. Jantunen (2015). Head movements in Finnish Sign Language on the basis of motion capture data: A study of the form and function of nods, nodding, head thrusts, and head pulls. *Sign Language & Linguistics* 18(1), pp. 41–89.

Rissanen, T. (1985). *Viittomakielen perusrakenne* [The basic structure of sign language]. Publications of the Department of Linguistics 12. Helsinki: University of Helsinki.

Velupillai, V. (2012). *An introduction to linguistic typology*. Amsterdam: John Benjamins.

11. Language Resource References

ELAN. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. URL: *http://tla.mpi.nl/tools/tla-tools/elan*.

ProGram-aineisto. Lumiukko- ja sammakkotarinat. Versio 1 [ProGram data. The stories Snowman and Frog, where are you?. Version 1]. University of Jyväskylä, Sign Language Centre, Finland. URL: *https://lat.csc.fi*.

SLMotion. Aalto University, Department of Computer Science, Espoo, Finland. URL: *http://research.ics.aalto.fi/cbir/software/slmotion/*.

Methods for Recognizing Interesting Events within Sign Language Motion Capture Data

Pavel Jedlička, Zdeněk Krňoul, Miloš Železný
University of West Bohemia, Faculty of Applied Sciences, NTIS - New Technologies for the Information Society
Univerzitní 8, Pilsen, Czech Republic
jedlicka@kky.zcu.cz, zdkrnoul@kky.zcu.cz, zelezny@kky.zcu.cz

Abstract

Rising popularity of motion capture in movie-production makes this technology more robust and more accessible. Utilization of this technology for sign language capturing and analysis is evident. The article deals with the usability of the motion capture in creating sign language corpora. A large amount of the data acquired by the motion capture has to be processed to provide usable data for wide range of research areas: e.g. sign language recognition, translation, synthesis, linguistics, etc. The aim of this article is to explore possible methods to detect interesting events in data using machine learning techniques. The result is a method for detection of the beginning and the end of the sign, hand location, finger and palm orientation, whether the sign is one or two handed, and symmetry in the two-handed signs.

Keywords: motion capture, machine learning, creating sign language corpora

1. Introduction

In these days signing language translation or TV broadcast is provided by humans. Sign language (SL) synthesis is considered as supplementary communication means of the deaf individuals. There are SL approaches intended for creating sign language in an artificial way. One perspective technique is using virtual 3D character animation as a signing avatar (Krňoul et al., 2008). However, there is still poor realism of such produced character animation compared to the standard SL video of signing subject causing overall unacceptability of the signing avatars by the deaf community. A huge disadvantage of image processing is that computer vision is a very complex task for the SL videos. Image recognition of the position of body parts such as arms, hands, and handshapes is possible, but accuracy is far worse than in using motion capture.

One reason is that artificial signing avatars are not able to sign fluently and naturally and, therefore, it is difficult or uncomfortable to understand them. On the other hand, replaying an utterance in 3D animation generated from the motion capture of sign language speaker provides very natural outcome because the captured motion copies movements of the SL subject. Such continuous data reflects a certain number of still unidentified phenomena of SL production system. Therefore, integration of high-quality motion capture data is essential for any further research and gives certain assumptions to provide accessible sign language synthesis (Huenerfauth et al., 2015).

The full body capture including hand, finger, facial expression and eye gaze movements is a condition to collect spatial-temporally synchronous records of all the channels (Gibet et al., 2015). However, for such complex recording, the motion capture hinders movements of subject's body so it does not have to compose a natural move. Moreover, an interconnection of SL annotations and motion capture data seem to be a crucial issue (Lu and Huenerfauth, 2012), (Gibet et al., 2015). Hereby, analyses of the motion capture data are often taken into account in limited short time intervals.

An analysis of 50 minutes of videos combined with motion captured data from French sign language corpora was conducted to extract low-level or high-level motor schemes (Gibet et al., 2012). There is incorporation of an automatic segmentation technique of the short hand-shape sequence (Heloir et al., 2006), a statistic analysis of phasing between hand motion and handshapes, categorizing of hand motion velocity profiles within signs and during sign transitions.

In the paper, we present initial experiences in the full body motion capture of Czech Sign Language interpreter. Each lexical item from a dictionary is produced when the signer's hands are returned to a relax-pose between the items during recording. The new technique for the motion capture data processing is presented to explore capabilities of automatic identification of start and end pose of the signs. In the context of the SL recording scenario, the experiment is uncovering helpful aspects that can lead to further inter-sing segmenting techniques of the SL motion capture or video data.

2. Sign language motion capturing

The popularity of using motion capture systems in many different tasks causes this system to be more accessible for non-commercial subjects. It also causes this technology to improve more and becoming more precise. There are more different systems using different technology for motion capturing (Hasler et al., 2009). These different approaches are optical, gyroscopic, mechanical, etc.

The optical system was chosen because the signing subject is not wearing any special suit that limits his or her natural movement. The marker-based system was chosen for its higher precision compared to non-marker approaches.

2.1. Initial experiences

The VICON system was chosen as a main motion capture technology of the data acquisition for the sign language synthesis task. The VICON motion capture system is based on the principle of high-frequency cameras measuring a motion of passive spherical retroreflective markers in the infrared spectrum. However, there are some limitation factors given by the capturing principle for the finger movements and the handshapes used in sign languages. The first factor is the number of cameras. We found that it is sufficient to use eight cameras for an accurate and robust motion capturing of the body torso, arms and head of a signing subject. In this case, according to our experience, a standard set of optical markers is sufficient to exactly capture overlapping arms as well as other hand/body contacts that widely occur in the sign languages.

However, the simultaneous capturing of the fingers of both hands and the rest of the body requires at least 30 additional markers. The markers have to be smaller compared to the different proportion of the fingers and the rest of the body and also they must be somehow rigidly attached. For example, they can be mounted on a conventional glove like the other body markers. But in this case, we observed negative effects of such fixation. Mainly for a smaller hand, the markers were not rigid to the particular finger segment during its bending. Although this is a relatively small movement, it causes an inaccuracy in the identification of a model internally used by the VICON system. The problem can be eliminated so that the markers are attached directly to the finger skin. In this case, however, their unwanted loss caused by frequent touches of the hands while signing is not excluded. It was also observed that there is higher speed of marker movements mainly for fingertips, which requires higher camera frame rate than that for tracking other parts of the body.

The main limitation factor is the tracking of the finger markers that are close to each other and significantly increase overlapping situations (frames with marker swaps), especially during the hand contacts. These problems can be partially solved by the good positioning of the cameras, but this leads to increasing the number of cameras to 20 or more which can be of course expensive and difficult solution with limited functionality in the case of the full marker occlusions.

2.2. Combining optical and data glove record

The combination of the aforementioned optical and the data glove motion capturing is an alternative recording technique. The measurement principle of the finger bending is based on the resistive sensors that provide robust measurements of finger contacts on one hand or mutually between hands. In addition, CyberGlove3 glove measures palm flex and wrist rotation (pitch and yaw). On the other hand, the reading of one sensor is relative to the reading of the preceding finger segment or the wrist and thus we do not get absolute 3D position. Thus, the CyberGlove3 motion capture data are relative to the 3D position of the forearm.

3. Dataset

Data were acquired by VICON motion capture system using 8 T-20 cameras. The T-20 camera has 2 Mpx resolution and it is possible to record at a speed up to 690 frames per second (fps). Recording, reconstruction, and data post-processing were made in Blade software from VICON. This software provided also a body model. Motion capture of the handshapes was recorded simultaneously using Cybergloves3 based on flex sensor technology. There was also the availability of facial motion data by VICON motion capture Cara. It is a marker-based motion capture system using 4 cameras aimed at the face of the signing subject. It is possible to track tens of markers placed on the face, lips, and even eyelids. But this was not involved due to higher demands on recording procedure and research purposes of the dataset.

The dataset used for this research contains two hours of signing. For motion-capturing were placed 53 passive 14 mm markers on the body of the signing subject. Used marker setup contains 10 markers on each arm and 15 on the torso and head providing the possibility to track any general movement of the whole upper body.

The subject signed about 1000 dictionary signs in Czech Sign Language and each individual sign was recorded separately starting and ending in the relax-pose. This restriction was chosen for more robust separating of single signs and it does not affect the quality of this particular research. Motion capture frame rate was set to 120 fps. This rate was accepted as sufficient because movements with faster changes were not observed. Higher frame rate is not requisite as the amount of data increases significantly. Part of the dataset was manually segmented by two different persons for further evaluation.

4. Automatic feature detection

The purpose of the first developed method is to automatically detect the relax-pose to separate individual signs. The sign segmented this way is surrounded by resting in the relax-pose and there is a characteristic movement of the signing subject when leaving the relax-pose and moving hands to start point and when returning back to the relax-pose from the end point. We developed the estimating method using this feature to determine the time stamp of the beginning and the end of the sign. The segment acquired by using this method was used for further analysis. In this article, we also focused on events in a starting point configuration (hand location). An important characteristic of sign language is the dominance of one the signing subject's hands. There is only one signing subject in used dataset and it is a priori known which of her hands is dominant. But it is also possible to recognize this information automatically simply by measuring the length of the trajectory of each hand. The dominant hand is apparently the one which moves the longer distance than the other.

4.1. Relax-pose detection

As it was mentioned, the first step of data processing was detecting the relax-poses to separate signs in the record. The relax-pose was defined as a position of hands freely hanging in front of the stomach. The beginning and the end of the relax-pose was detected by positioning the dominant hand in the expected area and by the decrease of the speed of this hand (particularly wrist joint) below the threshold. The speed v is measured as a difference of the position of dominant hand's wrist joint in two following frames:

$$dx(n) = x(n) - x(n+1)$$
$$dy(n) = y(n) - y(n+1)$$
$$dz(n) = z(n) - z(n+1)$$
$$v(n) = \sqrt{dx^2(n) + dy^2(n) + dz^2(n)},$$

where $x(n)$, $y(n)$, $z(n)$ are positions in frame n. Origin is placed on the ground, positive y-axis leads upwards, positive z-axis leads forwards, and x-axis leads on the right, all from the subject's orientation.

Area boundaries for each axis and the speed threshold were determined by supervised learning on the part of the data. An example of manual and automatic segmentation is in Figure 1.

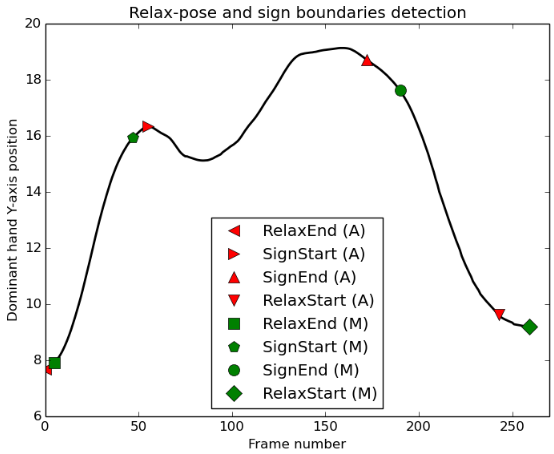


Figure 1: Automatic (A) and manual (M) detections of start and end points for both relax-pose and sign visualized for y-axis position of the dominant hand.

4.2. Sign-beginning detection

There is a specific movement from the end of the relax-pose to the beginning of sign. This movement seems to be more fluent than the movement during the sign. The automatic detection of the sign-beginning is based on measuring acceleration and deceleration of the dominant hand.

All data contains some low-level noise caused by the environment during recording. This noise doesn't affect detections based on position and speed but it causes problems in detection of the features in acceleration. It is necessary to filter acceleration signal before detecting points of interest with lowpass filter. Filter parameters were experimentally chosen corresponding to recording frame rate.

As the dominant hand is leaving the relax-pose acceleration increases. The hand is decelerating when approaching the hand location. Subject starts signing after that movement and, therefore, accelerates his or her hand again, in other words, the point of the second acceleration of the dominant hand is the hand location. Acceleration a is defined:

$$a(n) = v(n+1) - v(n).$$

Values of the speed and the acceleration for the same example as in Section 4.1. is shown in Figure 2. This acceleration-based approach was successful in most cases but there was a phenomenon in some signs which caused to trigger detection too early. It was caused by a sudden change of acceleration during the movement from the relax-pose to the sign location. This problem was solved by adding a maximum speed threshold as a parameter for the sign location detection.

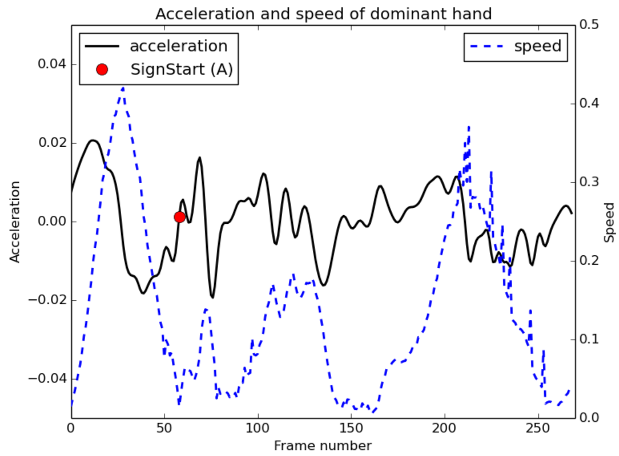


Figure 2: Speed and acceleration values.

4.3. Recognising features

There are some sign types which can be simply recognised e.g. one or two handed sign. Other observable specifications are different types of symmetry, static and dynamic signs, etc.

4.3.1. One handed sign detection

It was already mentioned that the signing subject has his or her dominant hand. Detection of the dominant hand is described in Section 4. The next step in feature recognition is to decide whether the sign is one or two handed. The detection is based on the same principle as the relax-pose detector but in this case it is focused on the secondary hand assuming that the dominant hand is signing.

4.3.2. Sign location detection

Many sign language notification systems describing handshape and location at the beginning of the sign. While the beginning of the sign is detected, it is simple to describe hand location and wrist orientation because the motion capture data already contains this type of information. There is no difference in one or two-handed signs, because the moment of the beginning of the sign is same for both hands. It is simple to acquire data for both hands.

4.3.3. Symmetry detection

Another interesting feature is the symmetry of two-handed signs. There are two types of symmetry. The natural type of symmetry is mirror movement of both hands. Inverse symmetry is when both hands start in mirror symmetry but each hand moves in opposite direction. In our research, it makes no difference whether the handshapes are the same or not, but this information can be added using data from finger motion capture provided by CyberGloves3. The initial experiment was made by correlating movements of both hands. The correlation alone is not robust enough for detection of symmetry and further research is needed.

5. Results and future work

The results were validated on two sets manually segmented by two persons. Each set contained 20 signs. Both manual segmentations were compared to each other for defining the cross-annotation difference. Only one of the sets (set 2) was used for supervised learning incorporated in detection techniques.

set	r-p end	sign start	sign end	r-p start
set 1	9.2	14.3	9.15	9.95
set 2	9.75	8.2	13.95	8.65

Table 1: Manual segmentations comparison.

Set 1	r-p end	sign start	sign end	r-p start
man 1	7.85	11.7	19.65	25.1
man 2	6.45	14.8	14.7	27.75

Table 2: Automatic segmentation validation on dataset 1.

Set 2	r-p end	sign start	sign end	r-p start
man 1	9.1	11.1	15.0	20.5
man 2	5.95	14.9	18.45	22.0

Table 3: Automatic segmentation validation on dataset 2.

The results are summarized in Tables 1, 2, and 3. The columns correspond to the relax-pose end, sign-beginning, sign-end, and relax-pose beginning. Values in the rows correspond to average absolute frame difference. It can be observed that the difference of two manual segmentations and the difference of automatic and manual segmentations are similar. It should be reminded that the frame rate of the record is 120 fps. This means that 1 frame difference equals 8.33 miliseconds. Human eye is not able to recognize framerate 24 fps which is framerate of video. Standard video frame length equals approximately 5 frames in used motion capture.

The results for recognizing important events such as an end of the relax-pose and sign beginning are very satisfactory because the difference between the automatic and the manual segmentation tends to be slightly lower compared to two manual segmentations. Worse results in the recognising end of the sign seem to be caused by not well-bounded signs at its end. Signing subjects tend to lose his or her hands during the end of the sign fluently. Human segmentation is more or less intuitive for this feature. On the other hand, automatic segmentation reflects more on distinct events in the data.

The relax-pose beginning detection results are not satisfactory. Automatic segmentation triggers when the dominant hand's speed decreases below threshold but manual segmentation tends to trigger earlier. This may be caused by the fact, that the human validator knows that the sign will end soon and he or she does not wait until hands stay still. The question is which segmentation is better and whether this difference means that automatic segmentation is better than human. Anyway, the beginning of the rest pose is the least important event of four evaluated features and does not cause any transferred inaccuracy.

In further work, we will focus on different approaches to sign segmentation as well as on sign location analysis. The next step is fluent sign speech analysis. The long term goal is data-driven sign language synthesis.

6. Acknowledgements

This research was supported by the Ministry of Education, Youth and Sports of the Czech Republic project No. LO1506 and by the grant of the University of West Bohemia, project No. SGS-2016-039.

7. Bibliographical References

Gibet, S., Marteau, P.-F., and Duarte, K. (2012). Toward a motor theory of sign language perception. In *Proceedings of the 9th International Conference on Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*, GW'11, pages 161–172, Berlin, Heidelberg. Springer-Verlag.

Gibet, S., Lefebvre-Albaret, F., Hamon, L., Brun, R., and Turki, A. (2015). Interactive editing in French Sign Language dedicated to virtual signers: requirements and challenges. September.

Hasler, N., Rosenhahn, B., Thormahlen, T., Wand, M., Gall, J., and Seidel, H. P. (2009). Markerless motion capture with unsynchronized moving cameras. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 224–231, June.

Heloir, A., Gibet, S., Multon, F., and Courty, N. (2006). Captured motion data processing for real time synthesis of sign language. In *Proceedings of the 6th International Conference on Gesture in Human-Computer Interaction and Simulation*, GW'05, pages 168–171, Berlin, Heidelberg. Springer-Verlag.

Huenerfauth, M., Lu, P., and Kacorri, H. (2015). Synthesizing and evaluating animations of american sign language verbs modeled from motion-capture data. In *Proceedings of SLPAT 2015*, pages 22–28, Dresden, Germany, September. Association for Computational Linguistics.

Krňoul, Z., Kanis, J., Železný, M., and Müller, L. (2008). Czech text-to-sign speech synthesizer. *Machine Learning for Multimodal Interaction, Series Lecture Notes in Computer Science*, 4892:180–191.

Lu, P. and Huenerfauth, M. (2012). Learning a vector-based model of american sign language inflecting verbs from motion-capture data. In *Proceedings of the Third Workshop on Speech and Language Processing for Assistive Technologies*, SLPAT '12, pages 66–74, Stroudsburg, PA, USA. Association for Computational Linguistics.

Centroid-Based Exemplar Selection of ASL Non-Manual Expressions using Multidimensional Dynamic Time Warping and MPEG4 Features

Hernisa Kacorri[1], Ali Raza Syed[1], Matt Huenerfauth[2], Carol Neidle[3]
[1]The Graduate Center, CUNY, Computer Science Program, New York, NY, USA
[2]Rochester Institute of Technology, Golisano College of Computing & Information Sciences, Rochester, NY, USA
[3]Boston University, Linguistics Program, Boston, MA, USA
{hkacorri,asyed2}@gradcenter.cuny.edu, matt.huenerfauth@rit.edu, carol@bu.edu

Abstract

We investigate a method for selecting recordings of human face and head movements from a sign language corpus to serve as a basis for generating animations of novel sentences of American Sign Language (ASL). Drawing from a collection of recordings that have been categorized into various types of non-manual expressions (NMEs), we define a method for selecting an exemplar recording of a given type using a centroid-based selection procedure, using multivariate dynamic time warping (DTW) as the distance function. Through intra- and inter-signer methods of evaluation, we demonstrate the efficacy of this technique, and we note useful potential for the DTW visualizations generated in this study for linguistic researchers collecting and analyzing sign language corpora.

Keywords: American Sign Language, Non-Manual Expressions, Dynamic Time Warping, Exemplar Selection, Animation Synthesis

1. Introduction

Technology to partially automate the process of producing animations of a virtual human character producing American Sign Language (ASL) could make it easier and more cost-effective for organizations to provide sign language content on websites. As compared to providing videos of ASL, animations that are automatically synthesized from a symbolic specification of the message would be easier to update and maintain, as discussed in (Huenerfauth, 2004; 2008). In this study, we examine whether multidimensional dynamic time warping (DTW) is suitable for evaluating the similarity of recordings of face and head movements of ASL non-manual expressions (NMEs).

For the purposes of generating animations of ASL, given a set of recordings of human face movements for ASL NMEs, it is valuable to identify an exemplar recording that could be used as the basis for generating the movements of virtual human character, to produce an understandable ASL animation containing an NME. The goal of the current study is to evaluate the potential of centroid-based exemplar selection for ASL NMEs. Specifically, we are investigating whether, given a set of recordings of humans producing some category of ASL NME, a multidimensional DTW metric can serve as the basis for selecting a "centroid," a member of the set with the minimum cumulative pairwise distance from the other members, such that this centroid serves as a representative exemplar of the set. Given idiosyncratic differences among individual productions, which are naturally found in any collection of recordings of ASL NMEs, it is reasonable to expect that some recordings may be outliers, and others may be more similar to the other items in the set, where "outlier" is defined as a member of the set with the maximum cumulative pairwise distance from the other members.

In this paper, we present an algorithm for selecting a centroid exemplar of a sign language NME from a set of human recordings (and for comparison sake, we also define a method for determining the maximal outlier in such a set of recordings). Through various forms of intra-signer and inter-signer comparison, we evaluate the effectiveness of this technique for identifying an NME performance that is typical of a specific linguistic type. In addition, we present several forms of data visualization based on this algorithm, some of which may be useful for linguistic researchers who are collecting or analyzing sign language corpora.

2. Background and Related Work

An ASL production consists of movements of the eyes, face, head, torso, arms, and hands; in fact, the movements of the face and head, during non-manual expressions (NMEs), are essential to the meaning of utterances. For example, these expressions can convey grammatical information about individual words or entire phrases or clauses during the utterance. The upper face and head movements of these NMEs occur in parallel with phrases containing manual signs (Neidle et al., 2000).

In current collaborative work involving RIT, Boston University, and Rutgers University, we are videorecording and annotating a set of human ASL productions, including several categories of syntactic NMEs. These recordings include markers of wh-questions, yes/no-questions, rhetorical questions, negation, and topics. This dataset has been essential for research on automatic ASL recognition technologies (Neidle et al., 2014), and RIT researchers are using this dataset for research on ASL animation synthesis. These recordings serve as the source of human movement data for the methods presented in section 3.

Several groups of researchers internationally have investigated how to generate animations of sign language that include linguistically meaningful facial expressions; we have compared and surveyed their methods and contributions in (Kacorri, 2015). In the work most closely related to the methods discussed in section 3, researchers have investigated the potential of clustering and centroid selection for identifying variants of German Sign Language (DGS) lexical items, based on the co-occurring lexical NMEs, primarily given differences in mouthing (Schmidt et al., 2013). However, the potential for centroid-based exemplar selection for syntactic NMEs (consisting of head and upper face movements and spanning over one or more glosses) has not been previously investigated.

2.1 Dynamic Time Warping (DTW)

DTW is utilized as the distance metric for the methods in section 3. DTW is a methodology commonly used to evaluate similarity among time-series data, e.g. (Sakoe and Chiba, 1978), and has been previously adopted to evaluate the similarity of animated characters' facial expressions (Mana and Pianesi, 2006; Ouhyoung et al., 2012).

The rate at which ASL NMEs change may vary by signer or context. For example, in Figure 1, the head roll movement in Recording 2 peaks first, then falls and rises faster than the movements in Recording 1. Comparing the two series requires alignment to match points by, for example, finding corresponding points between peaks and valleys as well as the rising and falling portions of the two series. The two curves can be aligned through a nonlinear stretching and compressing the time axis ("warping"). The distance between the two series is taken to be the sum of distances between the matching points in the warped time domain. DTW yields an optimal alignment by using dynamic programming to determine a warping function that minimizes the total distance between the series.

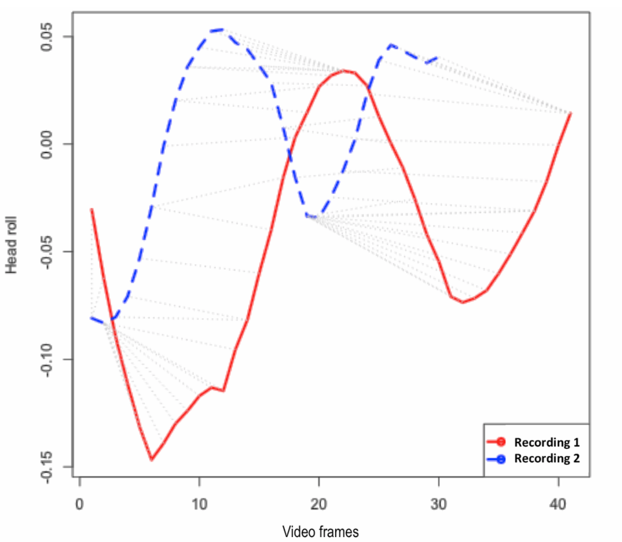


Figure 1: Example of DTW alignment between the "head roll" values detected during two Negative NMEs extracted from human ASL recordings in our dataset.

In a prior study (Kacorri and Huenerfauth, 2015), we demonstrated the potential of DTW for evaluating the similarity of face and head movements for ASL NMEs. We examined whether the judgments of native ASL signers about the quality of animations correlated with DTW-based similarity between that animation and a gold-standard human recording that had been analyzed to yield a stream of MPEG4 Facial Action Parameters (FAPs) (Pandzic and Forchheimer, 2003), representing the movements of the face. We found a significant correlation, indicating that DTW may be useful for automatically evaluating the similarity of ASL NMEs. In that work, we used the normalized distance from a multidimensional variant of DTW (Giorgino, 2009) on the parallel streams of MPEG4 values. This multidimensional DTW is also used in the centroid exemplar-selection methods in section 3.

3. Centroid-Based Exemplar Selection

This section describes our method for selecting an exemplar recording of a human ASL NME from a dataset containing multiple examples of each type of NME. As discussed in section 3.2, this method operates by identifying a centroid item within the set, using normalized multivariate DTW as the distance function.

3.1 NME Dataset

Our data for this study consists of the recordings of human ASL productions that were collected and analyzed at Boston University, as discussed in section 2. We analyzed 173 of these annotated video recordings of a female ASL native signer using an MPEG-4 face tracker (Visage Technologies, 2016), and we extracted the head pose and MPEG-4 facial features for each video frame. Since some recordings contained more than one NME, our dataset included a total of 199 multivariate time series of syntactic NMEs distributed in the following categories: wh-questions (14), yes/no-questions (21), rhetorical questions (13), negation (55), and topics (96). As shown in Table 1, there was a high variability across recordings in video length and number of manual glosses (individual signs performed on the hands) that occurred in parallel with each NME.

Given that the NME categories we are investigating mostly involve head and upper face movements, e.g. (Neidle et al., 2000), for this study we are interested only in a subset (a total of 9) of the extracted features by Visage face tracker, which includes:

- **Head orientation** (FAP48-FAP50): orientation parameters defined as pitch, yaw, and roll.
- **Eyebrow vertical displacement** (FAP31-FAP36): 6 parameters describing vertical movements of the inner, middle, and outer points of the left and right eyebrow.

NME Dataset (Num. of examples)	**Video Frames** min – max (mean)	**Num. of Glosses** min – max (mean)
Topic (96)	5 – 54 (15.5)	1 – 4 (1.43)
Negation (55)	10 – 76 (38.1)	2 – 7 (3.56)
Y/N-question (21)	9 – 78 (34.6)	2 – 6 (3.6)
Wh-question (14)	15 – 69 (31.2)	1 – 5 (2.2)
Rhetorical (13)	11 – 46 (28.3)	1 – 4 (3.0)

Table 1: NME Dataset Characteristics.

3.2 Selecting Centroid and Outlier

We used multivariate DTW (Giorgino, 2009) to obtain the normalized distances between all pairs of recordings in the same NME category. Within each NME category, we labeled one recording as the "centroid" and one recording as the "outlier", which were defined in the following way:

$$\text{centroid} = \arg\min_{u \in S} \sum_{v \in S} \text{DTW}(u, v)$$

$$\text{outlier} = \arg\max_{u \in S} \sum_{v \in S} \text{DTW}(u, v)$$

S is the set of all recordings within an NME category. Thus "centroid" and "outlier" are the recordings with the minimum and maximum cumulative DTW distance, respectively, to all other recordings within a category. The centroid recording is the most representative example from a given category since it characterizes the central (median) tendency of recordings within that category. Conversely, the outlier characterizes the least representative recording in a given category.

As suggested by prior work, e.g. (Gillian et al., 2011), preprocessing is necessary for DTW if either (a) the source range of the N-dimensional data varies or (b) if invariance to spatial variability and variability of signal magnitude is desired. To address the first case, we scaled all the features to the range [-1, 1] by dividing by the largest maximum value in each feature. For DTW analysis, such scaling is suitable for data that is already centered at zero, which is the case for our extracted MPEG-4 data, since 0 denotes a neutral pose for each feature. To address the second case, we performed z-normalization, so that the time-series for each FAP would have a zero mean and unit variance. Since DTW only performs alignment in the time dimension, series with very different amplitudes may not allow for proper comparisons when using the DTW similarity measure. Normalizing the amplitude values brings all series to a standard scale and allows for better similarity measures to be determined.

Table 2 provides details about the centroid and outlier recordings that were identified using the above procedure. The table includes information about the length of each NME, as indicated by the number of video frames in duration and the number of manual glosses that occur in parallel with each NME.

NME Category	Centroid		Outlier	
	v. frames	glosses	v. frames	glosses
Topic	19	2	13	2
Negation	13	3	30	3
Y/N-question	39	4	20	2
Wh-question	31	2	15	1
Rhetorical	12	1	40	4

Table 2: Number of frames and glosses for the centroids and outliers that were selected.

For example, for the "topic" category of NME, the recording that was selected by the algorithm as the centroid contained the utterance: "CAR BREAK-DOWN WAVE-NO," in which the topic NME occurred during the two glosses "CAR BREAK-DOWN."

3.3 Visualizing Centroids Versus Outliers

Figure 2a visualizes the DTW distance among all pairs of recordings in the set of "topic" ASL NME recordings. Each node in the graph represents a recording, and each edge, the DTW distance between the nodes. Nodes are numbered based on their listing in the dataset. Lighter colors for nodes and edges denote smaller DTW distances, thus more similar time-series values. The graphs were produced using the Python package NetworkX (Hagberg et al., 2008) with the Fruchterman Reingold layout and the Viridis color-map. Since the software default layout locates the nodes with the highest degree in the center, the input for the algorithm was the DTW distance matrix for a comparison set where each of the DTW distances is replaced with its absolute difference with the max distance. Thus the node in the center is the centroid with smallest total DTW distance to its neighbors. Figure 2b represents the cumulative distances for each node, with arrows indicating the centroid and outlier for this dataset.

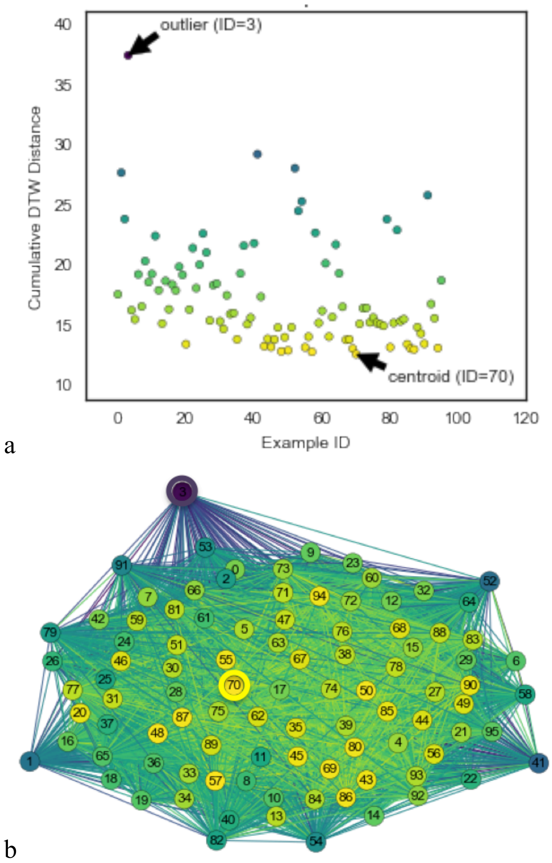


Figure 2: Visualization of (a) cumulative DTW-distances and (b) DTW-distance graph for the video recordings in the ASL Topic NME dataset. The centroid and the outlier are indicated in both graphs, by the brightest and the darkest color, respectively.

In Figure 2, the reader should note that the node with the brightest yellow color and most central location of the graph image visually indicates the "centroid" of this group, and the node with the darkest color and most remote location in the graph is the "outlier." In addition, to aid visibility, the centroid and outlier have been surrounded by a thick outline in Figure 2b.

4. Initial Confirmation of Centroid Quality Using Intra-Signer Data

As a preliminary assessment of the validity of the above procedure, we compared the "centroid" and the "outlier" we identified, using a methodology inspired by prior DTW research on template-based classification (Gillian et al., 2001). After identifying the centroid and the outlier for each ASL NME dataset, we constructed a classifier for assigning a label (wh-question, yes/no-question, rhetorical question, negation, or topic) to a given recording of unknown NME category. We treated the centroid for each ASL NME category (topic, negation, etc.) as an "exemplar" of that category. To classify some given recording of unknown category, we compared its distance to each of the five exemplars. The unknown recording was labeled with the category of the exemplar to which it had the minimum DTW distance. For sake of comparison, we also constructed a second classifier based on using the "outliers" identified for each NME category as if they were an exemplar of that category.

To evaluate these two classifiers, we removed the five centroids and the five outliers from the five NME datasets. Then, we used each of our two classifiers to assign a label to each of the recordings, and we calculated the accuracy of each classifier at this task. The results are shown in Figure 3.

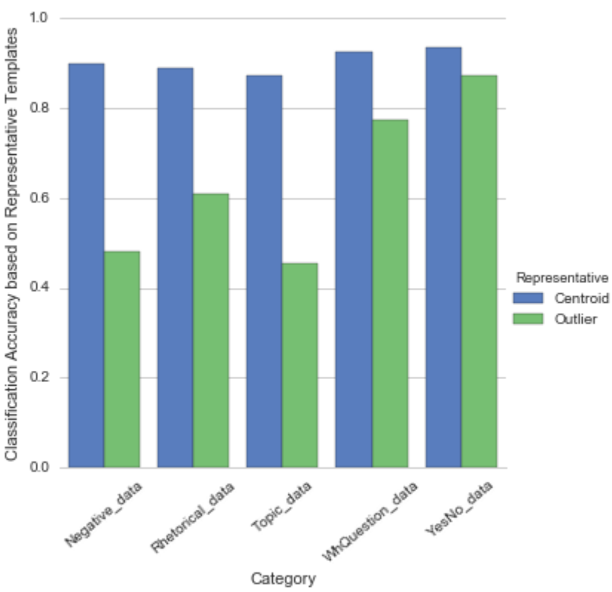


Figure 3: Accuracy results when using centroid versus outliers as exemplars for classification.

While we are not primarily interested in building an NME classifier in this paper, this form of comparison allows us to see how well the "centroids" identified using our above procedure are serving as representatives of each category of NME. We found that when the centroids are used as exemplars, a higher accuracy is achieved for all the NME types.

5. Inter-Signer Evaluation

While the preliminary assessment above suggested that the centroids were serving as effective exemplars of our ASL NME categories, the goal of this study is to determine whether our new centroid selection method, based on multidimensional DTW and MPEG4 facial features, would be useful for selecting an exemplar of human performance that could serve as the basis for animating a virtual human character. We note two key challenges in such a usage scenario:

- A virtual human signer may have different facial proportions than the human in the datasets.
- The specific sentence we wish to synthesize as animation may not have been performed by the human in the original ASL NME datasets.

Thus, we conducted a more rigorous form of evaluation to determine whether these centroids would be effective exemplars when considering data from a different signer performing a different sentence. Specifically, we compared the centroid and the outlier for each ASL NME category to two "gold standard" recordings of an ASL performance from a male native ASL signer performing the same category (topic, wh-question, etc.) of ASL NME.

Notably, the (male) human in the gold standard recordings is different from the (female) human in the recordings in the dataset used as the basis for centroid and outlier selection. Furthermore, the specific sentences used in the gold standard recordings did not appear in the original data set. Thus, this inter-signer evaluation is a more rigorous method for evaluating whether the centroid recordings identified in our original ASL NME datasets would be effective for animation synthesis.

NME Dataset	Example 1		Example 2	
	v. frames	glosses	v. frames	glosses
Topic	55	2	31	1
Negation	63	4	33	2
Y/N-question	73	5	45	2
Wh-question	35	1	55	3
Rhetorical	25	1	102	4

Table 3: Number of video frames and glosses for the two examples per NME that serve as gold standards.

The source of these gold standard recordings is the collection of ASL videos that we previously released to the research community as a standard evaluation dataset in (Huenerfauth and Kacorri, 2014). In that paper, we defined codenames for the individual recordings in that dataset; using that nomenclature, the recordings selected as gold standards were: W1 and W2 (wh-question), Y3 and Y4 (yes/no-question), N2 and N5 (negation), T3 and T4 (topic), and R3 and R9 (rhetorical question).

Figure 4 presents the difference between our gold standards and (a) the centroid selected using the method in section 3 or (b) the outlier selected using the method in section 3. Since we identified two gold standard recordings, the height of the bars indicates the average of the DTW distance from both of the gold standard recordings. As shown in the figure, the centroid outperforms the outlier.

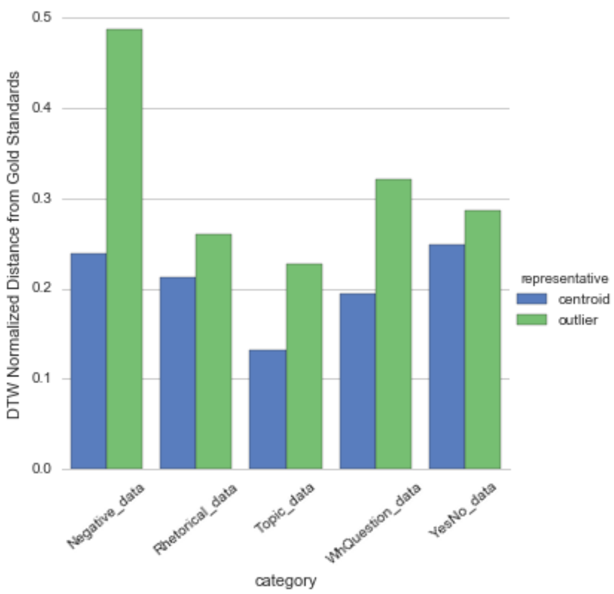


Figure 4: Average DTW distance of the centroid and outlier from two gold-standard recordings of a different signer performing ASL sentences not found in the original datasets, with a smaller distance indicating higher similarity.

6. Conclusions and Future Work

The results of this study indicate that centroid-based exemplar selection using multidimensional DTW operating in the space of MPEG4 FAP recordings is a viable method for selecting exemplar recordings of ASL NMEs. We have demonstrated a methodology for selecting human recordings that can serve as a basis for later stages of sign language animation synthesis. In so doing, this study has investigated a new methodology for utilizing sign language corpus data to advance research in sign language avatar technology, specifically for the selection of NME movements.

While this paper presented our results from analyzing nine MPEG4 facial action parameters (representing eyebrow height and head orientation), we plan to further investigate the utility of modeling additional facial parameters for components of ASL NMEs , such as brow furrowing or eyelid aperture. In addition, this paper presented a preliminary evaluation of the effectiveness of this exemplar selection algorithm. In future work, we intend to conduct a user study to evaluate the quality of animations of sign language generated using this technique, with ASL signers evaluating the animations.

Finally, we note an application of this research for researchers who are seeking to mine a sign language corpus. With the growth of sign language corpora, one challenge faced by linguistic researchers is visualizing patterns in this data to support hypothesis development. We note that the visualizations of "distance" shown in Figure 2 may have potential for assisting linguistic researchers in exploring the NMEs within recordings in a sign language corpus. For instance, graphs or plots of multidimensional DTW similarity of MPEG4 facial features may suggest neighbors, outliers, or clusters of similar recordings in a corpus. In fact, as investigated in (Schmidt et al., 2003), automated statistical clustering techniques could be used to identify variants within a set of recordings, or these graph-like visualizations could support discovery by linguistic researchers. Such graphs might suggest the existence sub-variants in a set of recordings of some ASL NME, which could be investigated in further linguistic work.

Furthermore, the type of visualizations in Figure 2 may be useful by researchers who are collecting sign language corpora so that they may quickly visualize the diversity of their collection during the recording and annotation process; this may indicate how diverse the corpus is. Therefore, such visualizations may help researchers to determine, during the collection of a corpus, whether further data are needed.

7. Acknowledgements

This material is based upon work supported by the National Science Foundation under award number 1059218, 1065009, 1065013, and 1506786. This material is also based upon work supported by the Science Fellowship and Dissertation Fellowship programs of The Graduate Center, CUNY. We are grateful for support from Dimitris Metaxas at Rutgers University.

8. Bibliographical References

Gillian, N., Knapp, R.B., O'Modhrain, S. (2011). Recognition of multivariate temporal musical gestures using n-dimensional dynamic time warping. In *Proc of the 11th Int'l conference on New Interfaces for Musical Expression.*

Giorgino, T. (2009). Computing and visualizing dynamic time warping alignments in R: the DTW package. *Journal of Statistical Software* 31(7), pp. 1–24.

Hagberg, A.A., Schult, D.A., and Swart, P.J. (2008). Exploring network structure, dynamics, and function using NetworkX. In *Proc of the 7th Python in Science Conference*, Pasadena, CA, pp. 11–15.

Huenerfauth, M. (2004). Spatial and planning models of ASL classifier predicates for machine translation. In *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation: TMI.*

Huenerfauth, M. (2008). Generating American Sign Language animation: overcoming misconceptions and technical challenges. *Univ. Access. Inf. Soc.* 6(4), pp. 419–434.

Huenerfauth, M, Kacorri, H. (2014). Release of experimental stimuli and questions for evaluating facial expressions in animations of American Sign Language. In *Proceedings of the 6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel, LREC 2014,* Reykjavik, Iceland.

Kacorri, H. (2015). TR-2015001: A survey and critique of facial expression synthesis in sign language animation. *Computer Science Technical Reports*. Paper 403. The Graduate Center, CUNY, New York, NY, USA.

Kacorri, H., Huenerfauth, M. (2015). Evaluating a dynamic time warping based scoring algorithm for facial expressions in ASL animations. In *Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT), INTERSPEECH 2015*, Dresden, Germany.

Mana, N., Pianesi, F. (2006). HMM-based synthesis of emotional facial expressions during speech in synthetic talking heads. In *Proc. of the 8th Int'l Conf on Multimodal Interfaces*, pp. 380–387.

Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., Lee R.G. (2000). *The Syntax of American Sign Language: Functional categories and hierarchical structure*, Cambridge, MA: The MIT Press.

Neidle, C., Liu, J., Liu, B., Peng, X., Vogler, C., Metaxas, D. (2014). Computer-based tracking, analysis, and visualization of linguistically significant nonmanual events in American Sign Language (ASL). In *Proc. 6th workshop on representation and processing of sign languages, LREC 2014.*

Ouhyoung, M., Lin, H.S., Wu, Y.T., Cheng, Y.S., Seifert, D. (2012). Unconventional approaches for facial animation and tracking. In *SIGGRAPH Asia*, pp. 24.

Pandzic, I. S., and Forchheimer, R. (2003). *MPEG-4 facial animation: the standard, implementation and applications*. Wiley.

Sakoe, H., Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing* 26(1), pp. 43–49.

Schmidt, C., Koller, O., Ney, H., Hoyoux, T., Piater, J. (2013). Enhancing gloss-based corpora with facial features using active appearance models. In *3rd Int'l Symp on Sign Language Translation and Avatar Technology (SLTAT)*. Chicago, IL, USA.

Visage Technologies. (2016). FaceTrack. Retrieved January 29, 2016, from http://www.visagetechnologies.com/products-and-services/visagesdk/facetrack/.

The Usability of the Annotation

Jarkko Keränen, Henna Syrjälä, Juhana Salonen, Ritva Takkinen
Department of Languages (Sign Language Centre), University of Jyväskylä, Finland
P.O. Box 35, FI-40014 University of Jyväskylä, Finland
E-mail: {jarkko.j.keranen, henna.t.syrjala}@student.jyu.fi & {juhana.salonen, ritva.a.takkinen}@jyu.fi

Abstract

Several corpus projects for sign languages have tried to establish conventions and standards for the annotation of signed data. When discussing corpora, it is necessary to develop a way of considering and evaluating holistically the features and problems of annotation. This paper aims to develop a conceptual framework for the evaluation of the usability of annotations. The purpose of the framework is not to give conventions for annotating but to offer tools for the evaluation of the usability of the annotation, in order to make annotations more usable and make it possible to justify and explain decisions about annotation conventions. Based on our experience of annotation in the corpus project of Finland's Sign Languages (CFINSL), we have developed six principles for the evaluation of annotation. In this article, using these six principles, we evaluate the usability of the annotations in CFINSL and other corpus projects. The principles have offered benefits in CFINSL: we are able to evaluate our annotations more systematically and holistically than ever before. Our work can be seen as an effort to bring a framework of usability to corpus work.

Keywords: usability, annotation, signed language, evaluation, corpus, framework

1. Introduction

Annotation conventions have been developed in various corpus projects for different signed languages (e.g. Johnston, 2016 Australia; Crasborn et al., 2015 Netherlands; Wallin & Mesch, 2014 Sweden). The corpus project of Finland's Sign Languages (CFINSL) started in 2014 (see also Salonen et al., 2016, in this volume). New corpus projects start by getting to know what annotation conventions are already in use in corpus projects which are further on in the process. For this reason, the documentation of these processes and interaction between researchers working on corpora is important. Recently, the Dutch and British corpus projects have proposed the standardisation of annotation conventions in order to enable cross-linguistic research (Crasborn & Bank & Cormier, 2015). In the *Digging Into Signs* workshop in London [1] several projects compared their annotation conventions.

Different corpus projects have discussed the problems of annotation and solved them in their own, different ways. In our project we have noticed that we need to be more efficient in our discussion of the problems and their solution. Three problems have come up when reading various annotation guidelines and workshop presentations. First, the discussion is sometimes fragmented, e.g. some discuss problems to do with memory (e.g. Johnston, 2016: 18**;** Santoro & Geraci, 2015: 11), and some talk about the intelligibility of the annotation (Wallin & Mesch, 2015: 10). In this paper, annotation is considered holistically, which means in this context that we look at annotation both as a whole and in terms of the interaction between its different aspects. Annotation has users, and users have their own particular goals; annotation is also used in conjunction with other tools such as ELAN [2] and Signbank [3]. The second problem is that different researchers use different terms for the same subject matter. In the Italian Sign Language (LIS) corpus project (Santoro & Geraci, 2015: 21–24), annotations are evaluated on the basis of whether they are simple or complex. These, however, are ambiguous concepts. Is it due to complexity that the user has difficulty remembering the conventions or that the convention is hard to find when someone tries to find it? Thirdly, it may be said that annotation is continuously evaluated but not clearly stated on what basis this evaluation is carried out. Is annotation evaluated intuitively or using a certain shared and established method?

From what has been said above, we see that what is needed is a conceptual framework which could be used to develop a more holistic and systematic way of considering and evaluating annotation conventions. With a framework we could develop our methods and talk about annotation using the same terms. In our experience the concept of usability is very useful because it makes reference to the basic problems of all corpus projects: how we can develop annotation that meets the users' needs and makes searching effective.

In this article, when we talk about annotation the focus is mostly on glosses, including the codes inside the annotation cell and not, for example, on the length of a sign or the annotation of non-manual features.

This article presents in Section 2 a usability framework, what it means, and how it could be applied to annotation. Section 3 introduces six principles which are tools for the

[1] http://www.bslcorpusproject.org/events/digging-workshop/

[2] ELAN. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. URL: *http://tla.mpi.nl/tools/tla-tools/elan*.

[3] "The FinSL Signbank has been developed on the basis of the NGT Signbank , which is a branch of the Auslan Signbank . The source codes for these three versions of Signbank are all available on Github (https://github.com/Signbank)."

consideration and evaluation of annotation practices. Section 4 presents an evaluation and comparison of different ways of annotating according to the six principles. In Section 5, the profile of the annotator is briefly discussed. The last section summarises what is dealt with in the earlier sections of this article.

2. The Relevance of the Usability to Annotation

Usability is a general term that can be applied to all kinds of products (see Figure 1): a product is usable if its users find it useful, easy and pleasing. Usability has three components: 1) the users of the product and their knowledge and skills, 2) the product itself as well as the functions and features it offers, and 3) the context in which the product is used and in which the users' functions and objectives when using the product are manifest. (Mäntylä, 2001: 128; ISO 9241–11, 1998.) Just as somebody designs, for example, a usable website, we are trying to design a usable annotation protocol.

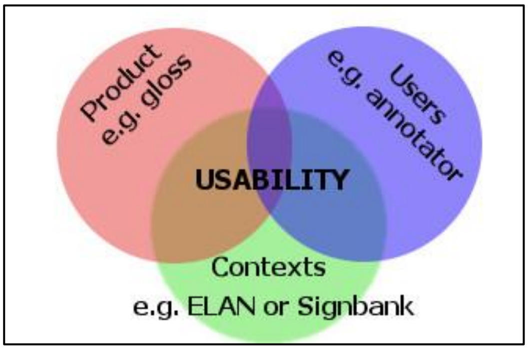


Figure 1: The aspects of usability

It makes sense to talk about the usability of annotation because an annotation has functions and features. As a product, for example, the letters of a gloss can be visually perceived; annotations connect a written word or other marks to a sign shown temporally in a video; annotations offer linguistic or phonetic information.

Thinking of the users of annotation, how well users can create or use annotations depends on the annotators' knowledge and skills. An expert annotator knows better about the possible uses of annotations and ELAN or Signbank than an inexperienced annotator. This raises two questions: one, of whether a native signer can annotate without any knowledge of linguistics (see Rutkowski & Filipczak & Kuder, 2015: 45), and two, of how well a sign language teacher without technological expertise knows how to use corpus material.

Annotation is always used in certain contexts. For instance, ELAN or Signbank– possible contexts of use – affect the annotation and the possibilities of how or what kinds of things are annotated.

Usability includes three elements: product, user, and context. The separate elements can be considered separately, but because the separate parts affect each other, the relations between them also need to be looked at. Thus, when the aim is to consider annotation holistically, one needs to examine it in relation to the user and the context.

Usability is evaluated or measured in terms of how well users can learn to use the product, and how effectively they can use it in order to achieve their goals. The usability of any tool, machine, programme or device which a person uses can be evaluated. (see Nielsen, 1993: 24–26.) There are ready-made methods of evaluation but in a corpus project an economical way of beginning is to start with a checklist (see Whitenton, 2015; NASA/FAA, 2000: 70).

The discussion above shows that usability as a framework has potential because it is useful from at least three points of view. One of the benefits is that the usability of the annotation can be considered and evaluated holistically. The second benefit is that it offers a common language between corpus researchers: it enables discussions about certain topics using the same concepts and ideas about usability, like memory, readability and efficiency. The third benefit of the framework is methodological: we can take advantage of ready-made methods and apply them to an evaluation of the usability of the annotation. In the next section, we will present six principles for the evaluation of the usability of annotation.

3. Six Principles

The principles for the usability of annotations that we use and present here have gradually developed as a result of our experience and our reflections, as well as from a review of discussions held by other corpus projects. The principles are based on the attributes of utility, which are that it is easy for users to achieve their goals and that at the same time they should find the product pleasant (See also Nielsen, 1995; 2012). Is it pleasing to use corpus material if e.g. reading it is difficult or the one does not find the information one is searching for?

The Six Principles for Usability of Annotations
Equivalence of meaning
Findability
Readability
Intelligibility
Consistency
Computer-Readability

Table 1: The list of six principles

The principles for the usability of annotations are shown in Table 1. The principles can be seen as a checklist (see Whitenton, 2015; NASA/FAA 2000: 70). Our aim in presenting these principles is to establish the necessary qualities of an annotation and make it possible to consider and evaluate the usability of an annotation one principle at a time. In that way, we can achieve a more holistic and systematic picture of the usability. We put forward the following principles concerning the usability of annotation.

The principle of **equivalence of meaning** means that the gloss and the meaning must be consistent, e.g. the gloss PALLO always refers to the meaning 'pallo' (ball). This principle concerns glosses written in capital letters.

Findability means that users can quickly and logically find the annotations for which they are searching. Weak findability makes the finding of the desired information slow or simply impossible. If annotations are designed and classified well, the search shows the desired information and filters out other, unwanted information (see an example in section 4).

Readability means that reading the annotation is technically easy and the annotation does not look messy. A readable annotation is clear, salient and distinguishable from another annotation. Readability helps one to read the search results because it means the results can be skimmed quickly and economically.

Intelligibility means that the meaning of the gloss is quickly identified. The user does not need to struggle to remember the system of annotation but can easily identify meaning (cf. Nielsen 1995).

Consistency means that the glosses are logical and comparable to each other. Johnston (2008) states that one can talk about a corpus only when the glosses and annotations are consistent and logical, and they consistently identify the signs. Different annotations and mistakes are a barrier to the functionality of a corpus. If the annotation practices are consistent, the glosses and the codes are always annotated in the same way and in the same order. For example, SSL corpus annotation (Wallin & Mesch, 2014) uses the symbol @ after gloss and after that the codes for linguistic categories.

Computer-readability refers to the need for the codes linked to the gloss to function technically without any problem with programmes such as ELAN and Signbank. Every programme has its limitations and advantages. Computer-readability is concerned with which codes are good for annotation from the point of view of the characteristics of the programmes with which they will be working.

The principles function as a tool for evaluating the annotation. Concerning usability, the basic idea is that *the more the principles are followed, the more usable the annotation will be*. Each annotation can be evaluated in terms of each principle, to see how and why any problems have arisen. This makes it possible to refine and improve the annotation.

4. Evaluating the Usability of the Annotation

In this section we introduce some examples of annotations. We use the above mentioned principles when we evaluate our own annotations and the annotations used in other sign languages. This section discusses different problems which we have met in our annotation work and in other corpus projects we have encountered.

In different corpus projects, sign variants are marked in different ways, for example with a letter (SIGN-A) or number code (SIGN1) (see Crasborn et al., 2015: 5; Cormier et al., 2015: 6). In the corpus of SSL (Wallin & Mesch, 2015: 7) as well as in CFINSL, handshapes are used to mark sign variants. In CFINSL we have also used other phonological parameters: place of articulation and movement. There are three reasons for this.

Firstly, in CFINSL we annotate both phonetic and lexical variations because CFINSL also has a pedagogical aim (cf. Salonen et al., 2016, in this volume). Our aim is not to annotate phonological features per se, but we annotate the structural differences between signs which have the same meaning, using phonological parameters. Examples include a handshape e.g. 'run' JUOSTA(BB) and JUOSTA(SS); a movement, 'name of a city' TAMPERE(bouncing) and TAMPERE(sliding); and place of articulation, 'black' MUSTA(forehead) ja MUSTA(eye).

Secondly, with phonological codes the relationship between the gloss and the meaning can be better clarified. For example, the Finnish word 'puoli' means two different things, in English 'half' and 'side'. FinSL has separate signs for those two meanings. Therefore we differentiate between these two meanings with handshape codes PUOLI(GG) 'half' ja PUOLI(B) 'side' .

The third reason is the principle of intelligibility. When the sign variants are coded with letters or numbers (see Crasborn et al., 2015: 5; Cormier et al., 2015: 6) it causes a memory load, because the signs cannot be identified according to the numbers or letters of the alphabet. The annotator may be forced to look at the forms of the signs again and again e.g. in Signbank or wherever the descriptions of the signs are stored. The SSL project has also adopted handshapes to indicate variants because of intelligibility (Wallin & Mesch, 2015: 10).

The choice of a certain phonological code is based on the salience of the parameters in a sign. Handshape is often the most salient. If the handshape of two variants is similar, the code of another salient parameter e.g. movement or place of articulation is taken into consideration. To take an example, the sign KOSKAAN-EI 'never' can be produced in three different ways, and they all have the handshape (B). Thus, it is impossible to distinguish between the sign variants with a handshape code. Instead, the more salient parameter in two of the variants is movement. However, in the third variant of the sign the most salient parameter is place of articulation because movement is minimal. Therefore we decided to annotate the variants as KOSKAAN-EI(sliding), KOSKAAN-EI(circle) and KOSKAAN-EI(mouth).

In our annotation we still most commonly use handshape, because it works well in distinguishing between the sign variants in most signs. We do not add any phonological code unless they are phonetic or lexical variants.

Next let us look at annotation from the viewpoint of readability. The annotation guidelines of the SSL corpus project show that before 2011 the annotation was, for example, (g-)LÅTA-VARA and after 2012 the order was LÅTA-VARA@g (see Wallin & Mesch, 2012; Wallin & Mesch, 2011;). The annotation guidelines did not give any reason for the change.

In CFINSL the order of elements in the gloss for such a sign is GLOSS@g because this order is more readable, especially in the list of search results. (We are still considering our use of the symbol @ because the annotation conventions of CFINSL are currently being reviewed). Another question is whether the symbol- @ or a colon (:) is more salient in the ELAN search list, e.g. ISÄ(S)@mon or ISÄ(S):mon 'father' . In SSL annotation @ has been used since 2011 because the symbol @ is more readable (Mesch, personal communication 17.12 2015).

To consider annotation from the viewpoint of findability, let us look at the sign for 'know'. We created the basic gloss TIETÄÄ 'know', plus an additional code in order to clarify the form and meaning of the sign. There are several glosses: TIETÄÄ(toisto) 'know', TIETÄÄ-PALJON 'know a lot' and TIETÄÄ-EI(55) 'not know'. This works well because in the search list the annotator can find several glosses that start with TIETÄÄ, and after the main gloss possible additional specifications. Earlier, misled by Finnish as a metalanguage, we used different glosses: TIETÄÄ 'know', TIETÄVÄINEN 'know a lot' and EI-OLE-TIETOINEN 'not know'. Those different glosses were difficult to remember and find because they were scattered in different places in the list of glosses.

Similarly, a few corpus projects (e.g. Johnston, 2016; Crasborn & Bank & Cormier, 2015) have decided to put the main gloss first and then any additional information e.g. KNOW-NOT because in the search list the annotator can find several glosses that start with KNOW.

Another example of a findability problem is the annotation of fingerspelling in the corpus project of Italian sign language (Santoro & Geraci, 2015: 23). The alphabet is annotated with hyphens, as in W-O-R-D. The problem is that the search cannot find all the fingerspelled glosses at the same time, because there is no categorising code for fingerspelling. In addition, it is difficult for users of the corpus material to search the finger alphabet if they must use the alphabet and hyphens. One solution is to add a code for fingerspelling to the gloss, as is done in the BSL corpus (FS:) (Cormier et al., 2015: 13) and in the NGT corpus (#:) (Crasborn et al., 2015: 17). Using the fingerspelling code it is possible to find all the fingerspelled glosses.

Another findability problem was found in the annotation of composed numerals in the NGT corpus project. They are annotated as numbers e.g. 128. The first problem is the lack of a code for classifying numbers. Another problem is that it is difficult to find large numbers like tens, hundreds or thousands etc. With the gloss for 128, for example, it is impossible to distinguish the 100 because the smaller number, 20, cuts across the hundred, and it is impossible to find the 20 because the smaller number, 8, cuts across the twenty. To get round this problem, the BSL corpus project (Cormier et al., 2015: 6) and the LIS corpus project (Santoro & Geraci, 2015: 24) annotate numbers in words in one annotation cell; in the case of BSL e.g. [NINETEEN^EIGHT^NINE]; and in LIS e.g. [MILLENOVECENTOOTTANTANOVE]. In this way it is possible to search for composed numbers.

Next we will look at two examples from the viewpoint of consistency. In the SSL corpus annotation we find the glosses SOVA and SOVA(Jv) 'nukkua' (SSLCO2_331; SSL02_409). One of the synonym signs is without any handshape code and the other gloss has one. In the CFINSL project we are adding a phonological code consistently to every variant, e.g. 'sleep': NUKKUA(LL), NUKKUA(B^B^) and NUKKUA(BB). Consistency is particularly important to new annotators, as it makes it easier for them to learn to annotate. Another reason for consistency is that it presents all the sign variants of a sign as equal.

It is also worth considering computer-readability, because annotating and searching for annotations is done by computer programmes. How well the annotation works also depends on the characteristics, restrictions and flexibility of the computer programme. We have met problems in the search system of ELAN. Our aim was to search using the handshape code (B). When (B) was written in the search field, all the glosses with (B) in parenthesis appeared in the results, e.g. (B), (BB), (GB) and (SB). We could not restrict the search so that it gave only results with just (B). This shows that it is important to know the characteristics, restrictions and possibilities of programmes like ELAN and Signbank from the angle of how computer-readable the annotation is. In future, it would be worth considering computer-readability from the user-friendly perspective.

In this section we have considered the usability of each of the principles except consistency of meaning. The next section deals briefly with the profile of an annotator, especially from the cognitive point of view.

5. Attention to the Annotator

When trying to create the ideal annotation, it is important to remember the question of the relationship between human beings and technology (Saariluoma et al., 2010). Human cognition causes difficulties which need to be identified. One of the difficulties is categorisation. Let us look briefly at how categorizing can be a problem, with reference to the prototype theory.

According to the prototype theory (Rosch, 1975), the prototype of a category represents culturally the most typical example of the category while the periphery represents a less typical example of the category. Cognitive categories are vague; i.e. how we categorise things fluctuates. The effect of a prototype is present also in language on all its levels: phonology, morphology, and syntax (Lakoff, 1987).

The prototype effect is present in linguistic annotation, too. For example, discrimination of a phonological handshape variant can be problematic because the borderlines between handshapes are vague. We have considered whether it is essential to discriminate between e.g. variants of the sign ISÄ(Ax) 'father' and ISÄ(S), because the handshapes closely resemble each other (see Figure 2). However, in FinSL both handshapes appear frequently in the sign. Are they lexicalised signs? In our work, a third variant with the handshape (G) was more easily annotated as a separate gloss because this handshape was clearly different from the handshapes (S) and (Ax) and because annotators can identified the sociolinguistic feature of the sign ISÄ(G): older signers typically sign with handshape (G).

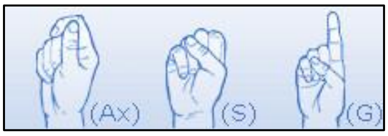


Figure 2: Handshapes (images from SUVI[4])

Additionally, the prototype of the category varies, depending on the context: as Labov (1973) puts it, the prototype is also context-bound. This makes categorisation more complicated and vague, and annotation decisions intuitive. Decisions about annotations can be difficult because different people have different intuitions and linguistic backgrounds, e.g. native signer with deaf parents or hearing parents, in different contexts. This shows that there is no such thing as perfect annotation; all we can do is to strive to achieve annotation that is as usable as possible, creating clear principles for the annotation in each metalanguage.

Even if the annotator has had training in linguistics and a lot of experience as an annotator, it is not possible to achieve perfect annotation. The question of how we categorise the world is also present in annotation work. The point of this section is to remind ourselves that annotators need to be aware that the human cognition affects annotation. With this awareness they will understand better why they sometimes have problems in the categorisation of signs and phonetic forms.

6. Conclusion

We have seen that usability as a conceptual framework allows us to consider annotation more systematically and holistically and therefore to better achieve the goals of the corpus. The framework sets the usability of the annotation in relation to the user and the context. Potentially, it also offers a common language to communicate using the same concepts and ready-made methods for working the usability of the product. Usability as the framework has potential to channel the discussion about problems with annotation into one and the same direction for all those concerned.

We have created principles in the form of a checklist for evaluating annotations. These principles are based on our experience of annotation so far, and will develop further during our annotation work and through feedback from colleagues. Coherent principles with clear explanations help annotators in their work. For efficient annotation we need principles that are easy for annotators to learn and remember. This will reduce the number of mistakes made during annotation. An ideal annotation is easy to learn, easy to read, easy to find, easy to understand and remember. The user experience - useful, easy and pleasing- is one of the most important elements of usability.

Documentation of the decisions behind annotation conventions is important because such information can and should be made available to others, for example to those working on new corpus projects and struggling with the same sorts of problems. Documentation would make it possible to develop common principles for evaluation, which in turn would make cooperation and standardisation easier. Finally, human cognition and its effects on annotation are topics we should discuss together in the future.

7. Acknowledgements

We authors want to thank Anna Puupponen for commenting on the article, and team working on the Finnish corpus projects for useful discussions. The authors also wish to thank Johanna Mesch for her personal communication (17.12 2015).

8. Bibliographical References

Crasborn, O., Bank, R. & Cormier, K. (2015). *Digging into Signs: Towards a gloss annotation standard for sign language corpora*. Final version. Radboud University: Sign Language of Faculty of Arts & Centre for Language Studies. URL: http://www.ru.nl/sign-lang/projects/digging-signs/

Crasborn, O., Bank, R., Zwitserlood, I., Kooij, E., Meijer, A. & Sáfár, A. (2015). *Annotation Conventions for The Corpus NGT*. Version 3. Radboud University Nijmegen: Centre for Language Studies & Department of Linguistics.

Cormier, K., Fenlon, J., Gulamani, S. & Smith, S. (2015). *BSL Corpus Annotation Conventions*. Version 2.. Deafness Cognition and Language (DCAL) Research Centre, University College London.

Digging into Signs Workshop: *Developing Annotation Standards for Sign Language Corpora*. -British Sign Language Corpus Project. URL: http://www.bslcorpusproject.org/events/digging-workshop/

ISO 9241-11 (1998). *Ergonomic requirements for office work with visual display terminals (VDTs)* Part 11: Guidance on Usability. ISO.

Johnston, T. (2008). Corpus linguistics and signed languages: No lemmata, no corpus. In the *Proceedings of the 3rd Workshop on the Representation and Pro-*

[4] Suvi, the on-line dictionary of Finland´s Sign Languages. URL: http://suvi.viittomat.net/

cessing of Sign Languages: Construction and Exploitation of Sign Language Corpora, Paris: ELRA, pp. 82–87.
Johnston, T. (2016). *Auslan Corpus Annotation Guidelines*. Macquarie University (Sydney) and La Trobe University (Melbourne). February 2016 version.
Labov, W. (1973). The boundaries of words and their meanings. In C-J. Bailey & R. Shuy (eds.), *New Ways of Analyzing English*, pp. 340–73. Washington: Georgetown University Press.
Lakoff, G. (1987). *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
Mäntylä, M. (2001). Käytettävyys ja kognitito. In P. Saariluoma, M. Kamppinen. & A. Hautamäki (eds.), *Moderni kognitiotiede*. pp. 128–151. Helsinki: Gaudeamus.
NASA/FAA (2000). *Developing operating documents: A manual of guidelines*. eVersion. URL:http://human-factors.arc.nasa.gov/opdoc-workshopIV/Opdocfinal.pdf
Nielsen, J. (1993). *Usability Engineering*. San Diego & London: AP professional.
Nielsen, J. (1995). *10 Usability Heuristics for User Interface Design*. URL: https://www.nngroup.com/articles/ten-usability-heuristics/.
Nielsen, J. (2012). *Usability 101: Introduction to Usability*. URL: https://www.nngroup.com/articles/usability-101-introduction-to-usability/
Rosch, E. (1975). Cognitive Representations of Semantic Categories. *Journal of Experimental Psychology: General*, Vol.104, No.3, p.192–233.
Rutkowski, P., Filipczak, J. & Kuder, A. (2015). PJM Corpus Annotation Guidelines. Presentation presented at the *Digging into Signs Workshop*, University College London, March 30–31, 2015.
Saariluoma P., Kujala T., Kuuva S., Kymäläinen T., Leikas J., Liikkanen L. & Oulasvirta A. (2010*). Ihminen ja teknologia − Hyvän vuorovaikutuksen suunnittelu*. Teknologiateollisuus: Helsinki.
Salonen, J., Takkinen, R., Puupponen, A., Nieminen, H. & Pippuri, O. (2016). Creating Corpora of Finland´s Sign Languages. To appear in *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining.* Paris: ELRA.
Santoro, M. & Geraci, C. (2015). Italian Sign Language (LIS) Corpus. Presentation presented at the *Digging into Signs workshop*, University College London, March 30–31, 2015.
Wallin, L., Mesch, J. & Nilsson, A-L. (2011). *Transkriptionskonventioner för teckenspråkstexter*. Version 3. Avdelningen för teckenspråk, Stockholms universitet.
Wallin, L., Mesch, J. & Nilsson, A-L. (2012). *Transkriptionskonventioner för teckenspråkstexter*. Version 4. Avdelningen för teckenspråk, Stockholms universitet.
Wallin, L. & Mesch, J. (2014). *Annoteringskonventioner för teckenspråkstexter*. Version 5. Avdelningen för teckenspråk, Stockholms universitet.
Wallin, L. & Mesch, J. (2015). Swedish Sign Language Corpus. *Digging into Signs Workshop: Developing Annotation Standards for Sign Language Corpora*. -British Sign Language Corpus Project. URL: http://www.bslcorpusproject.org/wp-content/uploads/MeschWallin_DiggingIntoSigns_London2015.pdf
Whitenton, K. (2015). *Menu Design: Checklist of 15 UX Guidelines to Help Users*. URL: https://www.nngroup.com/articles/menu-design/

9. Language Resource References

ELAN. Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands. URL: *http://tla.mpi.nl/tools/tla-tools/elan*.
Signbank: "The FinSL Signbank has been developed on the basis of the NGT Signbank, which is a branch of the Auslan Signbank. The source codes for these three versions of Signbank are all available on Github (https://github.com/Signbank)."
SSLCO2_331 & SSL02_409. In Mesch, J., Rohdell, M. & Wallin, L. (2015). Annotoringsfiler i ett komprimerat paket: Mesch_Rohdell_Wallin_201509.zip. Paketetets innehåll har utökats till 71 eaf-filer med senaste annoteringskonventioner och översättningar.
SUVI, the online dictionary of Finland´s Sign Languages. The Finnish Association of the Deaf. URL: http://suvi.viittomat.net/.

Transitivity in RSL: a Corpus-Based Account

Vadim Kimmelman
ACLC, University of Amsterdam
Amsterdam, the Netherlands
v.kimmelman@uva.nl

Abstract

A recent typological study of transitivity by Haspelmath (2015) demonstrated that verbs can be ranked according to transitivity prominence, that is, according to how likely they are to be transitive cross-linguistically. This ranking can be argued to be cognitively rooted (based on the properties of the events and their participants) or frequency-related (based on the frequency of different types of events in the real world). Both types of explanation imply that the transitivity ranking should apply across modalities. To test it, we analysed transitivity of frequent verbs in the corpus of Russian Sign Language by calculating the proportion of overt direct and indirect objects and clausal complements. We found that transitivity as expressed by the proportion of overt direct objects is highly positively correlated with the transitive prominence determined cross-linguistically. We thus confirmed the modality-independent nature of transitivity ranking.

Keywords: argument structure, transitivity, Russian Sign Language

1. Introduction

1.1. Theoretical Background

Transitivity is the possibility of a verb to occur with a direct object.[1] If we look at transitivity cross-linguistically, we can observe that different verbs are transitive to a different degree. In a recent study, Haspelmath (2015) analysed a sample of 36 languages in order to assess transitivity cross-linguistically. For each language, the possibility of using a transitive frame has been collected for a list of 70 verbs. Two main results came out of this study: (1) languages differ with respect to the proportion of transitive verbs they have and (2) verbs differ with respect to the proportion of languages in which they are transitive. This latter result generates a ranking of verbs by transitivity prominence, such that some verbs like *die* or *jump* are never transitive, other verbs like *break* and *tear* are transitive in all languages in the sample, and some verbs like *leave* or *know* are transitive in some languages but not others.

Several explanations have been offered for the fact that some verbs are more likely to be transitive than others. For instance, Malchukov (2005) proposed a semantic map for transitivity as in Figure 1. The verbal meanings to the left are more likely to be transitive than the verbs to the right. Verbs of effective action (such as *break*) are the most likely to be transitive because they involve an Agent (an active and volitional participant), who has an effect on the Patient (and the Patient changes as a result of the action). Verbs of contact also involve an Agent and a Patient, but the Patient is not changed by the actions of the Agent, and with verbs of pursuit there is also no contact between the Agent and the Patient. Verbs of perception, emotion, and sensation constitute a separate sub-part of the map because there the notions of Agent and Patient are less directly applicable, and one might speak of an Experiencer and a Stimulus, so they are less likely to be transitive than verbs of effective action, but not ordered with respect to verbs of contact or pursuit. Such a cognitive explanation of transitivity is likely to apply across modalities, so we expect to find similar patterns of transitivity in sign languages, too.

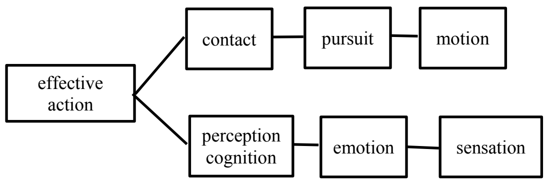


Figure 1: Malchukov's semantic map for transitivity.

Another explanation for the differences in transitivity between different verbs is frequency-based (Haspelmath, 1993; Haspelmath et al., 2014). It is intuitively true (albeit probably not explicitly tested) that some actions in real world frequently involve two participants (such as breaking), while others frequently involve one participant (such as freezing), so the former are more likely to be lexicalised as transitive and the latter as intransitive. Again, such an explanation for transitivity is likely to apply across modalities, so also to sign languages.

The two explanations are not incompatible, as for instance verbs of effective action describe the situations which necessarily frequently involve two participants. However, the frequency-based explanation makes a clear prediction with respect to transitivity in Russian Sign Language (RSL).

Although transitivity ranking is defined cross-linguistically, we think that there is a way of applying transitivity ranking to one language and testing it based on corpus data. We predict for an individual language (signed or spoken) that the frequency of overt object expression should reflect the transitivity prominence from Haspelmath (2015): If a particular event is more likely to have two participants in real

[1] By this definition we consider ditransitive verbs to be a subclass of transitive verbs: they have a direct object, and also an indirect object

world, the verb used to describe this event is more likely to be lexically transitive cross-linguistically AND to have an overt object in corpus data of a particular language.

At first sight it might seem that overt objects in corpus data do not directly reflect transitivity of the verbs, because even transitive verbs can occur without overt objects in languages which allow object pro-drop (and RSL is such a language). However, the possibility of pro-drop is orthogonal to the notion of transitivity. Imagine a language that has transitive verbs X and Y which are transitive to a different degree, so that X is sometimes used intransitively[2], and Y is always (semantically) transitive[3]; object pro-drop in this language is allowed. We then expect that pro-drop will cause X and Y to appear without objects in a certain percentage of cases (a%), but we expect the effect of pro-drop to be the same for the two verbs, because pro-drop depends on the definiteness of the object, and not on its presence in the lexical argument structure of the verb. However, the proportion of cases where the verb is indeed used intransitively and no object is implied depends on the verbal meaning of X and Y and reflects real world frequencies of the events with one or two participants, so it would be different for X and Y (say x% and y% of all cases). The observed percentages of overt objects for X and Y will be (100-(a+x))% and (100-(a+y))% respectively, and so despite the presence of cases of pro-drop, these percentages will also reflect transitivity of these verbs.

1.2. The Current Study

In order to test the relation between the transitivity prominence and overt object expression, we analysed corpus data of RSL. RSL is a language used by at least 120 000 people in Russian Federation. Although some research on linguistic properties of RSL has recently appeared, no studies of argument structure are available yet. Importantly, corpus-based investigations of transitivity have not yet been conducted for any other sign languages.

Apart from providing a basic corpus-based description of transitivity in RSL, in this paper we are interested in answering two more general questions: (1) Does transitivity prominence apply across modalities? and (2) Is transitivity prominence related to overt object expression?

2. Methodology

The RSL corpus (Burkova, 2015) contains naturalistic narratives and a small number of dialogues produced by 37 signers from different regions of Russia. The total length of the video files is 5h 28min. The whole corpus is annotated in ELAN with sign-by-sign translation (separately for the right and left hands) and with sentence-level free translation in Russian; the total number of annotations is over 65 000. Based on the number of annotations on the right hand tier, the estimated number of signs in the corpus is ≈25 000.

To study transitivity, verbal tokens expressing the 80 typical verbal meanings from the VALPAL project (Hartmann et al., 2013) have been identified.[4] Since the RSL corpus does not make use of lexical IDs, the same sign is sometimes translated with different glosses, and the same gloss is sometimes used for different signs; therefore, the search for tokens involved identifying all possible patterns with which a particular verbal meaning would be expressed in Russian, and using regular expressions to search for them. For instance, the meaning 'to dress' in Russian can be expressed by two verbs with different prefixes: *o-devat'* and *na-devat'*, and also the gloss might contain the corresponding noun *odezhda*, so a regular expression was used to identify all possible tokens.

Sometimes a meaning from the VALPAL list was matched to more than one sign in RSL, so labels like SPEAK1, SPEAK2, SPEAK3 have been created to reflect that different signs have the meaning 'to speak'. At first, all different forms were assigned different labels. However, sometimes the differences between the forms appear to be phonetic. For instance, RUN1 contains an alternating movement of the hands, RUN2 contains an identical but synchronized movement of the hands, and RUN3 contains the same movement as RUN2 but only involving fingers. In such cases we also looked at the argument structure of the different variants (as described below) in order to decide whether they should be analysed as one verb. In most cases it turned out that different variants also differed in argument structure, as was the case with SPEAK1, SPEAK2, and SPEAK3, but for the verbs RUN and COME the variants were not different, thus for further analysis we grouped them together.

We excluded the meanings that are expressed in RSL by classifier predicates, as classifier predicates have modality-specific properties, crucially also with respect to argument structure (Benedicto and Brentari, 2004; Kimmelman et al., 2016), and have to be analysed separately. However, we included two signs which can potentially be classifier predicates, namely GIVE and TAKE. In our data there are examples in which these signs are used with abstract objects (1) or with objects that do not correspond to the handshapes used in the predicates, so it is clear that these verbs at least have non-classifier usages. Note, however, that we then included all usages of these verbs, so some tokens of these verbs might in fact be classifier predicates.

(1) MORE INFORMATION GIVE
'[...] to give more information.'

Altogether, based on the 80 verbal meanings, we annotated 117 verbal signs (types) totalling 2248 tokens. For each token we created annotations on several tiers, including a tier for the verbal meaning and a tier for word order in the clause. Clause boundaries were identified on semantic grounds: the verb and all its arguments and adjuncts (similar to Hansen and Hessmann (2008)). We labelled the most agentive argument as S. For verbs which only take one object, we labelled the relevant argument as O (direct object). For ditransitives, such as SAY, SPEAK, TELL, TALK, TEACH we labelled the Addressee/Goal as O and the

[2] As for instance transitive verbs *to melt* and *to eat* can be used intransitively in English.

[3] As for instance the transitive verb *to give*, which can only be used transitively in English.

[4] The VALPAL project used a list of 80 verbal meanings; however, Haspelmath's transitivity prominence is only discussed for 70 of the 80 verbs in Haspelmath (2015).

Theme/Patient as O2. This is an unusual decision from a typological perspective, as the Theme/Patient is usually the direct object. However, in RSL all these verbs are agreeing and they agree with the Addressee/Goal argument, so based on the language-internal criterion of agreement we identified this argument as the direct object. We also identified possible clausal complements CO in cases like *He saw [that she left]*.

In order to compare transitivity of RSL verb to Haspelmath's transitivity prominence ranking, we calculated proportions of overt arguments per verbal type (e.g. number of tokens with overt O divided by the total number of tokens). For the sake of completeness, we calculated proportions of overt direct and indirect objects, clausal complements, and subjects. Since we are working with proportions, verbal types with a small amount of tokens could distort the picture, so we only included verbs with at least 25 tokens. This resulted in 29 verbal types totalling 1611 tokens. Note that these types includes some groups of verbs which correspond to one meaning in Haspelmath's ranking, such as SPEAK2 and SPEAK3, and also some verbs from the longer VALPAL list of meanings which are not included in Haspelmath's ranking. Thus for the comparison with this ranking we were left with 25 verbal types, but for the general overview of overt expression of arguments we report on all 29 frequent verbal meanings.

3. Results

Basic properties of argument structure of frequent verbs in RSL can be assessed by looking at the overt direct and indirect objects as well as clausal complements which accompany these verbs. In Figure 2, the frequent RSL verbs are ranked according to the proportion of overt direct objects. One can immediately see that this ranking looks quite reasonable, with verbs like HAPPY and LAUGH never occurring with a direct object, and verbs like TAKE and GIVE being the most frequently accompanied with an overt direct object.

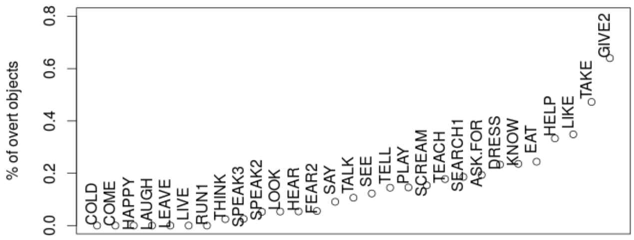


Figure 2: Frequent RSL verbs ordered by proportion of overt direct objects.

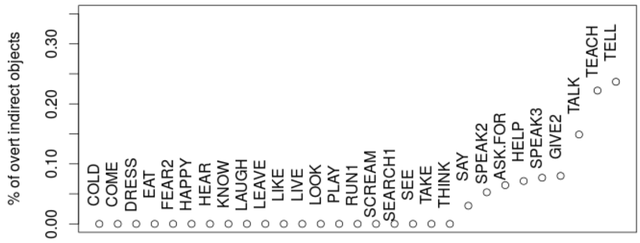


Figure 3: Frequent RSL verbs ordered by proportion of overt indirect objects.

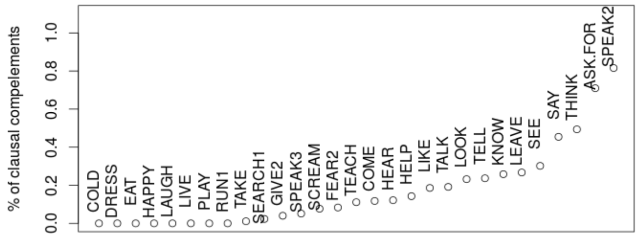


Figure 4: Frequent RSL verbs ordered by proportion of overt clausal complements.

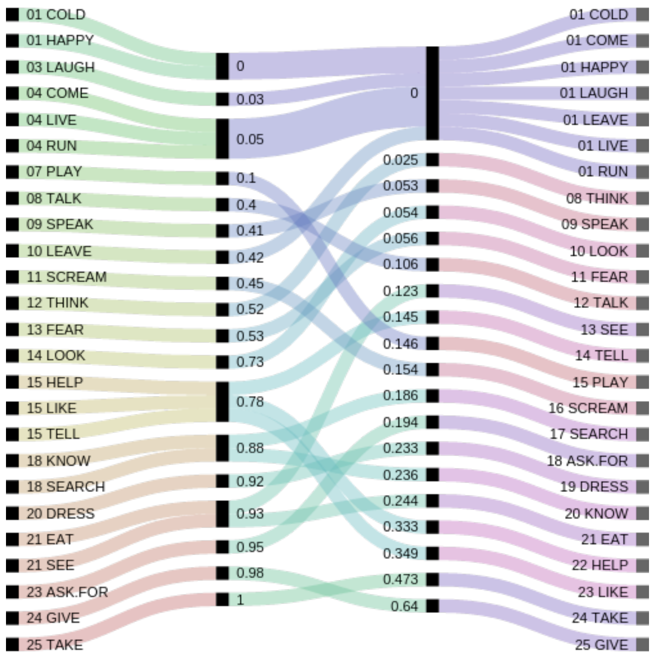


Figure 5: Haspelmath's ranking (left) and ranking based on overt objects in RSL corpus (right). Created with RAW (raw.densitydesign.org).

In Figure 3, the frequent RSL verbs are ranked according to the proportion of overt indirect objects. Again, this ranking intuitively makes sense, as the meanings which are typically ditransitive cross-linguistically, such as TEACH and TELL are high in this ranking.

Finally, in Figure 4, the frequent RSL verbs are ranked according to the proportion of overt clausal complements. Not surprisingly, verbs of speech, perception and cognition are high in this ranking, as semantically the complements of these verbs are often full propositions and not entities.

It should be clear that the best general measure of transitivity is the proportion of overt direct objects, because indirect objects and clausal complements target specific semantic subclasses of verbs. Thus we compared the ranking from Figure 2 to the ranking by transitivity prominence from Haspelmath (2015). For each verb, Haspelmath provided a number which represents the proportion of languages in which a particular verb is lexically transitive. In order to compare our ranking, we calculated the correlation between this measure and the proportion of overt direct objects in our RSL corpus. Since neither data sets are normally distributed, we used Spearman's rank correlation. It turned out that the two measures are highly positively correlated ($\rho = 0.849$) and that the correlation is highly significant ($p = 8.081 * 10^{-8}$).[5]

The relation between the ranking based on the RSL corpus and Haspelmath's ranking is provided in Figure 5 as an alluvial diagram. On the left, the verbs are ranked according to transitivity prominence (Haspelmath, 2015), and connected to the numerical values of transitivity prominence (center left). On the right, the verbs are ranked according to the proportion of overt direct objects in the RSL data, and connected to the numerical values of the proportion (center right). The numerical values of the corresponding verbs are connected to each other. This figure also clearly represents the ties, so it lets one see for instance the verbs which are never transitive cross-linguistically (COLD, HAPPY) and the verbs which never occur with overt direct objects in RSL (COLD, HAPPY, LAUGH, COME, LIVE, RUN).

From the graph and the high correlation coefficient it should be clear that transitivity prominence as defined by Haspelmath based on cross-linguistic data is also reflected by the proportion of overt direct objects in the corpus data of RSL. For the sake of completeness, we also calculated correlations between the proportions of overt indirect objects, clausal complements, and subjects, and only found low correlations which were not statistically significant. This is in agreement with our expectation, as indirect objects, clausal complements, and (especially) subjects should not reflect transitivity in general.

4. Conclusions

In this paper, we provided a first description of transitivity in RSL (that is, overt expression of objects of different types) based on the corpus of RSL. We have demonstrated that the presence and frequency of objects, indirect objects is related to the verb's semantics, as can be seen in figures 2, 3, and 4 above. RSL is thus compatible with cognitively-based accounts of transitivity, such as in Malchukov (2005). For instance, verbs of effective action such as TAKE are more likely to co-occur with an overt object than verbs of pursuit such as SEARCH or cognition, such as KNOW.

We have also shown that the transitivity ranking based on the proportion of overt direct objects in the corpus correlates well with the transitivity prominence ranking (Haspelmath, 2015). This has two theoretical consequences:

1. Transitivity is a modality-independent phenomenon, as the same verbs are likely to be transitive in spoken and signed languages.

2. The proportion of overt objects in a corpus of a single language reflects transitivity prominence.

The second consequence is also of practical importance. When analysing argument structure in sign languages based on corpus data, researchers are often limited by the lack of grammatical annotations in the existing corpora. However, if a simple measure – the proportion of overt objects – reflects transitivity in general, investigation of the basic properties of argument structure in sign language becomes relatively simple. The only annotations necessary for such an analysis are clause boundaries and labels for predicates and arguments.

In addition, if one finds a verb which, based on the proportion of overt objects, occurs in an unusual place in the ranking judging by comparison between the corpus-based and the transitivity prominence rankings, one might want to further investigate this verb as it might be a typologically exceptional item. For instance, in the RSL data analysed here, the verb SEE is unexpectedly infrequently used with an overt direct object. This can be explained by one of the two factors (or their cumulative effect): (1) this verb often occurs with clausal complements, which do not count as direct objects in our analysis and (2) the meaning of this verb might be closer to 'look' than to 'see' as it might not imply that an image of the object has been attained. Whether the latter explanation applies should be tested in future research.

5. Acknowledgements

I would like to thank Roland Pfau and Marloes Oomen for their comments on this paper. This project has been supported by the NWO (The Netherlands Organisation for Scientific Research), project number 360-70-520.

6. Bibliographical References

Benedicto, E. and Brentari, D. (2004). Where did all the arguments go?: Argument-changing properties of classifiers in ASL. *Natural Language & Linguistic Theory*, 22(4):743–810.

Burkova, S. (2015). Russian sign language corpus. http://rsl.nstu.ru/.

Hansen, M. and Hessmann, J. (2008). Matching propositional content and formal markers: Sentence boundaries in a DGS text. *Sign Language & Linguistics*, 10(2):145–175.

Hartmann, I., Haspelmath, M., and Taylor, B. (2013). *Valency Patterns Leipzig*. Max Planck Institute for Evolutionary Anthropology, Leipzig.

Haspelmath, M., Calude, A., Spagnol, M., Narrog, H., and Bamyaci, E. (2014). Coding causal–noncausal verb alternations: A form–frequency correspondence explanation. *Journal of Linguistics*, 50(03):587–625.

Haspelmath, M. (1993). More on the typology of inchoative/causative alternations. In Bernard Comrie et al., editors, *Causatives and transitivity*, pages 87–120. John Benjamins, Amsterdam.

Haspelmath, M. (2015). Transitivity prominence. In Andrej Malchukov et al., editors, *Valency Classes in the World's Languages*, pages 131–147. de Gruyter Mouton, Berlin.

Kimmelman, V., Pfau, R., and Aboh, E. O. (2016). Argument structure of Russian Sign Language classifiers: insights from corpus data. Poster presented at TISLR 12.

Malchukov, A. (2005). Case pattern splits, verb types and construction competition. In Mengistu Amberber et al., editors, *Competition and variation in natural languages: The case for case*, pages 73–117. Elsevier.

[5] Due to the presence of ties, the p-value is an approximation.

Automatic Alignment of HamNoSys Subunits for Continuous Sign Language Recognition

Oscar Koller[1,2], Hermann Ney[1] and Richard Bowden[2]
[1] Human Language Technology and Pattern Recognition Group - RWTH Aachen University, Germany
[2] Centre for Vision Speech and Signal Processing - University of Surrey, Guildford, UK
{koller,ney}@cs.rwth-aachen.de, r.bowden@surrey.ac.uk

Abstract

This work presents our recent advances in the field of automatic processing of sign language corpora targeting continuous sign language recognition. We demonstrate how generic annotations at the articulator level, such as HamNoSys, can be exploited to learn subunit classifiers. Specifically, we explore cross-language-subunits of the hand orientation modality, which are trained on isolated signs of publicly available lexicon data sets for Swiss German and Danish Sign Language and are applied to continuous sign language recognition of the challenging RWTH-PHOENIX-Weather corpus featuring German Sign Language. We observe a significant reduction in word error rate using this method.

Keywords: Sign Language Recognition, Subunits, HamNoSys, Hand Orientation

1. Introduction

Traditionally, sign language corpora intended for machine learning have been annotated at the gloss level as annotation is a time consuming and expensive process. However, glosses used as basic modelling units do not scale well with increasing corpus sizes. Furthermore, singleton signs, which have only a single token for training, make it difficult to find smooth model distributions reflecting the sign accurately. This problem, often referred to as one-shot learning, requires a single training instance to generalise over all possible variations to be encountered in the test data. Shared subunits across the different types of a corpus reduce the negative effect of singleton signs, as the composing subunits usually occur many times throughout the corpus and can therefore be robustly estimated.

Nowadays, several lexical corpus collections exist (Braem, 2001; Jette H. Kristoffersen et al., 2008 2016; McKee et al., 2015; Finish Association of the Deaf, 2015) comprising HamNoSys or other subunit transcriptions. In order to exploit and combine existing annotation efforts from different corpora, we perform automatic alignment on the subunit level. Specifically, this work explores cross-language-subunits (trained on Swiss German and Danish Sign Language) describing the hand orientation articulator. This modality has so far been mostly unexplored, due to the large variability attributed to it. In this way, we propose a method to solve the problem of missing subunit annotations, while still being able to train linguistically derived subunits. The subunit alignments may be used to train a deep convolutional neural network which model subunit representations across different data sets and even sign languages. The Convolutional Neural Network (CNN) is pre-trained and 22 layers deep. Finally, we apply the learnt models as feature extractors on our initial gloss annotated machine learning corpus and perform continuous sign language recognition of challenging real-life data on the publicly available RWTH-PHOENIX-Weather corpus (Forster et al., 2014). We observe a significant reduction in word error rate using this method.

This paper is organised as follows: after introducing the related literature in Section 2, we present the employed data sets in Section 3. We then present the proposed approach in Section 4 and evaluate it in Section 5. The paper closes with a conclusion in Section 6.

2. Related Work

There is a large body of research looking at sign subunits for sign language recognition. There are two broad classes of approaches: (i) data driven subunits, (ii) linguistically derived subunits. Both approaches have been compared to each other, with different outcomes. In Pitsikalis et al. (2011) phonetically derived subunits outperform data driven subunits by 7% on average. However, generally speaking, it is often due to missing subunit-level annotations that researchers opt for the data driven approach. Data driven approaches usually split the signs up by a segmentation algorithm, which is often based on discontinuities in hand movement velocity, such as in Theodorakis et al. (2014). In Bauer and Kraiss (2002), a limited number of signs is arbitrarily segmented which then serves as seed for either an Expectation Maximization (EM)-like iterative refinement of subunits or k-means to find subunit clusters (Kong and Ranganath, 2014). Other approaches use sparse coding to generate a sign dictionary (Yin et al., 2015).

The first sign language recognition system, presented in Tamura and Kawasaki (1988), employed linguistically derived subunits. Usually, linguistic subunit annotations provide a way to break whole signs up into constituent parts and construct a lexicon (Vogler and Metaxas, 1999; Pitsikalis et al., 2011). Other approaches use iterative EM to derive mouth-subunits from pronounced words (Koller et al., 2014). Similarly, available annotations can be aligned based on HamNoSys (Pitsikalis et al., 2011) or SignWriting (Koller et al., 2013; Koller et al., 2016) to the signed footage. The deployment of the subunit classifiers is handled differently. In Cooper et al. (2012) and Kadir et al. (2004) subunits classifiers are learnt and then combined into a second stage sign-level classifier. Systematic comparisons between subunit and whole sign modelling exist.

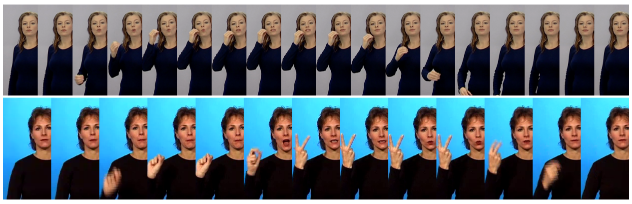


Figure 1: Showing employed data sets for training: Top to bottom, Danish sign language dictionary (Jette H. Kristoffersen et al., 2008 2016) and the Swiss German Sign Language dictionary (Braem, 2001).

In Vogler and Metaxas (1999), which is based on the movement and hold model with linguistic subunits, a powerglove hand tracker helps to perform continuous sign language recognition (CSLR). The authors conclude that sign level modelling slightly outperforms subunit modelling on a 22 sign vocab task trained on 400 sentences.
Hand location and movement are the most frequently encountered modalities used in subunit modelling schemes, closely followed by the handshapes. However, in Waldron and Kim (1995) they have been combined with 11 orientation subunits to recognise a 14 sign vocabulary.

3. Data Sets

Two different sign language dictionary data sets are employed for training the hand orientation classifier, which cover isolated signs. The first represents isolated signs from Danish Sign Language (Jette H. Kristoffersen et al., 2008 2016) with linguistic annotations, and the second features Swiss German Sign Language (Braem, 2001) with provided HamNoSys annotation (Prillwitz et al., 1989). The Danish data features high quality video footage recorded with 720×576 pixel, with very little motion blur. The Swiss German data originates from the year 2001 and is captured at a low resolution, the majority of videos being 320×240 pixel. It contains motion blur and the frames are interlaced. Figure 1 shows data examples of both sources. Both lexica provide hand orientation labels. The Danish data follows its own annotation scheme, which seems to be derived from HamNoSys. From a pattern recognition point of view annotations from both data sets are ambiguous, noisy and partly inconsistent. The chosen modality in this work are hand orientations. An isolated signed instance therefore consists of a finger orientation and a palm orientation annotation, sometimes sequences of two or more such annotations. This can be seen in Figure 1, where the top row depicts a signed instance comprising a single hand orientation and the bottom row shows an orientation sequence that transitions from “fingers:up palm:frontleft” to “fingers:up palm:back”. The signer brings his hands from a neutral position to the place of sign execution, while transitioning from a neutral hand orientation to the target hand orientation. The sign may involve a hand movement, a rotation of the hand and changes in hand shape. The annotation may represent any of these hand orientations or an intermediate configuration that was considered linguistically dominant during the annotation. It is also important to note that most linguistic annotations are done for the canonical form, which does not necessarily reflect the exact articulated instance we have access to in the video. Statistics of the two employed data sets are given in Table 1. Garbage and hand orientation frame counts are estimated automatically by our algorithm, which is done separately for finger and palm orientations. Both setups yield slightly differing numbers, which are both presented in Table 1. Both data sets jointly feature nearly $100,000$ frames of hand orientation performed by about 28 different signers.

	Danish	Swiss
duration [min]	97	200
# frames	145,720	299,864
↳ autom. orient.	32,574 / 44,432	60,643 / 55,005
↳ autom. garbage	113,146 / 101,288	239,221 / 244,859
# signed sequences	2,149	4,730
# signs	2,149	4,730
# signers	6	~ 22

Table 1: Corpus statistics: Danish (‘Danish’) and Swiss German (‘Swiss’) Sign Language data sets used for training the finger and palm orientation classifier. ‘orient.’ stands for orientation. The automatic frame counts are given for the finger orientation and the palm orientation. Therefore, two different estimated numbers are presented.

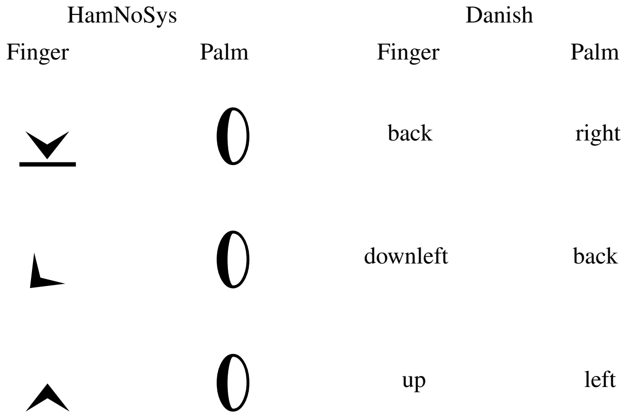


Figure 2: Showing an example mapping from HamNoSys to the Danish notation. It is apparent that in the HamNoSys annotation the palm orientation is coded in dependence of the finger orientation.

For the purpose of combining both lexicon data sets in the scope of this work we needed to create a mapping from HamNoSys to the Danish annotation. This was done manually and had to accommodate the fact that the Danish data set provided independent annotations for finger and palm orientation, whereas in HamNoSys the palm orientation is coded to be dependent on the finger orientation. This means that the same annotated palm orientation symbol can refer to different actual palm orientations depending on the current finger orientation. This is depicted in Figure 2.

After joining both annotation schemes, there is a total of 24 finger orientation classes and 26 palm orientation classes.

Finally, we evaluate on the publicly available continuous sign language data set benchmark RWTH-PHOENIX-Weather 2014 Multisigner corpus (Forster et al., 2014), which is a challenging real-life continuous sign language corpus that can be considered to be one of the largest published continuous sign language corpora. It covers unconstrained sign language of 9 different signers with a vocabulary of 1081 different signs. The data set is presented in detail in Koller et al. (2015).

4. Approach

This paper builds on our previous work (Koller et al., 2016), which is extended to the modality of hand orientations and to cover HamNoSys annotations. In the following subsections, we briefly explain the developed HamNoSys parsing, first introduce our weakly supervised learning framework and then describe how to incorporate the learnt subunit classifiers into continuous sign language recognition.

4.1. HamNoSys Parsing

The data set annotations are coded in HamNoSys, an established annotation scheme primarily developed for linguistic purposes. It contains sufficient detail to directly animate an avatar. Each sign described by HamNoSys is composed of clusters of handshape, orientation, place of articulation and movement. HamNoSys does not contain explicit segmentation information. Due to the economic writing style, HamNoSys is very minimalistic, but also needs a parsing that corrects missing information.

We first convert the HamNoSys annotations to SIGML (Glauert and Elliott, 2011). In order to be able to compensate for the palm orientations being dependent on the finger orientations, we need to ensure that a palm orientation occurs always in the context of a finger orientation. However, in transitions from a specific palm orientation to another, HamNoSys dismisses those modalities that do not change. The parser needs to take care of adding this missing information back in. After that, the mapping from HamNoSys to a non-dependant annotation scheme can be easily accomplished. An example of such a mapping is presented in Figure 2. Finally, finger orientation and palm orientation annotations are separated in order to train them as single classifiers.

4.2. Weakly Supervised Subunit Learning

Our weakly supervised CNN training algorithm constitutes a successful solution to the problem of weakly supervised learning from noisy sequence labels to correct frame labels. Figure 3 gives an overview of the approach applied to the learning of hand orientation subunits. The input images are cropped around the tracked hands, which forms the input to the weakly supervised CNN training. The iterative learning algorithm is initialised with a 'flat start', linearly partitioning the input frames to an available best guess annotation, usually a single hand orientation class preceded and followed by instances of the garbage class (as the orientation subunit is expected to happen in the middle of the sequence). The algorithm iteratively refines the temporal class boundaries and trains a CNN that performs single image hand orientation recognition (being a separate finger and palm orientation classification). While refining the boundaries, the algorithm may drop the label sequence or exchange it for one that better fits the data. The iterative process is similar to a forced alignment procedure, however, rather than using Gaussian mixtures as the probabilistic component we use the outputs of the CNN directly.

4.2.1. Problem Formulation

Following Koller et al. (2016), we have a sequence of images $x_1^T = x_1, \ldots, x_T$ and an ambiguous class label $\tilde{l}$ for the whole sequence, we want to jointly find the true label l for each frame and train a model such that the class symbol posterior probability $p(k|x)$ over all images and classes is maximised. We assume that a lexicon ψ of possible mappings from $\tilde{l} \rightarrow l$ exists, where l can be interpreted as a sequence of up to L class symbols k,

$$\psi = \left\{ \tilde{l} : l_1^L \mid l \in \{k_1, \ldots, k_N, \varnothing\} \right\} \quad (1)$$

Optionally, l may be an empty symbol corresponding to a garbage class. Each $\tilde{l}$ can map to multiple symbol sequences (which is important as $\tilde{l}$ is ambiguous and a one-to-one mapping would not be sufficient). In terms of sequence constraints, we only require each symbol to span an arbitrary length of subsequent images as we assume that symbols (in our application: hand orientation subunits) are somewhat stationary and do not instantly disappear or appear.

Due to the promising discriminatory capabilities of CNNs, we solve the problem in a iterative fashion with the EM algorithm (Dempster et al., 1977) in a Hidden-Markov-Model (HMM) setting and use the CNN to model the visual appearance of hand orientations. EM iteratively updates the assignment of class labels to images (E-Step) and then re-estimates the model parameters to adapt to the change (M-Step). We closely follow Koller et al. (2016) and, inspired by the hybrid approach (Bourlard and Morgan, 2012) known from Automatic Speech Recognition (ASR), we include the CNN's posterior output to likelihoods given the class counts in our data using Bayes' rule.

4.3. Convolutional Neural Network Architecture

Knowing the weakly supervised characteristics of our problem, we would like to incorporate as much prior knowledge as possible to guide the search for the true symbol class labels. Pre-trained CNN models constitute such a source of knowledge, which seems reasonable as the pre-trained convolutional filters in the lower layers may capture simple edges and corners, applicable to a wide range of image recognition tasks. We opt for a model previously trained in a supervised fashion for the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2014. We choose a 22 layer deep network architecture following (Szegedy et al., 2014) which achieves a top-1 accuracy of 68.7% and a top-5 accuracy 88.9% in the ILSVRC. The network involves an inception architecture, which helps to reduce the numbers of free parameters while allowing for a very deep structure. Our model has about 6 million free parameters. All convolutional layers and the last fully connected layer use rectified linear units as non-linearity. Additionally, a dropout layer with 70% ratio of dropouts is used to prevent over-fitting. We base our CNN implementation on Jia et al. (2014), which is an efficient C++ implementation using the NVIDIA CUDA Deep Neural Network GPU-accelerated library. We replace the last pre-trained fully connected layers before the output layers with those matching the number of classes in our problem (plus one garbage

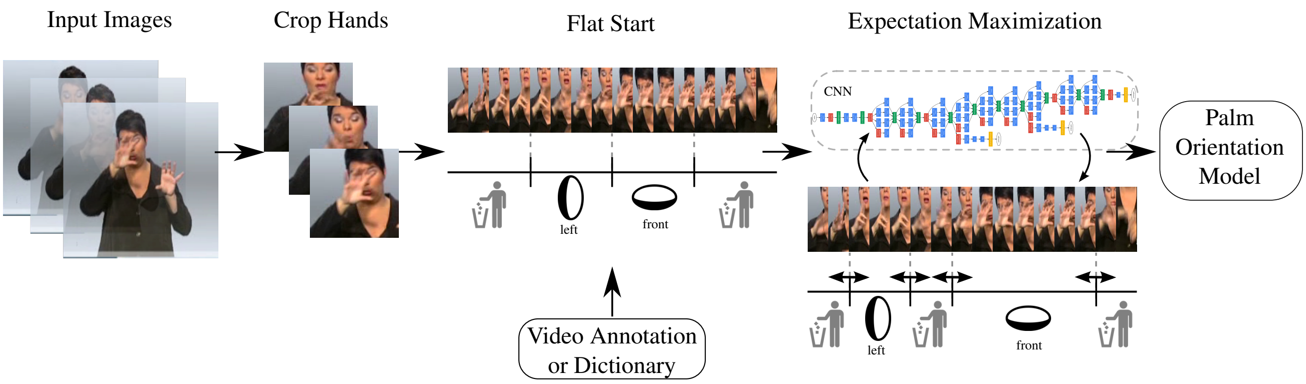


Figure 3: Overview of weakly supervised learning with HamNoSys subunits.

class), which we initialise with zeros. As a preprocessing step, we apply a global mean normalisation to the images prior to fine-tuning the CNN model with Stochastic Gradient Descent (SGD) and a softmax based cross-entropy classification loss.

4.4. Sign Language Recognition with Subunit Classifiers

In the previous subsections, we discussed learning a hand orientation classifier, based on available sign language lexicons with linguistic annotations. Continuous sign language recognition is the final task to be accomplished. However, suitable corpora (for machine learning), such as the RWTH-PHOENIX-Weather data set, do not provide subunit annotations. Therefore, we cannot apply the learned subunit classifiers directly, as there is no knowledge on how to break signs of the given corpus up into subunits.

A viable solution is to use the learned subunit classifiers as feature extractors and retrain a GMM system. This allows us to make use of the external subunit annotations (of corpora which are not intended for pattern recognition purpose) to improve the recognition on a given gloss annotated machine learning corpus, such as RWTH-PHOENIX-Weather.

The procedure is as follows: The hand orientation classifiers are trained to classify single images. During training, no sample from our target machine learning corpus was part of the training set. However, due to the CNN's ability to generalise, the unseen images forwarded through the trained network still provide good features for CSLR. With this work's experiments we will investigate if the classification results of the final softmax layer, the output scores of the last fully connected layer or the preceding last convolutional layers constitute the best features. We further evaluate how to preprocess these extracted features prior to modelling them in a standard HMM-GMM gloss-based CSLR system (Rybach et al., 2011). We compare no preprocessing to variance normalisation and dimensionality reduction by principal components analysis (PCA).

5. Experimental Results

We present the experimental evaluation in this section. In the first subsection, we focus on weakly supervised learning, whereas in the later subsection we apply the learnt subunit extractor to a state-of-the-art CSLR pipeline.

5.1. Weakly Supervised Subunit Learning

As described in the previous section, the task is to jointly estimate a good alignment for the noisy subunit labels and to model the given subunits robustly. The algorithm converges after a couple of iterations. For this work, we run it for 9 iterations. Figure 4 shows exemplar alignments of the palm orientation subunits in the initialising condition and after the last iteration. Looking at the initial alignments in the first and third line in Figure 4, we see that the majority of labels are already correctly aligned. However, at the positions where labels change, there are some alignment errors. After the convergence of the algorithm (row 2 and 4 in Figure 4), we see that all labels have been correctly aligned.

Figures 5 and 6 show the distribution of the aligned subunit classes across the nine iterations of weakly supervised learning for palm orientations and finger orientations respectively. We see that after a couple of iterations the palm orientations stabilise to four main orientation subunits (being 'left', 'front', 'down', 'back'). The training distribution of finger orientation subunits in Figure 6 look different. Here, the 'up' subunit dominates the others in terms of occurrence. Besides that, nine other finger orientations ('upleft', 'left', 'frontup', 'frontupleft', 'front', 'frontleft', 'down', 'downleft', 'backup') are less frequent in the data. This suggests that finger orientations are less stable than palm orientations. 'Stable' may refer to the production of sign language, to the annotation quality or to the modelling itself.

Within one iteration of weakly supervised learning, we continuously finetune the CNN model and measure the model's accuracy on a held out set (being 10% of the training data). We decide when to stop the CNN learning based on this accuracy. Figure 7 shows the correlation between the accuracy and the word error rate (WER). It has to be noted, however, that the WER is measured on a different data set and is therefore not directly comparable. We see a clear trend of increasing top-1 accuracy during the first training epoch (steps 1 to 8), from then onward the accuracy seems to oscillate a bit. The red WER on dev curve oscillates from the beginning, there is no clear trend visible. However, the green WER on test curve (lower is better) seems to con-

Figure 4: Palm orientation alignment visualisation. First and third rows show the sample alignments at the initialisation of the algorithm. Second and last rows show it after 9 iterations of the weakly supervised learning. It is visible how the learning helps to find a good frame alignment between the palm orientation subunits and the video footage. 'si' refers to the garbage class. Every fourth frame is shown.

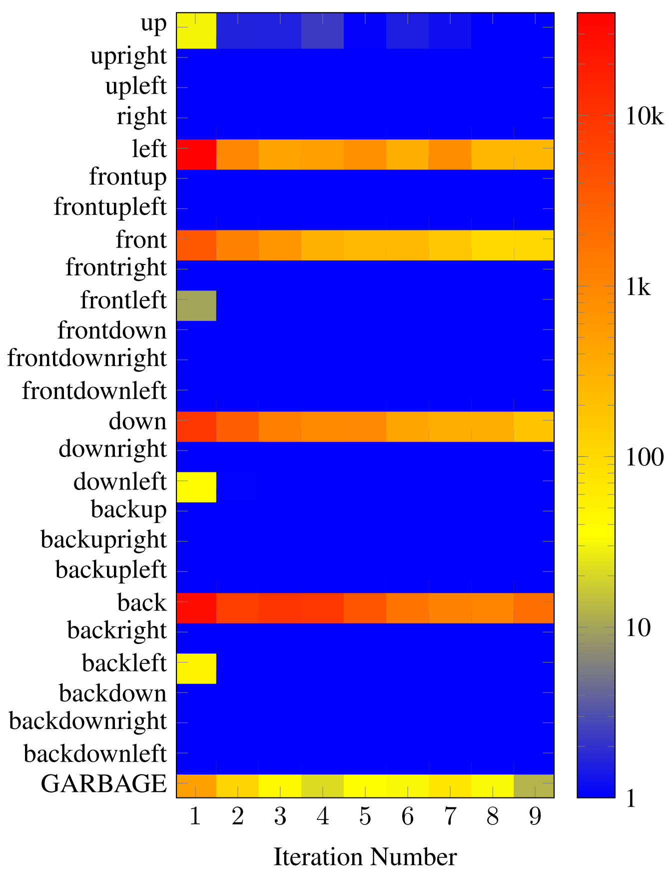


Figure 5: Figure shows unique training samples per palm orientation class as distribution across 9 iterations of the weakly supervised learning.

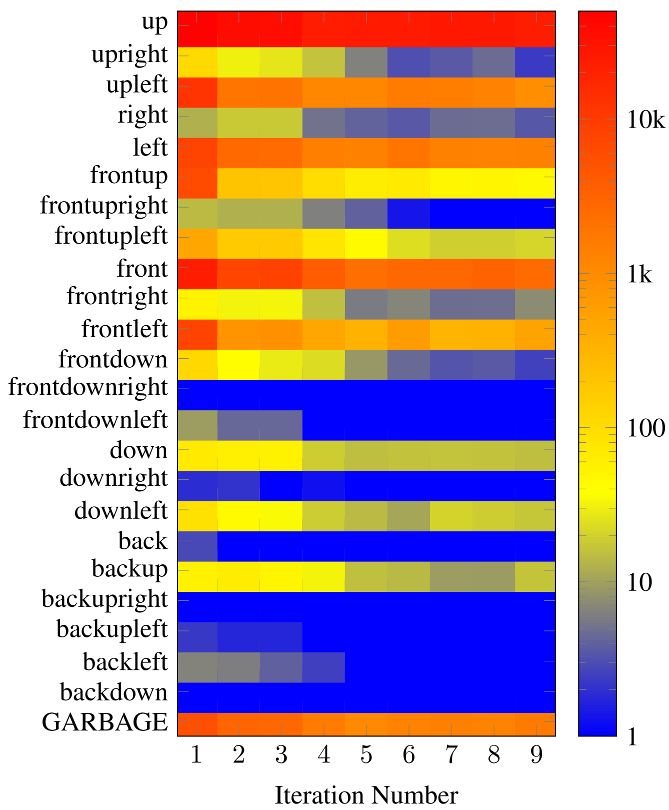


Figure 6: Figure shows unique training samples per finger orientation class as distribution across 9 iterations of the weakly supervised learning.

tinuously decrease with increasing epochs. Thus, it seems that the CSLR retraining and parameter tuning using extracted features on the development set allows us to fit the

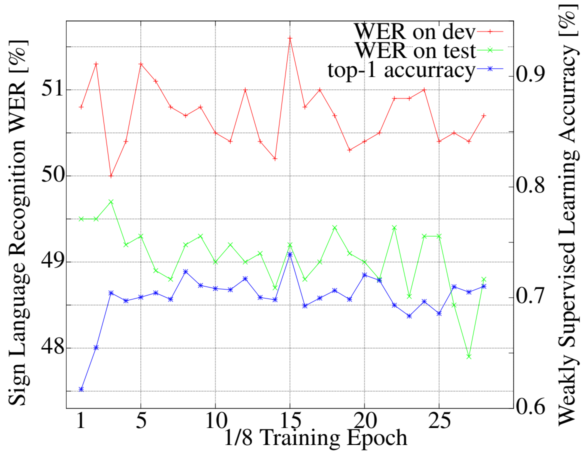


Figure 7: Correlation between WER measured on RWTH-PHOENIX-Weather test set (finger orientation + 1-Mio-Hands features) and accuracy 8 times per epoch of the weakly supervised finger orientation learning on 10% of the training data throughout the last iteration of the learning algorithm.

HMM-GMM model equally well to a worse subunit classifier. However, this then does not generalise to an unseen test set. It is good to see that with increasing CNN training this generalisation continuously improves (test WER is going down).

5.2. Continuous Sign Language Recognition

In this section we evaluate the trained hand orientation subunit classifiers integrated into a state-of-the-art continuous sign language recognition pipeline that predicts unseen sequences of RWTH-PHOENIX-Weather 2014 Multisigner. We employ the classifiers after 9 iterations of weakly supervised learning. As previously described, they have been trained solely with out-of-task (Danish and Swiss German) data. In the recognition pipeline, we use them as feature extractors, with a single input image from PHOENIX being forwarded through the CNN, which is composed of the different layers as mentioned in Section 4.3. The forward pass can be stopped at any of them and the output constitutes the extracted features. Table 2 compares the performance of using features originating from different layers of the trained subunit CNN. Not being trained on the task data directly (as no subunit annotations are available for PHOENIX), we expect the final classification output to be quite noisy. This assumption seems to hold, as the softmax features (lines 1-3 in Table 2) are all largely outperformed by the other layers. The last fully connected output (denoted as 'last FC' in Table 2) performs marginally better (when being variance normalised across the whole data set) than the PCA reduced output of the last convolutional layer (actually, the output of the last pooling layer, which has 1024 dimensions). For subsequent experiments, we perform recognition always with features originating from the last fully connected layer, which are then variance normalised.

We further analyse which CNN model initialisation scheme we should follow. Table 3 compares two different pre-training schemes, which either rely on the imagenet data set (over one million hand labelled objects from 1,000 categories, very diverse in size, appearance and capture conditions) or on the 1-Mio-Hands model, which was trained to distinguish handshapes orientation independently using more than one million hand images from the Danish, New Zealand and German Sign Language (see (Koller et al., 2016) for details.) We denote that the 1-Mio-Hands model helps to learn a better stand-alone subunit classifier. However, when combined with the original initialisation model and applied to the CSLR task, it lacks complementary information. We therefore use Imagenet to pre-train our CNNs in all subsequent experiments.

	Extraction Layer					WER	
	Softmax	Last FC	Last Conv	Var Norm	Dim	Dev	Test
1	X			no	26	65.0	63.7
2	X			yes	26	63.9	63.4
3	X			yes	26pca	52.1	51.3
4		X		yes	26pca	51.9	50.3
5		X		no	26	50.4	49.4
6			X	yes	50pca	50.4	**48.2**
7		X		yes	26	**50.1**	48.3

Table 2: Comparing different feature extractor layers. All experiments represent the palm orientation in feature combination (stacking) with the 1-Mio-Hands classifier (Koller et al., 2016). 'Dim' stands for dimension, 'var norm' for variance normalisation, 'fc' for fully connected layer, 'conv' for convolutional layer. WER in [%].

	Initialisation		WER [%]	
	Imagenet	1-Mio-Hands	Dev	Test
1-Mio-Hands alone	X		51.6	50.2
Subunit alone		X	53.1	53.0
Subunit alone	X		72.9	72.4
Subunit + 1-Mio-Hands		X	50.2	50.3
Subunit + 1-Mio-Hands	X		50.8	49.6

Table 3: Impact of initialisation. An initialisation from a better model trained on the same data yields a better stand-alone classifier, but lacks complementary information in combination with the original initialisation model. Results on RWTH-PHOENIX-Weather 2014 Multisigner. 1-Mio-Hands have been presented in Koller et al. (2016). '+' denotes feature stacking prior to GMM-HMM training.

Table 4 compares the finger and the palm orientation classifiers and their combination. We see that the palm orientation outperforms the finger orientation and as expected both contain complementary information with respect to each other, as their fusion is clearly better than each classifier alone. Moreover, both orientation subunit classifiers add complementary information to the strong 1-Mio-Hand handshape baseline, which improves from 51.6% → 49.6% on dev and from 50.2% → 48.2% on test.

Table 5 shows how much complementary information the hand orientation classifiers add to a strong multi-modal baseline consisting of jointly modelled (stacked) features from Koller et al. (2015) (being HoG3D, right to left hand distance, movement/trajectory of dominant hand, place of

	Dev		Test	
	del/ins	WER	del/ins	WER
HoG-3D	25.8/4.2	60.9	23.2/4.1	58.1
1-Mio-Hands	19.1/4.1	51.6	17.5/4.5	50.2
Finger orientation	33.0/3.1	72.9	31.3/3.1	72.4
Palm orientation	25.4/4.1	68.7	24.4/4.5	66.9
Finger + Palm	26.3/3.3	63.8	24.3/3.3	62.3
Finger + 1-Mio-Hands	16.3/5.3	50.8	15.0/5.6	49.6
Palm + 1-Mio-Hands	17.5/4.6	50.1	16.0/4.6	48.3
Finger+Palm+1-Mio-Hands	17.5/4.7	**49.6**	15.9/4.6	**48.2**

Table 4: Hand-only continuous sign language recognition results on RWTH-PHOENIX-Weather 2014 Multisigner. 1-Mio-Hands have been presented in Koller et al. (2016). '+' denotes feature stacking prior to GMM-HMM training. WER in [%].

articulation normalised by the nose and facial features) and the 1-Mio-Hand handshape features from Koller et al. (2016). We note, that the orientation subunits can improve the result on the dev set, but the improvement does not carry over to the test set. Most likely, more data would be required to achieve a better generalisation. Including RWTH-PHOENIX-Weather into the subunit training stage may also boost results (cf. (Koller et al., 2016)).

		Dev		Test	
		del/ins	WER	del/ins	WER
1	(Koller et al., 2015) cmllr	21.8/3.9	55.0	20.3/4.5	53.0
2	(Koller et al., 2015)	23.6/4.0	57.3	23.1/4.4	55.6
3	+ 1-Mio-Hands	16.3/4.6	47.1	15.2/4.6	**45.1**
4	+ Finger + Palm	18.0/4.5	**46.6**	16.5/4.8	45.5

Table 5: Multi-modal continuous sign language recognition results on RWTH-PHOENIX-Weather 2014 Multisigner. 1-Mio-Hands have been presented in Koller et al. (2016). '+' denotes feature stacking prior to GMM-HMM training. WER in [%].

6. Conclusion

In this work, we presented our recent advances in the field of subunit modelling for continuous sign language recognition. We demonstrated how generic annotations at the articulator level, such as HamNoSys, can be exploited to learn subunit classifiers. We explored cross-language-subunits, which were trained on isolated signs of publicly available lexicon data sets for Swiss German and Danish Sign Language. We therefore employed a weakly supervised learning framework that helped to jointly find those subunits that occur in the data and to model them robustly.

We analysed the alignment of the weakly supervised learning, finding that palm orientations seem to be more stable than finger orientations. Furthermore, we systematically determine the best extraction scheme to include the learnt CNN as feature extractors in a GMM-HMM system. Finally, we evaluated palm orientation and finger orientation subunits to perform CSLR on the publicly available RWTH-PHOENIX-Weather corpus (Forster et al., 2014). We find that the modalities improve a handshape only system by 2% absolute WER, while still improving a multimodal baseline (manual and non-manual features) by 0.5%.

7. Acknowledgements

We thank Penny Boyes Braem and Sarah Ebling for enabling the access to the Swiss German data which was crucial for this work. We further thank Thomas Troelsgård and Jette H. Kristoffersen for providing the employed Danish Sign Language annotations and videos.

Bauer, B. and Kraiss, K. F. (2002). Video-based sign recognition using self-organizing subunits. In *16th International Conference on Pattern Recognition, 2002. Proceedings*, volume 2, pages 434–437. IEEE.

Bourlard, H. A. and Morgan, N. (2012). *Connectionist speech recognition: a hybrid approach*, volume 247. Springer Science & Business Media.

Braem, P. B. (2001). A multimedia bilingual database for the lexicon of Swiss German Sign Language. *Sign Language & Linguistics*, 4(1/2):133–143.

Cooper, H., Ong, E.-J., Pugeault, N., and Bowden, R. (2012). Sign language recognition using sub-units. *The Journal of Machine Learning Research*, 13(1):2205–2231.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

Finish Association of the Deaf. (2015). Suvi Viittomat. `http://suvi.viittomat.net/`.

Forster, J., Schmidt, C., Koller, O., Bellgardt, M., and Ney, H. (2014). Extensions of the Sign Language Recognition and Translation Corpus RWTH-PHOENIX-Weather. In *Language Resources and Evaluation*, pages 1911–1916, Reykjavik, Island, May.

Glauert, J. and Elliott, R. (2011). Extending the SiGML notation—a progress report. In *Second International Workshop on Sign Language Translation and Avatar Technology (SLTAT)*, volume 23.

Jette H. Kristoffersen, Thomas Troelsgård, Anne Skov Hardell, Bo Hardell, Janne Boye Niemelä, Jørgen Sandholt, and Maja Toft. (2008-2016). Ordbog over Dansk Tegnsprog. `http://www.tegnsprog.dk/`.

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. (2014). Caffe: Convolutional Architecture for Fast Feature Embedding. *arXiv preprint arXiv:1408.5093*.

Kadir, T., Bowden, R., Ong, E.-J., and Zisserman, A. (2004). Minimal Training, Large Lexicon, Unconstrained Sign Language Recognition. In *BMVC*, pages 1–10. 00079.

Koller, O., Ney, H., and Bowden, R. (2013). May the Force be with you: Force-Aligned SignWriting for Automatic Subunit Annotation of Corpora. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 1–6, Shanghai, PRC, April.

Koller, O., Ney, H., and Bowden, R. (2014). Read My Lips: Continuous Signer Independent Weakly Supervised Viseme Recognition. In *Proceedings of the 13th*

European Conference on Computer Vision, pages 281–296, Zurich, Switzerland, September.

Koller, O., Forster, J., and Ney, H. (2015). Continuous sign language recognition: Towards large vocabulary statistical recognition systems handling multiple signers. *Computer Vision and Image Understanding*, 141:108–125, December.

Koller, O., Ney, H., and Bowden, R. (2016). Deep Hand: How to Train a CNN on 1 Million Hand Images When Your Data Is Continuous and Weakly Labelled. In *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, June.

Kong, W. W. and Ranganath, S. (2014). Towards subject independent continuous sign language recognition: A segment and merge approach. *Pattern Recognition*, 47(3):1294–1308, March.

McKee, D., McKee, R., Alexander, S. P., and Pivac, L. (2015). The Online Dictionary of New Zealand Sign Language. `http://nzsl.vuw.ac.nz/`.

Pitsikalis, V., Theodorakis, S., Vogler, C., and Maragos, P. (2011). Advances in phonetics-based sub-unit modeling for transcription alignment and sign language recognition. In *2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–6, June.

Prillwitz, S., Leven, R., Zienert, H., Zienert, R., and Hanke, T. (1989). *HamNoSys. Version 2.0.* Int'l Studies on Sign Language and Communication of the Deaf. Signum, Hamburg.

Rybach, D., Hahn, S., Lehnen, P., Nolden, D., Sundermeyer, M., Tüske, Z., Wiesler, S., Schlüter, R., and Ney, H. (2011). RASR - The RWTH Aachen University Open Source Speech Recognition Toolkit. In *IEEE Automatic Speech Recognition and Understanding Workshop*, Waikoloa, HI, USA, December.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2014). Going Deeper with Convolutions. *arXiv:1409.4842 [cs]*, September.

Tamura, S. and Kawasaki, S. (1988). Recognition of sign language motion images. *Pattern Recognition*, 21(4):343–353.

Theodorakis, S., Pitsikalis, V., and Maragos, P. (2014). Dynamic–static unsupervised sequentiality, statistical subunits and lexicon for sign language recognition. *Image and Vision Computing*, 32(8):533–549. 00003.

Vogler, C. and Metaxas, D. (1999). Toward scalability in ASL recognition: Breaking down signs into phonemes. *Gesture-Based Communication in Human-Computer Interaction*, pages 211–224.

Waldron, M. B. and Kim, S. (1995). Isolated ASL sign recognition system for deaf persons. *IEEE Transactions on Rehabilitation Engineering*, pages 261–271.

Yin, F., Chai, X., Zhou, Y., and Chen, X. (2015). Semantics constrained dictionary learning for signer-independent sign language recognition. In *Image Processing (ICIP)*, pages 3310–3314. IEEE.

Evaluating User Experience of the Online Dictionary of the Slovenian Sign Language

Ines Kožuh[1], Primož Kosec[2], Matjaž Debevc[1]
[1]University of Maribor, Faculty of Electrical Engineering and Computer Science
Smetanova 17, 2000 Maribor, Slovenia
[2]Nuimo
Koroška cesta 219, 2351 Kamnica, Slovenia
E-mail: ines.kozuh@um.si, primoz@nuimo.si, matjaz.debevc@um.si

Abstract

The extensive use of mobile devices and tablets has resulted in an increasing need for the ubiquitous availability of different types of dictionaries online. The purpose of our study was to evaluate the user experience and usability of the online dictionary of the Slovenian Sign Language. Six Slovenian hearing non-signers were included in the study. While using the online dictionary, participants were asked to complete six tasks: searching for a letter, a word, written explanation of the word, thematic section and particular fairy tale, as well as completing the quiz. In addition, the participants evaluated the usability of the online dictionary with the System Usability Scale. The findings revealed that participants perceived the tasks "searching for the word" and "searching for the thematic section" to be the most difficult in comparison with other tasks. In contrast, they found completing the quiz to be the easiest one. Regarding the time measured, the task "searching for the word" was the most time-consuming (29.17 seconds) and the task "searching for the letter" was the least time-consuming (10.75 seconds). This study provides interesting insights into how Slovenian hearing users perceive using the online dictionary of the Slovenian Sign Language, which could be a basis for future research with D/deaf and hard-of-hearing users of the Slovenian Sign Language.

Keywords: online dictionary, sign language, usability, evaluation

1. Introduction

In recent years, the need for the availability of different types of dictionaries online has increased. With the spread of mobile devices, the requirements for the availability of these dictionaries have expanded to them. For sign language users, dictionaries on mobile devices could thus be a significant education tool.

Previous studies, mainly conducted for American Sign Language, showed that d/Deaf people may have several problems when reading in the language of their local hearing community. For instance, Traxler (2000) determined that half of deaf high school students read at the 4th-grade level or below. It is thus necessary not only for those who use sign language as their primary mode of communication but also for those whose primary language is the same as the written language to have access to a sign language dictionary. It may improve communication between both parties and help people understand unknown written words and phrases.

Existing solutions for sign language dictionaries are available on web pages and in mobile applications (Jones, 2015). These on-line and off-line dictionaries are mostly used in such a way that one particular word is chosen among others, or it is inserted in a textbox for translation in sign language to be displayed. For instance, Jones (2015) introduced a mobile application that allows a user to point the mobile phone camera at a page of text, take a picture and then click on a word to access the definition. The definition is then displayed as a video streamed from a YouTube, a social video-based network.

Unfortunately, these dictionaries are mainly available only for American Sign Language and are web-based (i.e. only available for use on web pages). A lack of Android-based dictionaries is evident, especially in non-English speaking countries. For instance, in Slovenia, there was no dictionary of Slovenian Sign Language available as a mobile application.

Thus, in this paper, we present the online dictionary of the Slovenian Sign Language developed by the University of Maribor and the company NUIMO in close cooperation with the Slovenian Association of the Deaf and Hard of Hearing. The software allows searching for sign language interpretation of words in Slovenian Sign Language both from a list of words and by typing words in a textbox. Moreover, in contrast to other existing dictionaries, we created quick access to a set of words collected in the most important categories. These categories can be built either automatically by the system or by users themselves. This software is available for both websites (browsers) and mobile devices (apps).

Furthermore, in this paper, we present the user experience and usability evaluation of the online dictionary of the Slovenian Sign Language. The paper is organised as follows. First, we introduce the online dictionary, next, the methods and procedure of the study are described in detail. Moreover, the results are presented and discussed. Finally, some conclusions are based on the lessons learned.

2. The Online Dictionary of the Slovenian Sign Language

2.1 Front-End

There are three elements on the front-end: a search engine, a list of topics, and a quiz (see Figure 1).

The users can search for a term with the help of a smart search feature, which is able to automatically handle the user's input and show the available terms within the dictionary. When a term has been selected, the user is redirected to the video unit (Figure 2).
The video unit is the core element of the dictionary. It is composed of several elements: video links and metadata. The video links include a standard term interpretation in sign language and a list of links for different term variations. The metadata represent written information, such as synonyms, written sentences, sign language explanations in the online Dictionary of the Slovenian Sign Language, and explanations in written form from the Dictionary of the Slovenian Literary Language. The video player uses a two-fold video-playing technique. Primarily, video is played with a Flash player. For non-Flash compatible browsers, HTML5 video has been implemented.
Another possible scenario to find a video unit within the list of topics. On the index page, two lists are available: popular topics and selected topics list. Popular topics are shown based on the impressions (user clicks), while the selected topics are controlled from the back-end by a moderator. When the user clicks on a topic, a list of all videos within the selected topic is shown.
The quiz was designed for practicing the written and signed language. Each quiz unit is composed of a sign language video, which poses a question, and multiple written answers, of which only one is correct. The user selects an answer and is moved to the next quiz unit. At the end of the quiz, the score of correct vs. wrong answers is computed and displayed.

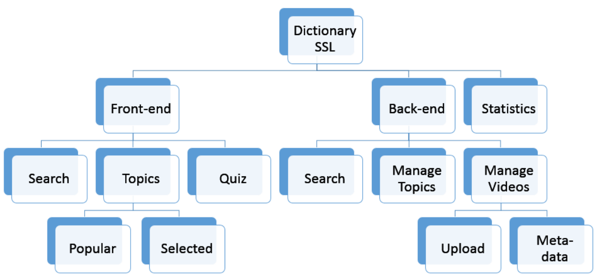


Figure 1: Architecture of the online Dictionary of Slovenian Sign Language.

2.2 Back-End

From the back-end, the moderator controls the topics, videos, and metadata. Videos and topics can be found in a search engine by using filters. The videos are uploaded to a dedicated video server from which they are streamed.

2.3 Statistics

The system uses internal and external statistics. The internal statistics log all user interactions such as video clicks, topic clicks, and participations in quizzes, while the external statistics use Google Analytics to obtain additional insight.

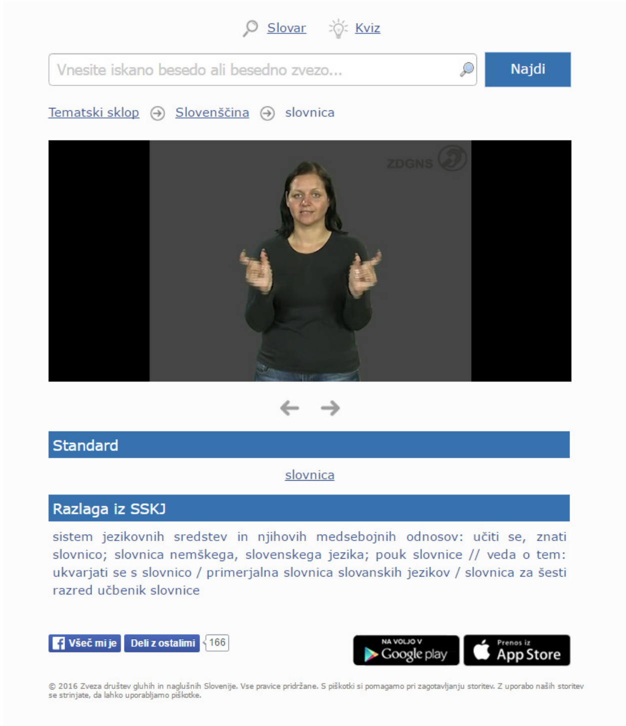


Figure 2: Video unit of the online Dictionary of the Slovenian Sign Language (http://www.szj.si; 2016).

3. Methods

3.1 Participants

Six participants were included in the study. They were master students of media communications, aged between 23 and 25 years; five were female and one male. They reported no hearing loss. All of them had graduated in media communications.

Skills	**M (mean)**	**SD (standard deviation)**
Written language skills	4.39	0.54
Sign language skills	0.56	0.64
ICT skills – PC	4.83	0.37
ICT skills – Tablet	4.50	0.76
ICT skills – smartphone	4.33	0.75

Table 1: Demographic data about the sample.

3.2 Measures

The measuring instruments used in the study were: a demographic questionnaire, the System Usability Scale (SUS) and an assessment instrument for observation.

3.2.1 Demographic questionnaire

The demographic questionnaire contained eight questions; participants provided information about their genders, ages, education, types of hearing loss if any, sign and written language skills, skills of using information-communication technology (ICT), and perceived ease of completing the tasks in the experimental session.

Hearing loss was measured according to the American National Standards Institute (ANSI, 2010), in which an unaided hearing loss of 27 dB is considered to be a threshold for hard of hearing people, while people with an unaided hearing loss of 91 dB are considered to be deaf.

Both sign and written language skills were assessed, each with three question items administered with five-point Likert-type response categories ranging from 1 (very poor) to 5 (excellent). The questions were developed on the basis of the adjusted Deaf Acculturation Scale (DAS) (Maxwell-McCaw & Zea, 2011). Item example: "How well do you understand Slovenian Sign Language?"

Skills of using ICT were assessed with three question items administered with five-point Likert-type response categories ranging from 1 (not using at all) to 5 (can use very well). The questions covered personal computers, tablets, and smartphones.

Perceived ease of completing the tasks in the experimental session was measured with six question items administered with five-point Likert-type response categories ranging from 1 (extremely difficult) to 5 (extremely easy). One question item was provided for each task described later in Section 3.3.

3.2.2 SUS Questionnaire

The purpose of the SUS method was to evaluate the usability of the online Dictionary of the Slovenian Sign Language. The SUS questionnaire, translated into the Slovene language by Assoc. Prof. Dr Matjaž Debevc and Dr Ines Kožuh from the University of Maribor, as published in Kožuh (2010), was used.

Every participant was asked to go through the system and evaluate it according to 10 Likert-type statements, ranging from 1 (strongly disagree) to 5 (strongly agree).

The calculation procedure according to SUS method (Brooke, 1996; as cited in Kožuh, 2010):

(1) Values of individual items are total.
(2) The value of an individual item is ranked on a scale between 0 and 4.
(3) For items 1, 3, 5, 7 and 9, the value is calculated so that 1 is subtracted from the answer value of the scale.
(4) For items 2, 4, 6, 8 and 10, the value is calculated so that the value (position) is subtracted from 5.
(5) The sum of all items is multiplied by 2.5 so that the overall value of system usage is obtained.

Based on the abovementioned procedure, the output is expressed in points between 0 and 100. The closer the grade to 100 is, the better the quality of the system is. Figure 3 shows the SUS scale between 0 and 100 with an explanation of scores (Kožuh, 2010).

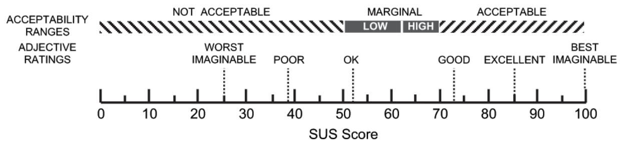


Figure 3: Explanation of the SUS score (Brooke, 1996; cited in Kožuh, 2010).

3.2.3 An assessment instrument for observation

The assessment instrument for observation aided the observer in noting the time each student spent on each task. All tasks were listed on the paper where the observer was allowed to write down number of seconds spent on each task.

3.3 Procedure

The experiment was conducted in January 2016 at the University of Maribor. Although the dictionary is available both as a mobile and web application, in our experiment we used only the latter. The experiment comprised three sessions:

(1) Introduction to the experiment.

Prior to the experimental session, the participants were informed about the procedure of the experiment. Next, they signed a written consent form.

(2) Experimental session.

Participants were asked to perform six tasks and the time of completing each task was measured. The tasks were:

- searching for the word "Australia",
- searching for the letter "F",
- searching for the thematic section "Children's signs",
- searching for a written explanation of the word "Academy",
- searching for the fairy tale "Peter Klepec", and completing the quiz "Animals".

While completing the tasks, the time spent for each task was measured.

(3) Evaluation session.

Participants completed two questionnaires. First, the SUS questionnaire, because it is known to be a reliable, quick tool for measuring usability (Brooke, 1996). Second, participants complete a questionnaire in which they provided information about their demographic data and background regarding sign and written language use, as well as ICT technology.

4. Results

The results from the experimental and evaluation session revealed that participants perceived the tasks "searching for the word" (M = 3.29, SD = 1.03) and "searching for the thematic section" (M = 3.29, SD = 1.16) to be the most difficult compared to other tasks (see Table 2). However, they found completing the quiz to be the easiest of all the assigned tasks (M = 4.43, SD = 0.90).

Task	Perceived ease of completing the tasks		Duration (seconds)
	M	SD	
1. searching for the word	3.29	1.03	29.17
2. searching for the letter	3.86	1.25	10.75
3. searching for the thematic section	3.29	1.16	15.00
4. searching for the written explanation of the word	4.00	0.53	18.17
5. searching for the fairy tale	3.86	1.36	14.20
6. completing the quiz	4.43	0.90	25.6

Table 2: Perceived ease of use and duration of completing the assigned tasks.

Regarding time measured, we found out that the task “searching for the word” to be the most time-consuming (29.17 seconds). This result can be understood in the context that this was the first activity conducted by participants; it may have taken some time for users to become familiar with the system. The least amount of time was spent on searching for a letter within the online dictionary (10.75 seconds), which was expected since this task was the least complex of all.

Intriguingly, participants perceived the task “completing the quiz” to be the easiest for use, although this task took them only 3.57 seconds less than the task “searching for the word”, which was the most time-consuming task. Also relatively well assessed were the tasks “searching for the written explanation of the word”, “searching for the letter”, and “searching for the fairy tale”. The task “searching for the thematic section” has the lowest score in perceived use of completing the task (M = 4.00, SD = .53).

Analysis of individual SUS scores revealed that the final SUS score is 69.2. According to Figure 2, the result can be interpreted such as marginally acceptable, although it is close to the level acceptable. Among individual SUS scores, the highest score was 95.0, while the lowest was 40.0.

5. Discussion and Conclusion

In this paper, an overview of the online dictionary of the Slovenian Sign Language along with lessons learned from usability and user experience evaluation of the system has been provided. The results showed that the first task, in which participants had to find a particular word, was the most time-consuming. This is reasonable due to the sequence of activities, although the task was not complex compared to other tasks, such as “searching for the fairy tale”.

Considering the perceived ease of completing tasks and the duration of completing these tasks, we found no consistency in terms of the principle of the better-perceived ease of completing tasks being reflected in the shorter duration of completing the tasks. The SUS method revealed that the system was marginally acceptable, while the score was close to the threshold for an acceptable system.

Our study has a few limitations. For instance, one stems from the self-reporting used in the questionnaire to collect data. Consequently, we cannot exactly know what language skills participants actually have and how difficult the tasks actually were. In addition, the study is limited in the number of participants included, which provides opportunities for future research where more participants are proposed to be involved. Furthermore, in the future, it would be intriguing to conduct the experiment with D/deaf and hard-of-hearing sign language users in which the mobile app would also be tested.

In the study presented in the current paper we did not include D/deaf and hard-of-hearing people since it was only a pilot study. Its intent was to solicit feedback from a small number of participants in terms of understanding the tasks and finding any ambiguity which may appear in instructions. These pieces of information may help us further improve the measuring instruments before the actual study with D/deaf and hard-of-hearing people would be held. For instance, in this study we received information on how the sequence of tasks should follow each other in order to proceed from less to more complex tasks.

6. Acknowledgements

We would like to thank the Slovenian Association of the Deaf and Hard of Hearing for the availability of the online dictionary of the Slovenian Sign Language for evaluation. Moreover, we acknowledge the students of Media Communications of the Faculty of Electrical Engineering and Computer Science at the University of Maribor for their participation in the study.

7. Bibliographical References

Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In: P. W. Jordan, B. Thomas, B. A. Weerdmeester in A. L. McClelland. *Usability Evaluation in Industry*. London: Taylor and Francis.

Kožuh, I. (2010). Načrtovanje in ocenjevanje spletnega učnega mesta za gluhe in naglušne [Design and evalution of the E-learning portal for the deaf and hard of hearing]. Diploma thesis. Maribor: University of Maribor.

Slovenian Association of the Deaf and Hard of Hearing. (2016). Online dictionary of the Slovenian Sign Language. Available on: http://www.szj.si/

American National Standards Institute. *Specifications for Audiometers*. New York, NY: American National Standards Institute; 2010. ANSI publication ANSI/ASA S3.6-2010.

Maxwell-McCaw D, Zea MC. The deaf acculturation scale (DAS): Development and validation of a 58-item measure. *Journal of Deaf Studies and Deaf Education* 2011;16:325–42.

Jones, M.D., Hamilton, H., and Petmecky, J. (2015). Mobile Phone Access to a Sign Language Dictionary. In Proceedings of the *17th International ACM SIGACCESS Conference on Computers & Accessibility* (ASSETS '15). ACM, New York, NY, USA, 331-332. DOI=http://dx.doi.org/10.1145/2700648.281136

Semiautomatic Data Glove Calibration for Sign Language Corpora Building

Zdeněk Krňoul, Jakub Kanis, Miloš Železný, Luděk Müller
University of West Bohemia, Faculty of Applied Sciences, NTIS - New Technologies for the Information Society
Univerzitní 8, 306 14 Pilsen, Czech Republic
zdkrnoul@ntis.zcu.cz, jkanis@ntis.zcu.cz, zelezny@ntis.zcu.cz, muller@ntis.zcu.cz

Abstract

The article deals with a recording procedure for sign language dataset building mainly for avatar synthesis systems. Combined data glove and optical capture technique is considered. We present initial experiences with the motion capture data produced by the CyberGlove3 gloves and a set of new tools to ease the recording process, glove calibration and proper interpretation by the 3D model. It results in a more flexible solution for the sign language capture integrating manual glove calibration with an automatic initialization, time synchronization and high-resolution sensor readings.

Keywords: data glove calibration, sign language, motion capture

1. Introduction

Whilst data processing as tracking of markers, 3D reconstruction or fitting a generic skeleton model provided by the VICON motion capture system are very beneficial for capturing of sign languages, there is no such option for a processing of the CyberGlove3 motion data. In this scenario, the capturing of sign languages consists of two steps: (1) the glove calibration to angular motion data of an internal hand model/skeleton at the beginning of the capturing session and (2) re-targeting of the angular motion data to a target avatar as the data post-processing step. Full automatic calibration approaches including the CyberGlove3 (Carmel et al., 2014) are often rejected as an incomplete solution (Wang and Neff, 2013), (Kahlesz et al., 2004), unaddressed visual fidelity of hand shapes (Griffin et al., 2000) and/or complexity of the calibration process (Elliott et al., 2008). Therefore, higher flexibility in the recording procedure using CyberGlove3 gloves is still a research issue.

The calibration process consists of finding the conversion relationship between raw data of the sensor and the actual bending (rotation in one axis) of the finger segment (in degrees). The CyberGlove3 software package includes Virtual hand tool (Carmel et al., 2014) that provides communication of the data glove with the computer. The main part of this tool is the Device Configuration Utility (DCU) that provides a user interface for glove calibration. The sensor maximum standard deviation nonlinearity is 0.6 % over the full joint range (Carmel et al., 2014) and the conversion relationship is approximated by a linear transformation, i.e. a scale (a) and an offset (b) of a line equation:

$$Y = a \cdot X + b \tag{1}$$

The DCU offers an automatic calibration of the linear relationship identified just from two predefined hand shapes. Theoretically, the technique is sufficient since two different finger bends clearly determine two points in a line (i.e. linear relationship of the sensor). But the result provided by the DCU is very inefficient. As was noted earlier in (Huenerfauth and Lu, 2010), the middle, the ring and the pinky finger abductions as well as flexion are incorrectly identified from the hand shapes (a flat hand and the touch of the thumb and the index finger) because there are two ambiguous sensor readings.

The second and preferred option of the DCU is a manual calibration. The initial experiment with sign language capturing shows that it is possible to calibrate basic (simple) hand shapes like a fist or a flat palm. However, the very laborious work is calibration of the touch of the thumb/index finger and the touch of the thumb/pinky finger was not achieved anyway. In this context, problems of sensor cross-coupling are often discussed that single finger segment bend can influence multiple sensors, or that some sensors measure a different motion of the hand. The cross-coupling is solved for the finger abduction sensors (Steffen et al., 2011) or the thumb roll/abduction (Wang and Neff, 2013) for CyberGlove2. The next issue is a reuse of the calibration parameters after re-dressing of the gloves by the same subject. In this case, an average standard deviation is only 3° for a single sensor (Carmel et al., 2014), nevertheless the previously identified calibration does not match key hand shapes. This inconsistency must be taking into account while creating of the sign language corpora.

In the paper, we present initial experiences and custom tools that address the weaknesses of the current sign language recording process: allow flexible setting of the data glove parameter; time synchronization between left, right glove and body motion capturing; an automatic initialization of the calibration parameters and its manual refinement; and high resolution sensor readings of CyberGlove3 with 12-bit A/D conversion instead of precision loss by the DCU.

2. Combining Optical and Data Glove Recording

The combination of the optical and the data glove motion capturing is one possible sign language recording technique. The measurement principle of the finger bending is based on the resistive sensors that provide robust measurements of finger contacts on one or mutually between hands. In addition, the CyberGlove3 glove measures palm flex and wrist rotation like pitch and yaw. On the other hand, the

reading of one sensor is relative to the reading of the preceding finger segment or the wrist and thus we never get absolute 3D positions. Thus, the CyberGlove3 motion capture data are relative to the 3D position of the forearm.
It is necessary to determine which part of the arm will be a link for two types of the motion capture data. The first option is to use a mapping of the wrist pitch and jaw sensor to the target model and the VICON system determining only the wrist position and the forearm twist. The second and also preferred option is tracking full/global wrist rotation as *finger direction* and *palm rotation* by the VICON system. In this case, at least two markers on the hand back have to be added to two markers placed on the wrist joint. The wrist pitch and yaw of the data glove are then ignored.

3. New Data Glove Recording Procedure

We consider the combined record dataset building in two phases: (1) recording phase for raw glove motion data recording without glove calibration, (2) processing phase allowing interpretation of the raw data and including the glove calibration. For the first phase, we developed a new tool for communication with the CyberGlove3 gloves. The tool provides an interface for recording with one or two (left and right) gloves at a time and also enables necessary time synchronization between the gloves and the VICON system. For the second phase of sign language dataset building, we develop scripts for converting the raw data records in the native format to a motion data standard file format.

3.1. CyberGlove3 Recording Tool

The CyberGlove3 recording tool consists of two Python scripts to control recording with CyberGlove3 gloves. The first one is for the recording with one glove and the second one allows time-synchronized recording of both gloves. A particular command is sent to the both gloves at the same time but it can be executed by each glove with a slightly different delay depending on the processing unit of each glove. To time-synchronize the data recording we set the same internal time for both gloves by one command and then start the recording with another one. A joint time setting allows us to reach the time difference between gloves in a range of a one time frame, i.e. 0.0333 s because there are 30 time frames for each second. But the time difference can be greater and we have to keep the internal time setting until the difference is acceptable. As soon as it is acceptable, we can start the recording.
Another useful and important feature of the script is a possibility of a supporting video recording of the calibration take by a connected web camera. The exact video recording of the calibration take is important for an accurate and clear glove calibration. The control script allows the following functionality: display battery voltage, display and set internal time, display and set active (recorded) file name, list of all memory card files, start and stop recording and enabling of the video recording. Records of the glove data are stored on an internal memory card in the processing unit of each glove. In general, glove motion data can be on-line streamed to host computer and or stored on the memory card. Nevertheless according to our experience, the Wi-Fi on-line stream is not robust enough for reliable recording of large data needed for the corpus. The data are in CyberGlove3 native binary format as raw responses of the glove sensors. In advance, this option provides the recorded data without transformation by the built-in calibration of the CyberGlove3 software.

3.2. Data Preprocessing Tool

We developped a Python script for converting the raw data records to a more suitable motion data file format as TRC (Track Row Column[1]). The TRC format enables easy joining of the raw motion data for the left and right hand and an easy import to the state-of-the-art 3D character animation software. In addition, the TRC recordings of both gloves can be joined to one TRC file by the next auxiliary python script.

3.3. Semiautomatic CyberGlove3 Glove Calibration

We consider the glove calibration during the data post-processing phase and to completely avoid the built-in glove calibration of the DCU. For this purpose, we used a professional 3D character animation software MotionBuilder and created a new calibration tool.

3.3.1. Calibration Tool

The tool has the form of several embedded templates. The main template integrates the graphical user interface similar to the manual calibration dialogue panel of the DCU, see Figure 1. All templates are built from *Relations constraints* elements and connected between themselves to model all needed relationships.

Figure 1: One user interface for manual calibration of two data gloves at a time. In the figure, only the half of the interface for the right hand is shown.

MotionBuilder also supports character skeletal animation. Hereby, we can directly refer from the templates to both the target avatar model and the object represents the motion capture data imported from the given TRC file, see Figure 2. All needed mathematical operators are created by the *Relations constraints* in a simple way, that in general

[1]http://simtk-confluence.stanford.edu:8080/display/OpenSim/Marker+(.trc)+Files

can be edited and used as building blocks for very specific actions.
Nevertheless, we assume linear relationships of the sensor raw data and angular rotations of fingers. We added blocks *Scale and offset (Number)* so that each block implements one scalar linear equation. Thus, it provides a baseline framework for the manual calibration technique similar to the one originally designed (Carmel et al., 2014).
In advance, only three prototype sub-templates: *finger*, *thumb* and *pinky* are needed for the glove calibration. The *finger* prototype sub-template models relationships of the finger bones of the target model and the raw sensor data labeled as *metacarpal*, *proximal* and *abduction*, and optionally *distal flexion* respectively. Thus, there are three (four) blocks for three (four) scalar linear equations depending on the number of used sensors (18 versus 22 glove sensors). For the 18 sensors, we interpolate the distal flexion from the proximal flexion in all prototype sub-templates. Generally, the *finger* sub-template is used for index, middle and ring fingers. The second distinct prototype sub-template has to be considered for thumb due to the different treatment. Nevertheless, we assume also one more prototype sub-template to enable different treatment for the pinky finger.

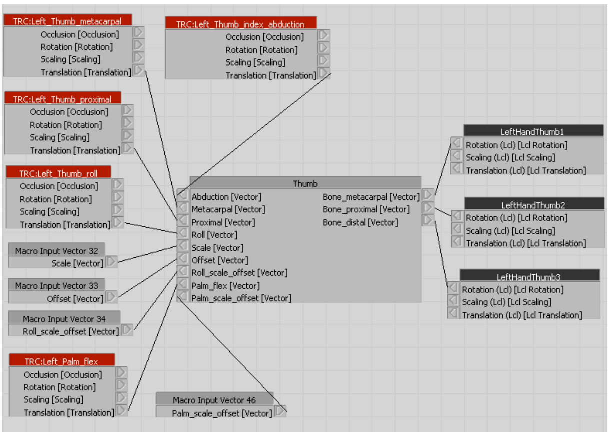


Figure 2: The prototype template for the thumb.

The prototype sub-templates are assigned to all fingers of the left and the right hand to assemble the main calibration template, see Figure 1. Each of the assigned sub-templates connects the relevant imported glove sensor data on its input with the joints of the target model skeleton on its output, see Figure 2. Each sub-template has also a user input form usable for the manual setting of the calibration parameters (the parameters controlling the scalar linear equation).

3.3.2. Target Model

The choice of the target model is done to overcome a limitation of the built-in hand model internally used by the CyberGlove3 that does not provide enough degrees-of-freedom for interpretation of sign language hand shapes. In addition, the new target model has to be appropriate also for the VICON motion capture data and potentially for facial motion capture data.
We used the 3D model for a whole human body automatically generated by a character generator[2]. The target model has all finger bones, facial bones, support for characterization of whole body motion capture data and also allows next data post-processing such as retargeting to different body proportions (sign language speaker/model).
In general, the property of such character models is the standard decomposition of its skeleton bone rotation to three basic operations: bone flexion, abduction and twist. It is done that one rotation axis of the bones in its local coordination system is in the direction of the bone head. We found that this feature is not suitable for mapping of thumb glove motion capture data. Therefore, we modified the local coordinate system of the thumb roll joints (left/right) of the generated model. We added a post-rotation of the local coordinate system so that there exists a direct mapping from the roll sensor and index/thumb abduction sensor each to one rotation axis, see Figure 3. The modification causes the rolling the thumb under the palm just by rotation of the x-axis and simultaneously *clear* side-to-side thumb/index abduction for second y-axis. For the such generated model, the post-rotation of the local coordinate system for the right hand was determined to be (-39°,-13°,27°) and mirrored values for the left hand.

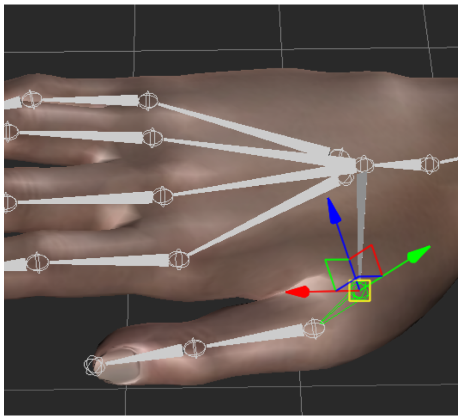


Figure 3: The modification of the local coordinate system for thumb metacarpal bone enabling mapping from thumb roll and abduction sensor.

3.3.3. Palm Flex Integration

In contrast to CyberGlove3 software, we integrate to the target model also a new relationship of the palm flex sensor. The palm flex sensor measures the flexion of the back of the hand. In particular, we consider the sensor to control the palm flexion of the target model by the pinky finger. We found that the motion capture data of the sensor are very important to reach hand shapes incorporating a touch of thumb and pinky of the target model.
Therefore, we extended the above-mentioned pinky prototype sub-template about linear mapping of the palm flex sensor data TRC_{px} to rotation of the pinky hand bone (*os metacarpi V*) as R_{ppx} (2). We added one more *Scale and offset (Number)* block and created one addition linear equation per hand. Now, the role of the pinky hand bone is

[2] https://charactergenerator.autodesk.com/

similar to the thumb metacarpal bone and it causes rolling of the pinky base joint in a direction to the palm.

$$R_{ppx} = a_{ppx}TRC_{px} + b_{ppx}, \tag{2}$$

where a_{ppx} and b_{ppx} are new calibration parameters for the pinky palm flex gain and the pinky palm flex offset.

However, this essential extension does not result in the thumb/pinky touch by the target model. We observed, for such hand shape, there is some kind of dependence of the palm flexion and the pinky metacarpal flexion. We experimentally extend the linear equation for the pinky metacarpal rotation (R_{pmx}) about 75 % of calibrated value of the palm flex sensor as:

$$R_{pmx} = a_{pmx}TRC_{pmx} + b_{pmx} + 0.75R_{ppx}, \tag{3}$$

and the equation for the metacarpal thumb abduction (R_{tma}):

$$R_{tma} = a_{tma}TRC_{tma} + b_{tma} \tag{4}$$

about an additional palm flex mapping through new *Scale and offset (Number)* block. We got replacement of (4) as:

$$R_{tma} = a_{tma}TRC_{tma} + b_{tma} + a_{tpx}TRC_{px} + b_{tpx}, \tag{5}$$

where thumb palm flex gain a_{tpx} and thumb palm flex offset b_{tpx} are next two new parameters.

3.3.4. Automatic Initialization Tool

The goal of the initialization tool is automatic estimation of all parameters of all sensor transformations, i.e. for the linear transformation to find all gains a and offsets b. To be able to find such parameters we need to know at least two different values of each sensor and their true reference values because there are two unknown parameters a and b for each sensor. Two sensor values guarantee the exact solution of each linear transformation, however, it is not possible to cover the full range of movements of all fingers only by two hand shapes. In addition, there are even four unknown parameters for the thumb metacarpal abduction conversion (5), i.e. we need at least four pairs of sensor reading and the reference value to find these parameters.

In the experiment, the calibration take consists of five hand shapes: a flat hand, a stretching of all fingers, a fist and two "o" hand shapes, the one with thumb – index touch and the second with thumb – pinky touch respectively. However, we need to solve an overdetermined system of linear equations. We used a well-known method of least squares to find an approximate solution. This solution found can then be used in the main MotionBuilder template as a starting point for the manual refinement of the calibration parameters. We need to acquire the necessary reference joint values of the target model for the automatic calibration too. This can be done by an animator or as we did by our calibration tool manually set for a one particular calibration take.

4. Conclusion

The time-consuming and laborious calibration of two gloves is moved from a recording session to the phase of an off-line data post-processing when the presence of the signer is not required. During the recording session, all data glove takes are always stored as raw data on glove SD cards. The proposed calibration procedure involves only a calibration sequence during the recording session. We chose five predefined hand shapes that have good visual interpretation and cover the measuring ranges of the sensors. Change or extension of additional calibration hand shapes depends on the task and requirements of the researchers on the accuracy of the target hand shapes.

The time required for the calibration sequence is very short, about 2 minutes, it can be captured simultaneously for both hands. Nevertheless, about 20 minutes is needed to reach the operating temperature of the CyberGlove3 sensors before the recording session. We advise this sequence to capture one more at the end of the recording session. We recommend also to use a webcam to store video of the calibration hand shapes. The video is particularly useful for the phase of the off-line data post-processing which will help resolve potential ambiguities between the calibration hand shapes and the captured hand shapes. All tools are freely available at *http://www.kky.zcu.cz/en/download*.

5. Acknowledgements

This research was supported by the Ministry of Education, Youth and Sports of the Czech Republic, the project No. LO1506.

6. Bibliographical References

Carmel, R., Ullrich, C., Silver, J., Discoe, B., and Sharma, S., (2014). *VirtualHand v3.1.3 - User's Guide*. CYBERGLOVE SYSTEMS LLC.

Elliott, R., Glauert, J. R. W., Kennaway, J. R., Marshall, I., and Safar, E. (2008). Linguistic modelling and language-processing technologies for avatar-based sign language presentation. *Univers. Access Inf. Soc.*, 6(4):375–391, January.

Griffin, W., Findley, R., Turner, M., and Cutkosky, M. (2000). Calibration and mapping of a human hand for dexterous telemanipulation. In *ASME IMECE 2000 Symposium on Hoptic Interfaces for Virtual Environments and Teleoperator Systems*, pages 1–8.

Huenerfauth, M. and Lu, P. (2010). Accurate and accessible motion-capture glove calibration for sign language data collection. *ACM Trans. Access. Comput.*, 3(1):2:1–2:32, September.

Kahlesz, F., Zachmann, G., and Klein, R. (2004). Visual-fidelity dataglove calibration. In *Computer Graphics International 2004 (CGI 2004)*, pages 403–410. IEEE Computer Society, June.

Steffen, J. F., Maycock, J., and Ritter, H. (2011). Robust dataglove mapping for recording human hand postures. In *International Conference on Intelligent Robotics and Applications (ICIRA)*, volume 7102. Springer.

Wang, Y. and Neff, M. (2013). Data-driven glove calibration for hand motion capture. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '13, pages 15–24, New York, NY, USA. ACM.

"Non-tokens": When Tokens Should not Count as Evidence of Sign Use

Gabriele Langer, Thomas Hanke, Reiner Konrad, Susanne König
Institute of German Sign Language (IDGS), University of Hamburg
E-mail: {gabriele.langer, thomas.hanke, reiner.konrad, susanne.koenig}@sign-lang.uni-hamburg.de

Abstract

Lemmatised corpora consist of tokens as instantiations of signs (types). Tokens usually count as evidences of the signs' use. Frequency of tokens is an important criterion for the lexical status of a sign. In combination with metadata on the signers' sociolinguistic backgrounds such as age, gender, and origin these tokens can also be analysed for regional and sociolinguistic variation. However, corpora may also contain instances of sign use that do not reflect the sign use of the person uttering them. This is particularly true for metalinguistic discussions of signs, malformed signing and slips of the hand as well as other phenomena such as copying/repeating signs of the interlocutors or from stimulus material. In our presentation we list and discuss different kinds of sign use (tokens) that should either not be counted as proof of a sign type at all or at least not as evidence of regular sign use by that particular person. Examples of these "non-tokens" are either taken from the DGS Corpus or from uploaded video answers of the *DGS Feedback*. We also discuss some implications on how to annotate these cases.

Keywords: sign language corpus, metalinguistic signing, lemmatisation, empirical status of tokens, annotation

1. Introduction

In a sign language (SL) corpus continuous stretches of signing are made accessible permanently (recorded on film) and lemmatised as running text from end to end without gaps. By this signs, linguistic phenomena and patterns can be analysed in context. Each individual manual activity that can be interpreted as a sign is tagged and has to be dealt with in annotation (lemmatisation). Occurrences of signs in signed utterances are usually taken as evidence for a sign's existence and use. The advantage of having a large corpus is that ideally one can have many tokens from different signers as evidence for one particular sign. Metadata of the signers (e.g. their regional rooting, SL acquisition age, age, gender etc.) can be used to analyse the distribution of a sign and its use or of other linguistic phenomena and patterns across different regions and groups of signers (e.g. see Langer et al. 2014).

However, some signs (tokens) in a corpus are not evidence of regular sign use by that particular person, for example when citing others in a metalinguistic comment on signs, malformed sign use or copying signs from the interlocutor or from stimulus material. Other (manual) activities appearing to be signs are not signs at all or signs from other sign languages. Depending on their number, it seems reasonable and useful to classify and tag these kinds of tokens in order to be able to include or exclude them from analyses according to the particular research question.

1.1 Background

The DGS Corpus Project is a long-term project with two major goals: building a lemmatised and annotated reference corpus of German Sign Language (DGS) and compiling a general DGS Dictionary on the basis of this corpus data. A major focus is on how to annotate a corpus in a way that best serves as a multi-purpose language resource for all kinds of research questions. At the same time there is also a strong lexicographic interest focusing on individual signs, their forms, variants, use, and distribution.

1.2 Sources of Examples

Examples for this paper are drawn from two quite different sources of data: annotated corpus material including free conversation as well as highly metalinguistic and context-dependent signed comments on stimuli presented in an online survey (called *DGS Feedback*) and recorded with a webcam.

1.2.1 DGS Corpus Material

The DGS Corpus contains over 1150 hours of footage and about 615 hours of filmed signing. A considerable portion of this data is being lemmatised, annotated and made accessible to become a reference corpus for German Sign Language (DGS)[1]. During the data collection phase (2010-2012), 330 signers where filmed in 12 different locations across Germany. The filming session for each pair of informants (signers) took place at one day and lasted about 6-7 hours (including breaks). In order to encourage the signers to keep interacting and talking to each other, a third signer (a trained deaf moderator) led through the sessions and through up to 20 different tasks. All tasks with exception of elicitation of isolated signs (see 2.1.1) were designed to record fluent natural (or near-natural) signing in context. Some tasks were more pre-structured and staged than others, several tasks involved talking about deaf-related topics or telling about one's life and personal experiences. One task is a free conversation in absence of the moderator on anything the informants liked to sign about. (For more detailed information on data collection and technical aspects cf. Nishio et al. 2010, Hanke et al. 2010).

Some of the filmed conversations contain tokens that have to be dealt with in annotation but should not be regarded as normal sign use of this person. Some of these "non-tokens" are metalinguistic sign use and appear either spontaneously or in the conversation task *Young*

[1] A representative part of the data is published from 2015 on as a subcorpus (DGS Corpus Project, 2015-2016).

vs. old signs which was specifically designed to provoke metalinguistic utterances.

1.2.2 Answers to DGS Feedback

In the project, one intermediary step towards a corpus-based dictionary is the compilation of a preliminary basic dictionary (for more detail cf. Langer et al. 2014) In the process of compiling the basic vocabulary lemma sign candidates (including their variant forms and presumed meanings/senses) are presented in an online survey called *DGS Feedback* to be commented upon by signers. The answer options include the possibility to record a signed answer or comment via a webcam and upload it to our server. Why is this material interesting? These signed answers and comments are highly meta-linguistic in content and very context-dependent in that they directly refer to the stimuli presented on the web page – often by a form of citation of or reference to the stimulus sign. On the one hand, the answers contain valuable information on sign use, and on the other hand, they are densely packed with all kinds of tokens that are some sort of "non-tokens". In order to make use of this information the uploaded signed answers have to be made accessible by some kind of annotation.[2] Whether they are selectively lemmatised and annotated or continuously as running text – in both cases one has to decide which occurrences of signs qualify to count as evidence of a sign and which do not ("non-tokens"). Therefore, this material is a valuable source, a testing ground, and ideal example for the "non-token" issue discussed here.

When using webcams for recording, the signing may be adapted to the limited field of view (and speed) of the webcam and provoke other kinds of special, non-standard tokens of signs (for this aspect see 4.2.4).

2. Influence of Elicitation Task

A good corpus should cover a variety of text types showing different uses of signs including metalinguistic ones. It is common practice in SL corpus building to have data collection sessions with more or less staged communicative events[3] to record signing. For this aim different tasks can be designed that may influence the signs' use.

2.1 Narratives and Conversation

Signed narratives, conversations or discussions can contain incidences of "non-tokens" that in corpus annotation might be useful to tag separately. In normal conversation talking about signs is just one of the many topics that can come up. Depending on introductory questions or given topics – e.g. acquisition of sign language, communication in school, family or at work, experiences with hearing persons, or communication while travelling in foreign countries – the documented signed texts are more or less likely to contain metalinguistic use. In the DGS Corpus we found examples for this kind of singing in several tasks including free conversation, young vs. old signs, and elicitation of isolated signs.

2.2 Elicitation of Isolated Signs

Eliciting individual signs by asking informants what sign they use for presented concepts is a elicitation method that has been widely used in SL research, specifically in studies on lexical variation and for compiling dictionaries. It is discussed here separately because it prompts metalinguistic responses from interviewees that often include one particular kind of tokens showing signs that the informant knows of but normally does not use in a natural signing environment, except when talking about these signs (see 3.1, category (b)). Stimuli for eliciting isolated signs are e.g. a picture, a fingerspelled word, a written word, sometimes with a short contextual hint, or a combination thereof. Ideally, one stimulus should evoke a single concept to be expressed by a (lexical) sign. The spontaneously given response to this kind of stimulus normally is an isolated sign that is used by the interviewee to express the intended concept. However, answering to prompts for isolated signs is a very unnatural communication situation in which the interviewees are highly aware of its sole purpose of collecting signs as signed representations of concepts or signed equivalents of written words. In their responses, signers sometimes show not only their own sign but also other lexical variants they know of. These (lexical or phonological) variants may or may not be used by the interviewees themselves. Sometimes informants provide explicit or implicit contextual clues to indicate whether they use these additional variants themselves or not. Often they just list variants with no indication whether they use the signs actively or not. One solution to this problem is to only consider the first – presumedly most spontaneously given – answer to a stimulus and leave out the following signs in the analysis because their use or non-use is not explicitly made clear.[4]

Although untypical for corpus data (usually aiming at documenting connected natural signing) this elicitation method has been used as one of several tasks in some SL corpus data collections[5] and has also been used in the DGS Corpus.[6]

3. Kinds of "Non-Tokens"

The issue of this paper has been called "non-tokens" but actually the signs discussed are tokens, but are special with regard to their empirical status – whether they

[2] Furthermore, such a collection of answers and comments could be regarded as a corpus of a very specific (metalinguistic) form of signing that might be worth of lemmatisation and analysis in its own right. For example, it seems to be great material to examine how metalinguistic citations and references are marked (e.g. body body shifts, facial expressions). If one were to lemmatise and annotate material of this kind the issue of how to handle different kinds of "non-tokens" would be a very prominent one.

[3] On the concept of staged communicative events cf. Himmelmann 1998, 185-186).

[4] Cf. for example Stamp (2014, 5): "Many participants produced multiple examples of signs and, as a result, either the variant stated to be the sign, or if not stated, the first variant produced, was coded."

[5] ASL Corpus: Lucas et al. (2001, 40), BSL Corpus: Schembri et al. (2013, 140).

[6] The elicitation task is only a small portion of the data: With 22 hours it accounts for 3.5 per cent of the filmed signing.

should count as proof of the sign or the sign's use or not in the light of a given research question. Whether a token is a "non-token" or not is not an on-off decision but a matter of degree and perspective.

A metalinguistic token of the sign X as in "I don't use X" might be taken as evidence that the sign X exists (token status) and that the signer knows this sign, it might also be counted in a general token count for frequency of the sign, but it may be a "non-token" of sign use by this particular person as a representative of e.g. a certain region when doing a distributional analysis of where a certain sign is used.

Malformed tokens of signs produced as slips of the hand might be "non-tokens" in an investigation of normal sign form, but might be the only tokens of interest in a study on slips of the hands.

These examples show that it would be very useful to identify, classify and label tokens with various "non-token" potential in a corpus as part of the such enriched annotation. This enables researchers to include or exclude tokens of these special kinds and thus to be more precise in the use of the corpus and avoid certain groups of noise in the data they choose for their analyses. In the following sections we will discuss different types of potential "non tokens".

3.1 Metalinguistic Reference

Any kind of metalinguistic reference to individual signs does not necessarily reflect typical sign use of the signer and should be identified and labelled.

3.1.1 Metalinguistic Reference to an Existing Sign

The first type of our catalogue of "non-tokens" is a metalinguistic reference to an existing DGS sign. The reference consists of an execution of the sign that is being referred to. Such a token does not necessarily indicate that the signer would use this sign in non-metalinguistic signing. Metalinguistic references could be either:

(a) a reference to an existing DGS sign also used by the signer, or
(b) a reference to an existing DGS sign that is normally not used by the signer.

In many cases the context will clarify which one of the above cases apply. (Think for example about the following utterances: (a) "I always sign X" vs. (b) "In Bavaria they use the sign X" (signed by a non-Bavarian signer).

Tokens of the type (a) are pieces of conscious introspective information rather than an unconsciously and spontaneously produced evidence of sign use by a particular person. These tokens still could count as evidence of sign use by that person but nevertheless they should be labelled as metalinguistic reference.[7]

Tokens of the type (b) could be interpreted as a conscious introspective piece of information on the existence and use of the sign itself but it would be misleading to include tokens of this kind for example in a distributional analyses of tokens linked with the signers' individual metadata (e.g. regional rootedness for regional distribution).

In cases where it cannot be decided whether (a) or (b) applies, one should stay on the safe side and assume (b) or open up a third category c) for these unclear cases.

Consider example 1. In the free conversation task with the moderator not being present, two female informants (A and B) discuss the use of the DGS sign WOMAN1. The sign's form is iconically motivated: The B-handshape traces the breast of a woman. Some signers, especially women, dislike the sign because they feel that it is not politically correct. First B asks A for her sign for 'woman' by fingerspelling the German word. A's answer is WOMAN3 – one example for the above-mentioned category (a). Then B starts talking about the debate showing the sign WOMAN1. A knowing about the debate anticipates the sign so that both nearly synchronically make the sign WOMAN1.

Informant A	**1**	**2**	**3**	**4**
Gloss	YES1A	WOMAN1	YES1B	WOMAN 1
Mouthing		frau		frau
Translation	Yes, the sign WOMAN1.			
Comment		reference (b)		reference (b)

Example 1: Metalinguistic Reference (b)

Token 2 and 4 are examples of category (b), but a little later in the conversation there are also examples of category (c) when A repeats two other signs for 'woman' that B has shown before A's utterance (see example 2).

1	**2**	**3**	**4**	**5**	**6**
THERE-IS1A	AND2	ALSO1A	WOMAN4	WOMAN2	WOMAN4
es-gibt	auch		frau		
There are also the signs WOMAN4 and WOMAN2.					
			reference (c)	reference (c)	reference (c)

Example 2: Metalinguistic Reference (c)

Metalinguistic references of the type (b) also cover cases when signers talk about what signs other groups or individual signers use or have been using (for example old, young, hearing, or deaf signers, signers from certain regions, CODAs, interpreters, father, grandmother etc.). In example 3, when discussing old versus young signs, informant A contrasts the sign of her grandmother with her own sign for 'banana'.

1	**2**	**3**	**4**
INDEX1	BANANA1A	I1	BANANA2
	banane		banane
She signs BANANA1A, I sign BANANA2.			
	reference (b)		reference (a)

Example 3: Metalinguistic Reference (a, b)

This kind of reference where the signer recalls a sign from memory can be distinguished from a reference (d) where the signer refers to a sign from his/her interlocutor as shown below in example 9, sign 2.[8]

[7] Not always does introspective information of one's own language use correspond with the unconscious language use in real life.

[8] Of course, signers like speakers adapt unconsciously their language use to their interlocutor. Often this remains unnoticed and wouldn't be labeled.

When signers metalinguistically refer to signs of a foreign sign language this is also a case of metalinguistic reference (e).[9]

1	2	3	4	5
ENGLAND3	INDEX1	YES1B	$GEST-HEAD-SHAKING	YES-CORRECT
englisch				yes
In England they don't sign YES1B, but YES-CORRECT.				
		token		reference (e)

Example 4: Metalinguistic Reference (a, b)

In this example, it clearly is not a case of regular use of a borrowed sign of foreign origin but a metalinguistic reference to the foreign sign normally not used by the signer. It would be useful to be able to distinguish metalinguistic reference to a foreign sign from regular use of a borrowed sign.
A special case of metalinguistic reference (f) is when signers talk about signs that they have been using in the past, but do not use anymore at the present time.

3.2 Un-Tokens (often contrastive)

A very common communicative pattern or strategy in DGS is to name the opposite or complementary fact or thing with negation – either by headshake and or facial expression alone or in addition by an explicit negation sign – followed by the fact or thing that one wants to express positively. This pattern works with contrasts between the negated and the positively expressed parts of the message. (The following made-up examples illustrate this pattern: CHEESE-neg NOT – SAUSAGE 'not cheese but sausage' or FREE-neg NOT – WORK 'not having spare-time but having to work'.) This strategy can also be used in metalinguistic signing about a non-sign vs. sign or non-use of a sign vs. use of a sign in a specific sense resulting in what we call *contrastive un-tokens* in the negated part.
In the *DGS Feedback* answers we find many examples for these kinds of un-tokens. The negated token can concern the sign as a whole ("non-sign"), a particular wrong execution of the sign ("non-form") or a particular use of the sign with regard to meaning ("non-use").
In example 5, taken form the *DGS Feedback* answers, the signer refers to a presented signed question item with his own sign for cheese (1, regular token), then copies the stimulus sign (3, un-token) marking it with a headshake and the following sign NOT1 in the sense of 'not used' and then repeats his own sign (6) as information of what sign he uses.

1	2	3	4	5	6
CHEESE2B	I1	CHEESE2A	NOT1	I1	CHEESE2B
		headshake			
käse					käse
My sign for 'cheese' is CHEESE2B, not CHEESE2A.					
regular token / reference to item		un-token / copy of stimulus			reference (a)

Example 5: Un-Token

[9] This has to be differentiated from cases where signers use signs of foreign origin non-metalinguistically in their normal signing – those cases could be either indications of borrowing or instances of code mixing.

Not all un-tokens are contrastive tokens, in principle tokens could also be negated without being followed by their positive counterpart. The point here is that the token in question is deliberately and clearly negated or declared as wrong or not being used to indicate that this in not the correct sign, sign form, or sign use.
In the following example 6 informants A and B are talking about their experiences as being deaf, in particular about their relation to and communication with their (hearing) parents. A cites a malformed sign her mother uses for 'important' – with U- instead of V-handshape – and her attempt to teach her mother the correct form IMPORTANT1:

1	2	3	…
MOTHER 2	SIGN1	IMPORTANT1*	
mutter		wichtig	
My mother signs IMPORTANT1*.			
		un-token (malformed)	

4	5	6	7	8	9
SIGN1	NORMAL1	SIGN1	IMPOR-TANT1	IN-DEX1	IMPOR-TANT1*
	normal		wichtig		wichtig
Normally one signs IMPORTANT1, but she signs IMPORTANT1*.					
			reference (a)		un-token (malformed)

Example 6: Un-Token

Example 7 is a *DGS-Feedback* comment to the stimulus WOMAN1/frau with the meaning of 'woman, female'. The stimulus sign (1) is copied as reference to the stimulus and then marked as un-token by the following comment ("I don't use this."). This is an example of an un-token with regard to meaning because the signer uses this sign, but in combination with another mouthing only to express the meaning 'breast' (example 7, sign 6).

1	2	3	4	5	6	7
WOMAN1	SIGN1A	I1	USE1	NOT1	WOMAN1	YES1A
frau		ich	benutze	nicht	busen	ja
I don't use the sign WOMAN1 for 'woman', but for 'breast'.						
copy of stimulus / un-token					metalinguistic reference (a)	

Example 7: Un-Token and Metalinguistic Reference

3.3 Copy of Stimulus (Reference)

When signed stimuli are used for prompting signers to comment about these signs – as it is done in the *DGS Feedback* – signers in their metalinguistic comments or answers often refer to these prompts either by copying the stimulus sign – here called *copy of stimulus* – or by using their own sign (see example 5, sign 1) for the same concept. Tokens that copy the presented stimulus sign for the sole purpose of referring to them are tokens that cannot count as evidence for a sign's existence or use, because they are pre-specified by the context.[10]

[10] It does not matter here whether the stimulus is a real sign or a non-existing, made-up sign form or whether an existing stimulus sign is presented correctly (e.g. with a correct meaning) or incorrectly (e.g. with an incorrect, unusual or unknown meaning). In the *DGS Feedback* we present non-existent made-up signs as distractors along with existing signs.

Often the copy of a stimulus is not only used to refer to the presented item but is also commented upon which either makes the copy a un-token sign 3 in example 5 and sign 1 in example 7 are not only un-tokens but also at the same time copies of the stimulus.

4. Non-Signs of the SL

Some manual activities within the continuous sign stream look like signs and are part of the utterances but in fact are not part of the respective target sign language.[11] However, such tokens may be part of the utterances, are needed for a correct interpretation and cannot be left out without loosing content of the utterance. This means that they cannot be skipped for annotation.[12]

For a corpus on DGS, all tokens that are not DGS sign use should be tagged. We propose the following types of non-signs or non-DGS signing respectively.

4.1 Gestures of Hearing Persons

Sometimes signers talk about gestures used by hearing persons either in communication attempts with deaf or among themselves. In the following example informant A tells informant B about her colleague who had learned the gesture for 'tasty' during his trip to Italy:

1	2	3	4	5
TASTY3	ITALY1	BELONG-TO1	SAY1	$GEST-TASTY
lecker	italien			lecker
In Italy this gesture means 'tasty'.				
regular token				cited gesture

Example 8: Copy of Gesture

In the course of the conversation, both signers use the gesture and also use their own signs for 'tasty'. Informant A marks the difference between her own sign TASTY3 and B's sign TASTY1:

1	2	3	4	5
INDEX1	TASTY1	ICH1	TASTY*	TASTY3
	lecker		lecker	lecker
You sign TASTY1, I sign TASTY*, TASTY3.				
	reference (d)		slip of the hand	reference (a)

Example 9: Metalinguistic Reference (a, b)

Interestingly, just before showing her own sign TASTY3 (5), A produces a mixed form (4) with the location of TASTY3 (belly) but the circling movement of TASTY1 (instead of repeated tapping). It is a malformed metalinguistic reference and interpreted as a slip of the hand (see below 4.2.2) that is immediately followed by a self-correction (5).

4.2 Malformed or Deviated Sign Forms

The next group consists of tokens that are in some way malformed or have deviated sign forms. They appear in normal (i.e. non-metalinguistic) signing contexts. Sometimes these tokens slip by unnoticed by the signer (cf. sign 5 in example 9), sometimes they are noticed and self-corrected.

4.2.1. False Starts and Aborted Tokens

Some tokens are instances where the signer starts to produce the sign and the hands move to the beginning of the sign but do not finish executing the sign completely. Sometimes one can guess at the intended sign, sometimes not.

Unfinished signs can be false starts or aborted signs. Often false starts are indicated by a facial expression or a subsequent headshake when the signer becomes aware of the near-mistake and produces a correction. Annotation guidelines have to specify how to treat false starts and aborted tokens and whether to tag them at all. If they are tagged, they should be labelled in a specific way in order to search for them separately or to exclude them from token counts and analyses.

4.2.2. Slips of the Hand

Sometimes signers accidently use the wrong sign or execute a sign with a wrong parameter. These instances are known as slips of the hand. They can pass by unnoticed or they can be self-corrected by the signer. Sign 4 in Example 9 is an example of slip of the hand followed by the correct sign.

While slips of the hand are an interesting topic of investigation in their own right (e.g. cf. Leuninger et al. 2004), for most other analyses one would rather exclude them from all counts and therefore need to label them accordingly.

4.2.3. Trying out or Searching for Signs

Sometimes in conversation but even more so in elicitation tasks for single signs, a signer is searching for the right sign or sign form he/she wants to use – visibly thinking "out loud" and in the process trying out several different signs or slightly different sign forms before arriving at the searched-for sign. These forms should not be regarded as tokens in their own right and be labelled accordingly.

In the conversation task *Young vs. old signs* informant A and B (from example 3) try to remember an old sign for 'parents'. B first shows a form like scratching her cheek with the fingertips of her 4-bent-handshape. Immediately, B reacts in raising her arm and wiggling with her index and middle finger. While B is in doubt and shows her own sign for 'parents', A continues to try out, moving both fingers to her cheek, then making small movements downwards while bending the fingers slightly as B did with four fingers. B copies the sign from A and shows a negative facial expression while A confirms that she at least knows this sign.

4.2.4. Adapted Tokens (Limited View of Webcam)

In cases where the camera field of view is limited to a small window – for example when recording with a webcam – signers sometimes subconsciously or consciously adapt their signing to fit into the window. They may for example change the place of articulation. In one feedback answer the signer showed his sign for 'hunger' twice: The first token of HUNGER was executed at the stomach – the normal place of articulation for this sign –

[11] As opposed to other manual activities that are not part of the linguistic utterance such as scratching one's nose, pushing one's hair back, rubbing or tapping nervously one's fingers or shaking the hand to loosen tense muscles. These manual activities are generally not annotated.

[12] As one would probably do in a selective spot transcription coding only tokens of interest, leaving out all others.

and therefore almost completely outside the view of the camera, the second token of HUNGER was executed at the chest – wrong place of articulation – to show handshape and the movement within the view of the camera. The second token of HUNGER is an example for an adapted token. Another example was the answer to the stimulus 'cheese' already discussed in example 5. The answer continued after a short break with another token of the sign CHEESE2B but this time adapted in form (both orientations changed in relations to each other) to show the handshapes clearly in the view of the camera and make sure that the handshape is visible from the front. These adapted sign forms should not be treated as normal instances of the sign with regard to form but should be labelled as adapted or annotated as deviant forms when lemmatising.

5. Consequences for Annotation

The various examples in this paper show that some tokens in running signed texts should not qualify as evidence for a sign's existence or for the use of a particular sign by the particular signer. It would be wise to label cases that are coming to one's attention accordingly. This allows for conscious decisions on whether to include or exclude tokens of certain kinds from analyses. One could argue that the percentage of non-tokens is not large enough to worry about them but our impression is that the percentage might be much larger than originally thought. SL corpora are much smaller than written text corpora and rely on only a relatively small number of individual signers. Therefore, even a single discussion like that described in example 1 and 2 can influence or distort results of analyses when not labelled properly and "non-tokens" not being excluded from certain kinds of analyses.

For the time being, we annotate all cases of "non-tokens" with metadata to the token tags, using an open vocabulary corresponding to the categories presented in sections 3 and 4. The vocabulary also contains some grouped values in order not to complicate the annotation too much for the first-pass annotators in cases where a categorisation is not straightforward. This approach is a preliminary version only. While it allows us to keep track of the cases already identified and to do some analysis, it does not yet allow iLex to automatically integrate the "non-token" flags into account when computing token counts. For this, a more sophisticated solution needs to be implemented once we have enough evidence that the current categorisation is both stable and manageable for the annotators.

6. Acknowledgements

This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.

7. Bibliographical References

DGS Corpus Project (2015-2016): Subcorpus. Available at: http://meine-dgs.de.

Hanke, T., Storz, J., Wagner, S (2010). iLex: Handling Multi-Camera Recordings. In *Workshop Proceedings. 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies Language Resources and Evaluation Conference (LREC) Valletta, Malta, May 2010,* pp. 110--111. [URL: http://www.lrec-conf.org/proceedings/lrec2010/workshops/W13.pdf].

Himmelmann, N. (1998). Documentary and descriptive linguistics. *Linguistics,* 36, pp. 161--195.

Langer, G., König, S., Matthes, S. (2014). Compiling a Basic Vocabulary for German Sign Language (DGS) – lexicographic issues with a focus on word senses. In A. Abel, Ch. Vettori and N. Ralli (Eds.), *Proceedings of the XVI EURALEX International Congress: The User in Focus, July 15-19 2014 in Bolzano/Bozen – Italy*, pp. 767--786. [URL: http://euralex2014.eurac.edu/en/callforpapers/Documents/EURALEX_Part_2.pdf].

Langer, G., König, S., Matthes, S., Groß, N., Hanke, T. (2016). Variation of DGS lexical items. What sign language lexicography can gain from a mixed method approach: Corpus data supplemented by crowd sourcing. Poster presented at the *Theoretical Issues in Sign Language Research (TISLR) 12 Conference, Jan 4-7, 2016 in Melbourne, Australia*. [URL http://www.tislr12.org/posters/; no. 117].

Leuninger, H., Hohenberger, A., Waleschkowski, E., Menges, E., Happ, D. (2004). The impact of modality on language production: Evidence from slips of the tongue and hand. In T. Pechmann, C. Habel (eds.). *Multidisciplinary approaches to language production*. Berlin: de Gruyter, pp. 219--278.

Lucas, C., Bayley, R., Clayton, V. (2001). *Sociolinguistic Variation in American Sign Language.* Washington, DC: Gallaudet University Press.

Nishio, R., Hong, S.-E., König, S., Konrad, R., Langer, G., Hanke, T., Rathmann, C. (2010). Elicitation Methods in the DGS (German Sign Language) Corpus Project. In *Workshop Proceedings. 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies Language Resources and Evaluation Conference (LREC) Valletta, Malta, May 2010,* pp. 178--185. [http://www.lrec-conf.org/proceedings/lrec2010/workshops/W13.pdf].

Schembri A., Fenlon J., Rentelis R., Reynolds S., Cormier K. (2013). Building the British Sign Language Corpus. *Language Documentation and Conservation* 7, pp. 136--154.

Stamp, R., Schembri, A., Fenlon, J., Rentelis, R., Woll, B., Cormier, K. (2014). Lexical Variation and Change in British Sign Language. *PLoS ONE* 9(4): 094053, pp. 1--14. [doi: 10.1371/jpournal.pone.0094053].

Designing a Lexical Database for a Combined Use of Corpus Annotation and Dictionary Editing

Gabriele Langer[1], Thomas Troelsgård[2], Jette Kristoffersen[2], Reiner Konrad[1], Thomas Hanke[1], Susanne König[1]

[1]Institute of German Sign Language (IDGS), University of Hamburg, [2]Center for Sign Language, UCC
E-mail: {gabriele.langer, reiner.konrad, thomas.hanke, susanne.koenig}@sign-lang.uni-hamburg.de, {ttro, jehk}@ucc.dk

Abstract

In a combined corpus-dictionary project, you would need one lexical database that could serve as a shared "backbone" for both corpus annotation and dictionary editing, but it is not that easy to define a database structure that applies satisfactorily to both these purposes. In this paper, we will exemplify the problem and present ideas on how to model structures in a lexical database that facilitate corpus annotation as well as dictionary editing. The paper is a joint work between the DGS Corpus Project and the DTS Dictionary Project. The two projects come from opposite sides of the spectrum (one adjusting a lexical database grown from dictionary making for corpus annotating, one building a lexical database in parallel with corpus annotation and editing a corpus-based dictionary), and we will consider requirements and feasible structures for a database that can serve both corpus and dictionary.

Keywords: corpus building, sign language corpora, lexicography, sign language lexicography, annotation tools, lemmatisation, sense discrimination, dictionary writing system

1. Introduction and Backgrounds

The German Sign Language (DGS) Corpus Project and the Danish Sign Language (DTS) Dictionary Project both have the aim to work on corpus-based lexicography within the iLex environment and want to have dictionary and corpus information in the same database.

Some signs are highly polysemous and have many phonological variants, while others have lots of variants and only a few senses, and yet others have only one form, but lots of senses. This is no problem in a database that serves exclusively as a type inventory for corpus annotation dealing with language documentation and description, yet it causes trouble for the lexicographer, who often needs to work with both flexible and in some cases pragmatic principles in order to produce dictionary entries that are human-readable as well as fairly homogeneous in appearance. In this paper, we will exemplify the problem and present ideas on how to model structures in a lexical database that facilitate corpus annotation as well as dictionary editing.

1.1 DTS Dictionary

The DTS Dictionary (Center for Tegnsprog 2008-2016; Kristoffersen & Troelsgård 2012) is a general-purpose dictionary describing the basic sign vocabulary of DTS. The dictionary has search facilities allowing for lookups based on sign form, Danish equivalent or topic (or a combination of these).

The core of the dictionary-making process is a semantic analysis of each selected sign – a task that so far has been performed partly based on introspection by staff members who are native signers, partly based on evidence found in video recordings. As the DTS group is now starting a corpus project, the aim is to build a tool that will on one hand facilitate the editing of new sign entries, and on the other hand supply tools for "retro-corpus-basing" existing entries, e.g. by checking for missing word-senses, retrieving collocation information, or finding better usage examples. Finally, a corpus will be an essential tool in connection with future lemma selection, and could be used for other linguistic research outside the lexicographic context.

The aim is, as an obvious starting point, to re-use the sign lemmas, which are already uniquely glossed, as the core type vocabulary for the token-type matching during annotation of corpus texts, adding new signs along the road. Furthermore, the aim is to also exploit the word-senses defined in the dictionary entries, so that a token – if a suitable sense is at hand – can be matched directly to a sign type with a specific meaning.

1.2 DGS Corpus

The DGS Corpus Project is a long-term project with two major aims: building a reference corpus for DGS[1] as a multi-purpose resource for research on DGS and compiling a general dictionary of DGS on basis of the corpus data collected. Coming from a background of compiling German – DGS language for specific purposes dictionaries (1993-2010)[2] the annotation tool and integrated lexical database iLex has been developed in the context of these previous projects to facilitate the lemmatisation and annotation of recorded signed data (Hanke 2002; Hanke & Storz 2008). This database containing type entries and lemmatised sign data from previous projects has been carried over and is being used and further developed alongside with the iLex program to suit the needs of the DGS Corpus Project.

In the first stages of the project the focus has been on data collection and annotation, the latter will continue for several years to come to provide the data for general re-

[1] A representative part of the data is published from 2015 on as a subcorpus (DGS Corpus Project, 2015-2016).

[2] For more information on the LSP dictionaries cf. Konrad (2011) and Konrad & Langer (2009).

search and lexicographic analysis. Thus, though the Corpus Project aims at compiling a dictionary as well as building a corpus, up to now the main focus on the development and use of iLex has been on annotation rather than lexicographic description. In the near future iLex structures have to be developed further to better support analysis, the various stages of working out lexicographic descriptions of sign uses as well as the writing of dictionary entries.

2. iLex

The iLex program is a database and annotation environment developed at the Institute of German Sign Language (IDGS, University of Hamburg,) especially for annotating sign language data. Within the annotation environment of iLex video files can be viewed and tagged as in other annotation tools like e.g. ELAN (cf. Crasborn & Sloetjes 2008), where media and annotations are time-aligned. Unlike ELAN, iLex combines a lexical database with transcript views for annotation of video segments. Lemmatisation as a process of identifying tokens as instantiations of sign types (token-type matching) is done by establishing a direct and dynamic link between type and token via drag & drop. Thus, consistency is supported by the database structure and does not rely on the use of ID-glosses (see Johnston 2010). [3] Type entries include information on the presumed lexical types and allow direct access to all tokens of the respective type. Furthermore, iLex provides a type hierarchy with several levels that allows modelling relevant differences in iconicity, form, and use of a sign and tagging the tokens accordingly. Each type of a lower level is attached to exactly one type of a higher level and is considered to represent a subset of the tokens and uses belonging to the superordinate type.

2.1 Use of iLex Structures by DGS Group

In the DGS Corpus Project, four type levels are being utilised. A type at level 3 – hereafter called *supertype* – represents a sign as an abstract linguistic entity (with focus on form and – if existent – iconicity). Types at level 1 – hereafter called *subtypes* – are defined to distinguish different established or conventional uses of a sign with regard to meaning. Conventional uses of a sign typically consist of regular and therefore expectable sign-mouthing combinations. A subtype is directly attached to its supertype if they share the same citation form.
Tokens that show productive or novel uses of a sign or not yet identified conventional uses are matched to the supertype directly. Productive uses are for example occasional or ad-hoc sign-mouthing combinations (cf. König et al. 2008, 398-400).
With the implementation of qualifiers (Konrad et al. 2012), also word forms and other form differences within one supertype or subtype can be classified and labelled. For this purpose the type levels 2 (qualified supertypes) and 0 (qualified subtypes) were introduced. Qualified types allow distinguishing and coding modified forms and also “minor variants” (Johnston 2016, 19-20) of the sign form as part of a more detailed analysis of a sign’s use.[4] The goal of this coding is to determine the range of form variation and modification within the given supertype or subtype.[5] On the subtype level, qualified forms (level 0) are either candidates for word forms or phonological variants or they may be just performance phenomena. On the supertype level form variation (modelled by level 2) can be cases of modification, phonological variation, derivation or performance phenomena.

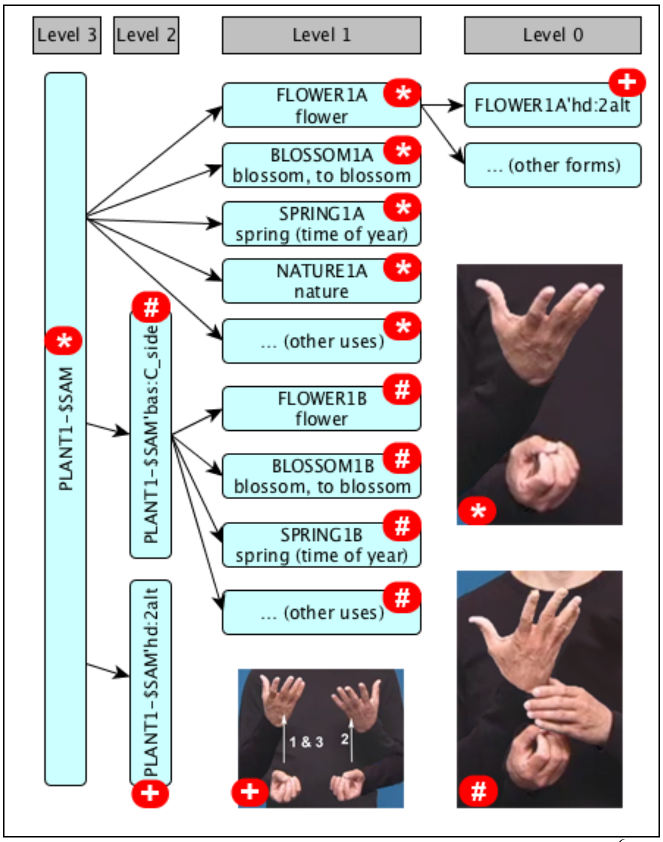


Figure 1: Type hierarchy of DGS sign PLANT1-$SAM[6]

Supertype entries are considered to be lexical entities whereas subtypes group together conventional sign uses, often triggered by mouthing. Roughly speaking, the supertype and subtype structure is used to model polysemy. For lexicographic descriptions the conventional sign uses have to be further analysed for different senses

[3] The iLex database uses type IDs for identification and linking. However, for the ease of human annotators each type is assigned a unique gloss in the database that functions like an ID-gloss and as a mnemonic aid e.g. when reading transcripts or referring to signs in communication. iLex blocks attempts to name a new type entry with a gloss already used and thus makes sure that glosses are unique on each level (cf. 2.1).

[4] This coding is not part of the basic annotation. It can be done completely for all tokens of a sign or selectively at a later annotation pass (lemma revision or detailed transcription).

[5] The focus here is on the individual types because of the lexicographic perspective. However, coding the same modifications or form deviations across different types in the same way will also allow to run analyses across a number of types.

[6] In order to distinguish supertypes from subtypes, glosses of supertypes always have the suffix “-$SAM” (abbreviation of ‘Sammelglosse’ (collective gloss)).

of a sign (sense discrimination).
Figure 1 exemplifies the type hierarchy structure in iLex as it is used by the DGS group. In the lexical database everything that presumably belongs to one sign is in some way hierarchically attached to the same supertype. Depending on their form and contextual meaning, tokens can be attached to types on each level. Tokens attached to types on lower levels are always considered to be at the same time instantiations of the superordinate types ("double glossing"). Supertypes and subtypes have regular glosses while qualified types have codes and values attached to the gloss of the superordinate type.

2.2 Use of iLex Structures by DTS Group

The DTS group already has dictionary entries with a semantic differentiation and intends to re-use these entries as sign types for annotation. For this purpose they have imported part of their dictionary entry structure into iLex.

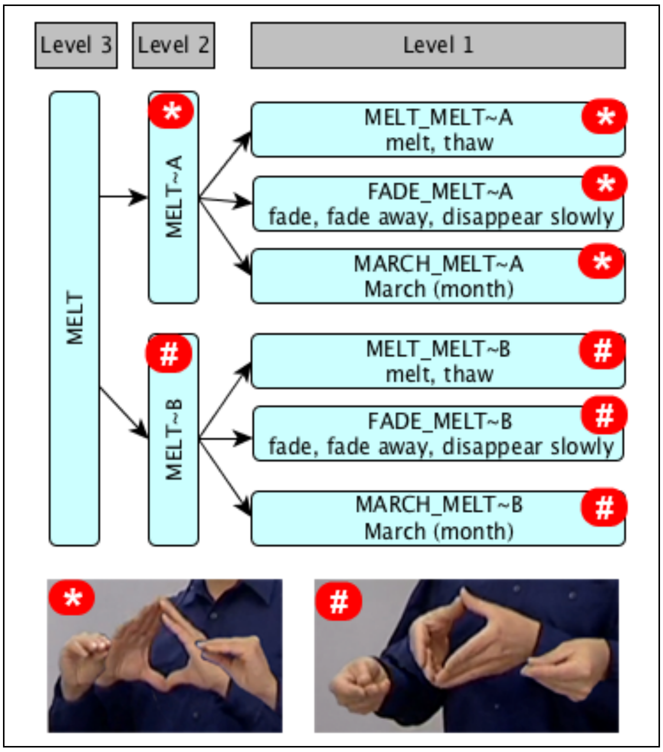


Figure 2: Type hierarchy of the DTS sign MELT

In the DTS Dictionary one sign entry may include different variant forms (one of which functions as citation form of the sign) and several senses of a polysemous sign. Three type levels have been used to model this structure in iLex. Types at level 3 (supertype) represent the whole lemma sign and therefore the whole entry. The different forms of a sign (GLOSS~A = citation form, GLOSS~B = variant) are represented as types at level 2. These form types are attached to the supertype. Subtypes (level 1) are used to represent the different senses as conventional uses of the sign in its respective variant form (see figure 2).

2.3 Differences in the Use of iLex Structures

The DGS and DTS groups use iLex structures in a quite similar way: Supertypes (level 3) represent the lemma sign, types of level 2 distinguish form variations, and subtypes (level 1) represent different conventional uses of the sign form (i.e. roughly meanings (DGS) or senses (DTS)). However, the DGS group does not repeat the supertype form (technically functioning also as citation form) on level 2 but links conventional uses of that form directly to the supertype, whereas the DTS group has all form types on level 2 (replicating the citation form) and therefore no direct linking from subtype to supertype. Apart from glossing conventions for phonological variants, the DGS group codes form types via qualifiers (and their values) to categorize form differences with regard to the citation form across types, while the DTS group does not use qualifiers at the moment.
The iLex database does not dictate how the type structures should be used, and in addition to the two models described above, the system can be designed to work with any model from a one-dimensional type list to a complex multi-level structure.

3. Lexicograhic Needs

3.1 Corpus Data as Basis for Sign Description

Annotated sign language (SL) texts should serve as the basis for the different kinds of analyses performed during the lexicographic description of a sign, e.g. establishing overviews of phonological variants and modified forms, of meanings, of usage (collocations, grammatical functions), or of distribution (with regard to region, age, gender etc.). Therefore, tokens that are likely to end up in one dictionary entry (or in one sense) should be tagged uniformly during corpus annotation.
Furthermore, the corpus system should provide tools for performing these analyses, e.g. tokens in context-view [concordance view], frequency lists, collocation statistics (Mutual Information, T-score etc.). In addition, the system should facilitate access to information from outside the corpus itself, e.g. data from informant surveys.[7]

3.2 The Lexicographic Workbench

When determining the final meanings' structure of a dictionary entry, it is good practice first to get an overview by describing the occurring senses at a rather high level of detail, and only in a second step to lump together closely related senses, preferably preserving the preliminary, fine-grained analysis to be consulted in connection with later revisions of the entry (cf. Atkins & Rundell 98-101, 268). The ideal integrated corpus-dictionary system should hence accommodate a preliminary, "full" set of senses, as well as a "cleaned-up" set that constitutes the meanings' structure of the final entry. Furthermore, these sets should be linked together

[7] The DGS Corpus Project uses an online survey called *DGS Feedback* to gather further information on signs and their use (cf. Langer et al. 2014, Langer et al. 2016). The results of this survey are complementary to corpus data and should be easily accessible when making lexicographic analyses.

in order to keep track of to where which senses go, and to more easily change the structure at later revisions. On both processing stages, there should also be place for storing further information regarding the decisions made during the analyses, e.g. hypotheses, questions, and comments.

In addition to this, a joint corpus-dictionary database should obviously accommodate all information kinds one would like to have in the dictionary's entry structure, as well as the needed meta-data such as markers for status, workflow control etc.

3.3 Corpus Data as Empirical Evidence

A dictionary entry is a sort of claim of giving an accurate description of the use of a sign, and links to actual corpus occurrences will provide accessible evidence for these claims. Thus, the system should allow for linking from each sense or grammatical or pragmatic function described in the dictionary data to corpus data, both on the higher levels (types in the lexical database for annotation), and on the lower (specific token or phrase tags in the annotations). Similarly, base form, variants and modified forms shown in the dictionary can be supported by evidence via links to corpus occurrences or links to types on various levels.

For some dictionaries, you might want to present authentic usage examples to the users. These could be taken directly from the corpus videos, or they could be adapted and re-recorded, e.g. for anonymisation reasons, or for making the examples more accessible for L2 learners (cf. Kristoffersen, 2010). In both cases, there will be a need for linking from a particular place in the dictionary data to a corpus occurrence. If example sentences are re-recorded, you might incorporate these recordings into the corpus system as a separate sub-corpus, in order to be able to link from the relevant dictionary sense both to the original source, and to the final version of the sentence.

4. Corpus Needs

4.1 Annotation and Lemmatisation

In addition to translation, the core task of basic annotation of SL texts is lemmatisation (cf. Johnston 2016, 13-48), also called token-type matching. Here the focus – the first criterion for matching – is on form, meaning being secondary and only rather roughly distinguished. As lemmatisation is very time-consuming it is essential that the annotator can find and identify relevant types as easily and fast as possible and with a reliable result.[8] One prerequisite for this is access to the up-to-date state of lexical entries (type entries). The system should also provide a number of easy-to-do searches via form, gloss, meaning, mouthings and combinations thereof across type entries and already lemmatised tokens.

For a fast check whether the found type is the correct one, the system should provide easy access to the citation form of the sign – for example by offering a representative video clip to be played (either a studio recording or an already lemmatised representative token), and also provide fast access to other tokens of that type for comparison. Also, when there is no fitting type to be found, annotators should be able to add a preliminary new type to the system.

Annotators should not be left in doubt what to do with tokens that are unusual with regard to their contextual meanings (productive uses or not yet identified conventional uses) or that are ambiguous in their meaning. The annotation conventions should cover these cases, and ideally the annotation tool should provide a mechanism to link them to a suitable type and at the same time keep them separate for further analysis, as it is the case when attaching all these tokens directly to the supertype. In this way annotators do not need to brood over meaning differences and the discrimination of various senses in the process of basic annotation.

Depending on lemmatisation rules it may be the case that two or more supertypes entries share the same citation form (homophony). In these cases, if it is unclear to the annotator which of these supertypes to choose, any of the possible supertypes could be regarded as suitable in the first annotation pass, and the decision of choosing a more specific type could be deferred to a later stage[9], see 4.2 below.

4.2 Lemma Revision

In order to insure consistency and quality of the lemmatisation, the DGS group found it helpful to establish a step they call *lemma revision* (cf. Konrad & Langer 2009). Here the focus is shifted from sequential text annotation to the single supertype and its forms and to some degree meanings. The token-type matching is checked in comparison to other tokens and the citation form. The tokens attached to the supertypes (productive uses) are checked for repeated occurrences of use in order to identify further conventional uses and establish new subtypes. The type structure is reviewed in the context of other types with related and similar forms and also taking into account sets of variants and modification behaviour. If necessary, the type structures are corrected or expanded. At the same time sign forms (modifications and variants) can be further distinguished (detailed annotation: levels 0 and 2). Cross-references between similar types are added. The result of the lemma revision is then a good basis for the ongoing lemmatisation. An annotation tool should allow one to conveniently access and collectively view all tokens of one (supertype) sign and compare them looking from different perspectives (form variation, meaning, relations to other signs etc).

5. Divergent Structural Needs

The type structure is the result of the lemmatisation process (including the lemma revision, cf. 4.2). Ideally, it

[8] The DGS Corpus database contains several thousands of type entries. In order to be able to find and identify the right supertype or subtype effective search strategies are necessary. iLex supports the lemmatisation by searches for and easy entering of the correct types into the transcript in various ways. When the annotator finds a good supertype candidate for token-type matching, the type hierarchy allows for getting a quick overview of the range of form and meaning aspects connected with one type to choose the best match.

[9] As the type hierarchy in the iLex database is flexible, another approach to solve the supertype homophony problem could be to introduce a formal, "form only" type above the supertype level, to be chosen when in doubt.

facilitates not only the ongoing annotation of corpus texts but also the building of a pre-structure of the data to be used as a basis for linguistic research, including lexicographic analysis and description of signs, their forms and uses. Lemmatisation in annotation focuses mainly (but not exclusively) on form in order to collect all instances of one sign under one label (be it an ID-gloss or an ID in a database), while establishing lemmas in lexicography focuses much more on the meanings of signs and has to consider additional factors (see 5.4). A rather complex example is illustrated in Fenlon et al. (2015, 196-198) where seven form variants are grouped into four separate lexemes in the BSL Sign Bank.

As implied above, the different sets of requirements might not always be fully compatible. Incompatibility problems may be due to annotation guidelines focussed on facilitating type search, to a clearer picture of the situation (e.g. different modification behaviour or different sets of variants) after analysis, or to pragmatic lexicographic decisions, e.g. keeping entries from becoming too large, complicated or counter-intuitive for the user. As a result, one will most likely in a number of cases end up with a different division or grouping of the data and types into separate or collective dictionary entries, thus diverging from the pre-structure built during the annotation.

In the following we will have a closer look at some of the factors that lead to structural incompatibility between corpus and dictionary, and at some concrete examples. Finally, we will shortly look at one of the challenges arising when combining both structures in interrelated products.

5.1 Ongoing Changes of Type Structures

Annotation is to be seen as an ongoing process of multiple passes each adding to and correcting the annotation while lemmatising or analysing or using the annotated data. In principle, this work is never finished as each look at the data with different goals or research questions in mind reveals new insights and can lead to an enrichment of the annotation.

In the DGS Corpus Project, up to 40 people are working at the same time in iLex on the data. New types are added as needed and type structures may be changed, expanded, rearranged or split as it seems necessary in the light of new data or analyses. Each change is available in real-time to all others. Therefore, type structures are at least in principle constantly changing.

At one point in time the scope of a dictionary entry has to be defined and fixed in order to be able to summarise and describe the data available. We are convinced that for this purpose we need a structure that allows us to do that and to fill in our information on this subset of data in a structured way without imposing dictionary writing decisions back onto the annotation database.[10]

5.2 Two Supertypes – One Dictionary Entry

Consider the two type structures of WORK2-$SAM (1) and END1-$SAM (10) in the annotation database (fig. 3).

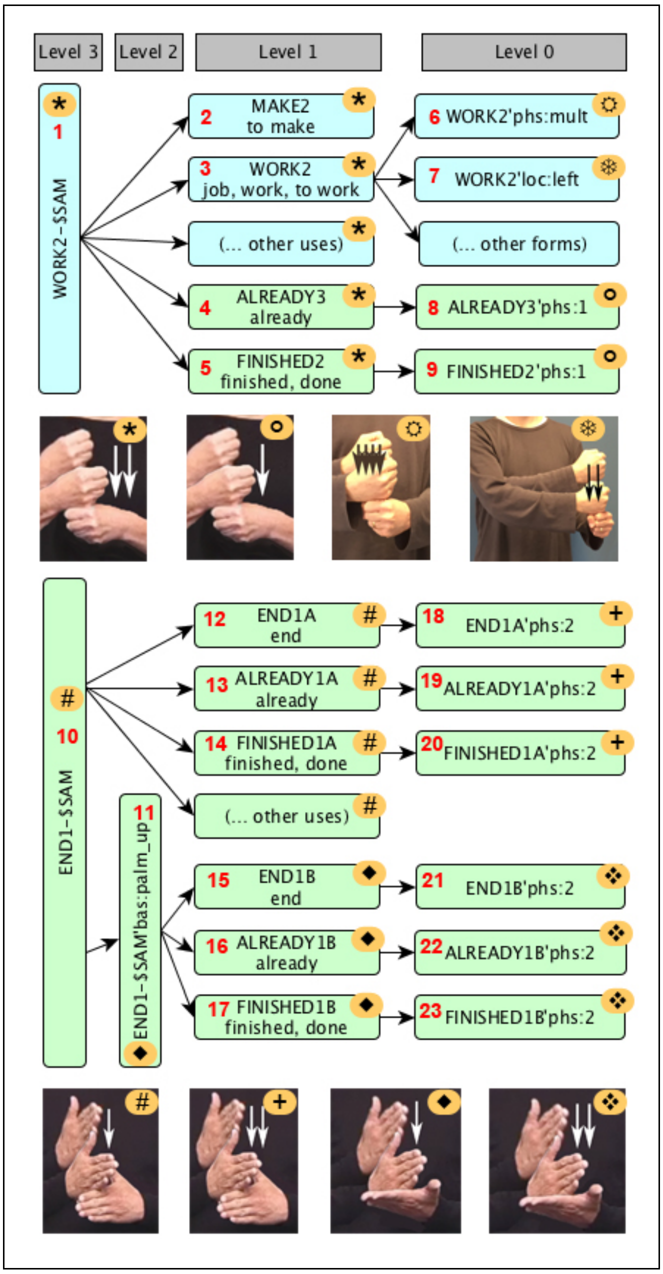


Figure 3: Type hierarchies of WORK2-$SAM and END1-$SAM (DGS)

The green parts of the WORK2-$SAM type structure are completely fitting with regard to form and one can also easily see a semantic closeness of the different uses of WORK2-$SAM which include among others 'to work' (3), 'to make' (2), 'done/finished' (5), 'already' (4) along the lines of '*work that has been done is already finished*'. For annotation this structure based on form as a first criterion is very convenient. However, the blue subtypes ('work', …) and the green subtypes ('finished', …) differ with regard to variants and modification behaviour, a fact that can be noticed when reviewing all the available data after lemmatisation. Another entry

[10] However, sometimes dictionary analyses and decisions may reveal problems in the annotational type structure of a sign and may lead to a restructuring of it as well. Our point here is not that this should not happen – in fact this happens a lot in the DGS Corpus Project and is even an appreciated way of quality control – but that the system should be flexible enough to have the freedom of diverging groupings of data for annotation and lexicography.

WORK1-$SAM (circular movement to hit the base hand repeatedly) is very similar in form to WORK2-$SAM and also shares most of its meanings in the *work* domain, but not the senses 'already' and 'finished, done'. Both signs share modification behaviour, for example they can be used in an intensified form with a smaller movement including several repetitions (6). On the other side there is END1-$SAM with its two variant forms (10) and (11) being very similar to the green parts of the WORK2-$SAM structure. (10) and (1) differ only with respect to the parameters handshape and ±repetition. These forms could be considered phonological variants on basis of similarity of form and meaning, especially since they both can be made with repetition (4, 5, 18-20) and without repetition (8, 9, 12-14). This is also true for the third form (11), which differs from (10) only in the orientation of the base hand.

Considering the differences in modification behaviour of WORK2 (3) and other uses in the *work* domain and ALREADY3 (4) and FINISHED2 (5) as well as the formational similarity and semantic overlap of ALREADY3/FINISHED2 with the END1-$SAM sign uses, it seems reasonable and adequate to virtually re-group the material into to two different dictionary entries WORK and FINISHED, as can be seen in figure 4.

WORK	FINISHED
Citation form: * ⇨1 Variant: [HamNoSys] ⇨WORK1-$SAM Modification behaviour: can be modified for loci ⇨7, … intensive: ⚙ ⇨6 … Senses: 1. work (task), to work ⇨3 2. work, job ⇨3 3. to make ⇨2 4. …	Citation form: # ⇨10 Variants: ◆ ⇨11, ○ ⇨8, 9 Note on form: All 3 forms can be made with repetition ⇨4, 5, 18-23 … Senses: 1. end ⇨12, 13 2. finished, done ⇨5, 14, 17 3. already ⇨4, 13, 16 4. …

Figure 4: Pre-dictionary entries[11] of WORK, FINISHED

5.3 One Supertype – Two Dictionary Entries

In the DGS group there is a tendency in the first pass of basic annotation to formationally group derived forms under the more common or basic sign form (supertype), even in cases where the lexicographer would arrive at the interpretation of two related but independent signs. An example for this is STAMP1-$SAM.

The sign type hierarchy of the sign STAMP1-$SAM (30, (see figure 5) would be split into two entries, one covering the blue and the other covering the green part of the structure. The common citation form (30) could be interpreted as an iconic resemblance of pressing a stamp onto a sheet of paper (fist onto flat hand), and it has several conventional uses including: 'civil servant', 'agency', 'patent', 'visa', 'authentification', 'to stamp'. It has a form variant (31) where the non-dominant hand is not a flat hand but also a fist.

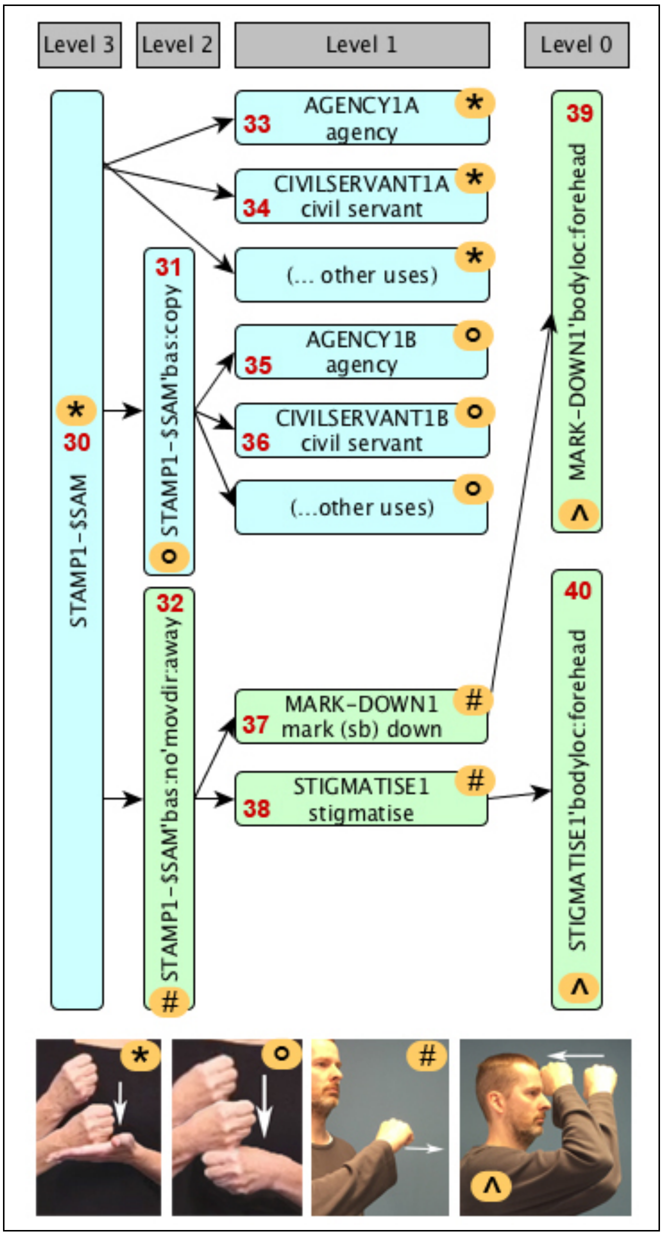


Figure 5: Type hierarchy of STAMP1-$SAM (DGS)

There is a derived form of the sign STAMP1-$SAM made with only one hand that moves forward instead of downward as if to stamp somebody else on the forehead (32). This form is conventionally used with the meanings 'mark somebody down' (37) or 'stigmatize' (38), both figurative uses of 'stamp'. Both (37) and (38) presumably[12] can appear in context also in the form of stamping the signer's own forehead (39, 40). Even though (32) is most likely derived from (30) it would get its own dictionary entry MARK-DOWN independent of the dictionary entry STAMP (see figure 6) because the uses of the sign differ with regard to form and meaning[13] from

[11] Pre-dictionary entries are content of the pre-dictionary database as described by Atkins & Rundell (2008, 98-100) and therefore of the Dictionary Writing System module planned for iLex. Everything red in the entries refers to evidence in the corpus (types or particular tokens of that type in the annotation database) and could be instantiated by links to the corpus data and annotation types. Red numbers and orange codes for form refer to numbers and codes in figure 3.

[12] No corpus data yet to back up this claim.

[13] Cf. Battison (2005, 240): "If two signs are made differently,

the other uses of the sign STAMP1-$SAM (e.g. 33-36) and also because it exhibits different modification behaviour.[14]

STAMP	MARK-DOWN
Citation form: * ⇨30	Citation form: # ⇨ 32
Variant: o ⇨31	Word forms:
…	1st person: ^ ⇨39, 40
Senses:	…
1. end ⇨12, 13	Senses:
2. finished, done ⇨5, 14, 17	1. mark somebody down ⇨37
3. already ⇨4, 13, 16	2. stigmatise ⇨38
4. …	

Figure 6: Pre-Dictionary entries[15] of STAMP and MARK-DOWN (DGS)

5.4 Editorial Principles

The task of defining exactly what constitutes a dictionary lemma differs from the corresponding task performed in connection with annotation (lemmatisation). In lexicography, the focus is more towards the meanings of signs and has to consider additional factors, including even practical matters such as entry size and user-friendliness. Typically, dictionaries have their own "lemmatisation rules" describing which kinds of words (or signs) to include into the dictionary, which to give their own independent entries and which to treat as run-ons or sublemmas, specifying how the particular dictionary treats cases of homonymy and polysemy. Svensén calls this step the *establishment of lemmas* and distinguishes it from lemmatisation (cf. Svensén 2009, 94).[16]
The DTS group works with principles for establishing lemmas that are partially based on phonological, partially on semantic criteria (cf. Kristoffersen & Troelsgård, 2010). One criterion is that a figurative use of a sign will be described as a sense in the main entry only if the semantic relation is synchronically transparent. Otherwise, the figurative use will be established as a separate (homophone) sign entry. An example of this is a sign that can mean 'red' as well as 'social'. As a consequence of the editorial principles of the dictionary, the sign is formally split into two entries, RED and SOCIAL (fig. 7).

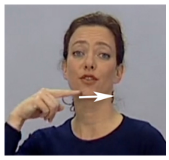
RED Senses: red		SOCIAL Senses: social	

Figure 7: Dictionary entries of RED and SOCIAL (DTS)

Another principle is that a sign described as having several variant forms can include only senses that can all be expressed through one particular variant (shown as the citation form in the dictionary). Any variant-specific sense gets its own independent sign entry in the dictionary. An example is the sign PRAY (see figure 8), with two variants, of which only one can mean 'sorry'. Hence, although 'sorry' could be considered as semantically transparently related to 'pray' or 'beg' (= 'beg for forgiveness'), it is established as a separate lemma.

SORRY	PRAY	
Sign form:	Sign form 1:	Sign form 2:
		
Senses: sorry, pardon	Senses: pray, prayer	

Figure 8: Dictionary entries SORRY and PRAY (DTS)

These examples from the DTS Dictionary show that the needs of lexicographers and annotators are different due to their differences of viewing the same phenomena from different perspectives. It would be unwise to force the dictionary decisions onto the annotation database structure for reasons of resources but also because different dictionaries made of the same corpus lexical database / corpus may work with different lemma definitions targeting their specific user groups.

5.5 The Glossing Challenge

In spite of the incompatibility problems mentioned above, a corpus project and a dictionary based thereon could easily coexist if they were regarded as two independent language resources. However, having two such interlinked resources gives you some advantages that you would obviously like to pass on to the users, e.g. by showing corpus data directly in the dictionary, by linking from the dictionary to corpus occurrences of a sign or a sense, or by linking from corpus searches to relevant dictionary entries. And here – when presenting the resources as two interrelated products (or perhaps even as one integrated website) – the challenge arises of how to represent the signs in a user-friendly way. For a DGS sign entry like FINISHED (see figure 4), the dictionary entry as a whole cannot unambiguously refer back to the right corpus type gloss. The corpus on the other hand will contain many sign types that are not covered at all by the dictionary. If both products work with glosses as labels, using different gloss systems might be confusing for the user and a merging of the glosses might prove difficult.

and have different meanings, this is good evidence that they are separate signs."

[14] Actually, during lemma revision the branch starting with (32) was taken out of the type hierarchy of STAMP1-$SAM and got its own supertype entry also in the annotation database.

[15] Red numbers and orange form codes refer to figure 5. See footnote 13 for further explanations.

[16] "…ESTABLISHMENT OF LEMMAS. This operation is not simply the same thing as lemmatization… Naturally, the establishment of lemmas assumes previous lemmatization, but it also includes deciding how lexical items having identical base forms are to be presented in the dictionary, and to what extent word elements and multi-word lexical items are to be accorded lemma status." (Svensén 2009, 94)

6. Conclusions

6.1 Structural Requirements

As we have shown above, a lexical database structure built for corpus annotation and one built for dictionary-making serve different purposes, and may therefore arrive at different suggestions for a structured description of the form and meaning of a particular sign. Where the lexicographic side typically is governed (to a higher degree) by semantic criteria, as well as by editorial rules and pragmatic decisions, the annotation side is mainly form-oriented, and should ideally facilitate that the lookup of a particular sign form always leads the user to the same type entry. Trying to perform a concurrent adjustment of the annotation base structure in order to match the dictionary structure at any time would eventually obscure the originally form-based structure, and hence hamper the lemma identification.[17] We therefore suggest a model with different structures for annotation and dictionary editing, that each serves its primary purpose optimally on one hand, and on the other hand are sufficiently similar to be related to each other – and to be interlinked.

6.2 Linking Corpus and Dictionary

As shown in section 5, divergences between annotation principles and editorial principles can lead to a need of linking from one entry in the annotation database to two or more entries in the dictionary, as well as the other way round. Hence the suggested structure should allow for a rather free linking structure. Furthermore, linking will be needed between a series of places in the dictionary structure and different places in the lexical database used for annotation.

If we walk through the entry structure of a comprehensive SL dictionary, links to corpus evidence could be relevant in many places. On the entry level, you would probably like to refer to the sign itself, as evidence of its phonological variants, just as you typically would list the citation form and variants of the headword in a written/spoken language dictionary. These references could ideally be established as links to one or more types (on any level) in the type hierarchy of the annotation base, suggesting that the sign in question matches the types perfectly. Knowing that in many cases this will not be possible, we suggest that the structure allows for linking not only to types, but also to individual tokens in the transcripts.

Similarly, where a written/spoken language dictionary shows inflected forms of the headword, you might want to show frequent modifications of the sign, with no regard to meaning (linking to level 2 types) or to a group of tokens being instantiations of this modification (linking to level 0 types).

On the meaning level, links to evidence for the described senses would obviously be desirable and appropriate, as they document the performed semantic analysis. This documentation could be valuable not only from the lexicographic perspective, e.g. in connection with later revisions of the dictionary (based on an updated and extended corpus), but also from other perspectives, e.g. in connection with other linguistic research based on the corpus. Just as the linking on entry level, this linking could have either a type (all tokens of that type are covered by that sense) or a single token as target in the annotation database.

On the meaning level, you might also need evidence of particular modified forms of the signs, used in a particular sense. Again, linking both to type and token could be needed.

On the meaning level, it would also be obvious to place links to good usage examples found in the corpus (as well as to reproductions of these, cf. section 3.3). Evidence of frequent collocations with the sign would also be a linking candidate. These latter types of linking could be done to a phrase/utterance in a corpus transcript, instead of to a single token. Figure 9 shows a linking model that accommodates the basic needs as described above.

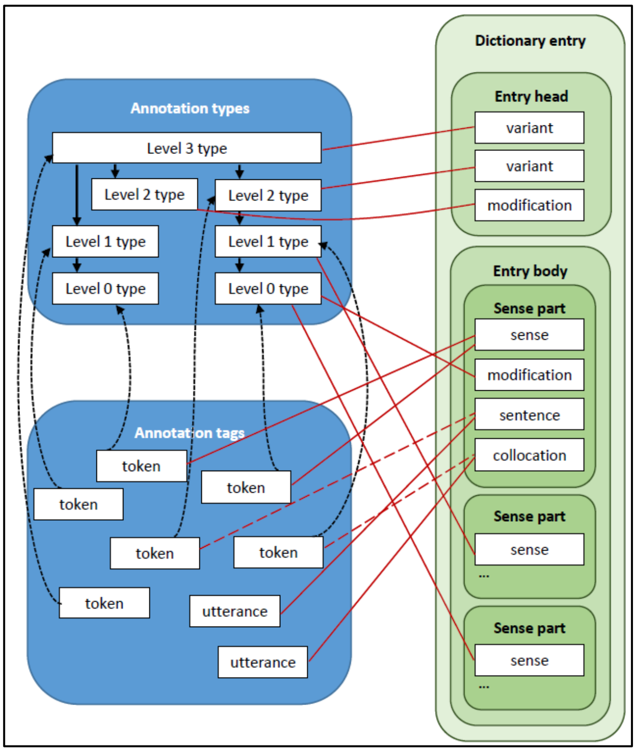


Figure 9: Suggested model for linking between corpus and dictionary

6.3 Future work

In the future, the two project groups will continue the work, aiming at adapting the iLex system to accommodate dictionary data, and to facilitate linking

[17] In the LSP dictionary projects in Hamburg, up to 2010, sign entries had been produced directly from the iLex type entries via an export routine. Prior and close to the production, changes to iLex type entries had to be halted and in some cases type structures had to be adapted to suit the intended outcome as dictionary entries. Based on this experience the DGS group thinks it advisable to have a separate structure for the preparation of dictionary entries. Atkins & Rundell (2008, 98-100) also make a strong case for such a structure, which they call *pre-dictionary database*. Ideally, this structure should be part of the iLex environment and be interlinked with the annotation types.

between these data and appropriate types and tokens in the corpus data, as outlined above.

7. Acknowledgements

(Hamburg Group) This publication has been produced in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.

(Copenhagen Group) The work at the DTS Dictionary project is supported by funding from the "Diversity and Social Innovation" research fund of UCC (Professionshøjskolen University College Capital, Denmark), from Alfred Jacobsens Foundation (Denmark), and from the Jascha Foundation (Denmark).

8. Bibliographical References

Atkins, B.T.S, Rundell, M. (2008). *The Oxford Guide to Practical Lexicography.* Oxford: Oxford University Press.

Battison, R. (2005). Signs Have Parts. A Simple Idea. In C. Valli, C. Lucas, & K.J. Mulrooney (Eds.), *Linguistics of American Sign Language. An Introduction*. 4th ed. Washington, DC: Gallaudet University Press, pp. 230--241.

Crasborn, O., Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood, & E. Thoutenhoofd (Eds.), *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages*. ELDA, Paris, pp. 39--43.

Fenlon, J., Cormier, K., Schembri, A. (2015). Building BSL SignBank: The Lemma Dilemma Revisited. *International Journal of Lexicography*, 28(2), pp. 169--206. doi:10.1093/ijl/ecv008.

Hanke, T. (2002). iLex. A tool for Sign Language Lexicography and Corpus Analysis. In M. González Rodriguez, & 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, pp. 923--926. [Online resource; URL: http://www.lrec-conf.org/proceedings/lrec2002/pdf/330.pdf ; last access: March 22, 2016]

Hanke, T., Storz, J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In O. Crasborn, T. Hanke, E. Efthimiou, I. Zwitserlood, & E. Thoutenhoofd (Eds.), *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages.* Paris: ELRA, pp. 64--67.

Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, 15(1), pp. 106--131.

Johnston, T. (2016). Auslan Corpus Annotation Guidelines. February 2016 version. [Online resource; URL: http://www.academia.edu/12360442/Auslan_Corpus_Annotation_Guidelines_February_2016_version_; last access: March 22, 2016].

König, S., Konrad, R., Langer, G. (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, pp. 379--404.

Konrad, R. (2011). *Die Erstellung von Fachgebärdenlexika am Institut für Deutsche Gebärdensprache (IDGS) der Universität Hamburg (1993-2010).* Universität Hamburg. [Online resource; URL: http://www.sign-lang.uni-hamburg.de/projekte/mfl/konrad_2011_fachgeblexika.pdf; last access: March 22, 2016].

Konrad, R., Hanke, T., König, S., Langer, G., Matthes, S., Nishio, R., Regen, A. (2012). From form to function. A database approach to handle lexicon building and spotting token forms in sign languages. In O. Crasborn, E. Efthimiou, S.-E. Fotinea, T. Hanke, J. Kristoffersen, & J. Mesch (Eds.), *Interaction between Corpus and Lexicon. Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages. 8th International Conference on Language Resources and Evaluation, LREC 2012, Istanbul, Turkey*. Paris: ELRA, pp. 87–94. [Online resource; URL: http://www.lrec-conf.org/proceedings/lrec2012/workshops/24.Proceedings_SignLanguage.pdf; last access: March 22, 2016].

Konrad, R., Langer, G. (2009). Synergies between transcription and lexical database building: The case of German Sign Language (DGS). In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.), *Proceedings of the Corpus Linguistics Conference (CL2009). University of Liverpool, UK, 20-23 July 2009*. [Online resource; URL: http://ucrel.lancs.ac.uk/publications/cl2009/346_FullPaper.doc; last access: March 22, 2016].

Kristoffersen, J.H. (2010). From utterance to example sentence – a lexicographic approach to exploitation of corpus. [Online resource; URL: http://www.tegnsprog.dk/hovedside/litteratur/SLCN_Berlin_2010_handout.pdf; last access: March 22, 2016].

Kristoffersen, J.H., Troelsgård, T. (2010). Making a dictionary without words: lemmatization problems in a sign language dictionary. In S. Granger, M. Paquot (Eds.), *eLexicography in the 21st century: New challenges, new applications. Proceedings of eLex 2009, Louvain-la-Neuve, 22-24 October 2009.* Louvain-la-Neuve: Presses universitaires de Louvain, pp. 165--172.

Kristoffersen, J.H., Troelsgård, T. (2012). The electronic lexicographical treatment of sign languages: The Danish Sign Language Dictionary. In S. Granger, M. Paqout (Eds.), *Electronic Lexicography.* Oxford:

Oxford University Press, pp. 293--315.
Langer, G., König, S., Matthes, S. (2014). Compiling a Basic Vocabulary for German Sign Language (DGS) – lexicographic issues with a focus on word senses. In A. Abel, C. Vettori, & N. Ralli (Eds.), *Proceedings of the XVI EURALEX International Congress: The User in Focus, July 15-19 2014 in Bolzano/Bozen – Italy*, pp. 767--786. [Online resource; URL: http://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/LangerKoenigMattesEURALEX2014.pdf; last access: March 22, 2016].
Langer, G., König, S., Matthes, S., Groß, N., Hanke, T. (2016). Variation of DGS lexical items. What sign language lexicography can gain from a mixed method approach: Corpus data supplemented by crowd sourcing. Poster presented at the 12th International Conference on Theoretical Issues in Sign Language Research (TISLR12), Jan 4-7, 2016 at Melbourne, Australia. [Online resource; URL: http://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/TISLR_poster_mixedmethods_18k_korr_Version.pdf; last access: March 22, 2016].
Svensén, B. (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making.* Cambridge: Cambridge University Press.

9. Language Resource References

Center for Tegnsprog (2008-2016). *Ordbog over Dansk Tegnsprog*. Available at: http://www.tegnsprog.dk.
DGS Corpus Project (2015-2016): Subcorpus. Available at: http://meine-dgs.de.

A New Tool to Facilitate Prosodic Analysis of Motion Capture Data and a Data-Driven Technique for the Improvement of Avatar Motion

John McDonald[1], Rosalee Wolfe[1], Ronnie B. Wilbur[2], Robyn Moncrief[1], Evie Malaia[3], Sayuri Fujimoto[1], Souad Baowidan[1], Jessika Stec[1]

[1]DePaul University, Chicago, USA

[2]Purdue University, West Lafayette, USA

[3]Netherlands Institute for Advanced Study, the Netherlands

E-mail: jmcdonald@cs.depaul.edu, wolfe@cs.depaul.edu, wilbur@purdue.edu, rkelley5@mail.depaul.edu, evie.malaia@nias.knaw.nl, sfujimoto23@gmail.com, sbaowida@mail.depaul.edu, jessika.stec@gmail.com

Abstract

Researchers have been investigating the potential rewards of utilizing motion capture for linguistic analysis, but have encountered challenges when processing it. A significant problem is the nature of the data: along with the signal produced by the signer, it also contains noise. The first part of this paper is an exposition on the origins of noise and its relationship to motion capture data of signed utterances. The second part presents a tool, based on established mathematical principles, for removing or isolating noise to facilitate prosodic analysis. This tool yields surprising insights into a data-driven strategy for a parsimonious model of life-like appearance in a sparse key-frame avatar.

Keywords: motion capture analysis, sign language synthesis, avatar technology, noise

1. A Simple Case Study as Motivation

Noise is an unwanted modification to motion capture data that occurs during recording. The following example illustrates how noise poses barriers to the analysis of prosodic structure. Figure 1 is a time graph taken from a motion capture session (Wilbur and Malaia In Press). It displays the y-coordinate (height) of the right wrist over a two-second period at the beginning of the sentence 'Newspaper said there was an awful storm in Florida where homes, cars, and trees were destroyed.' The first two seconds contain the signs 'NEWSPAPER READ'. Although the signal looks smooth to the casual observer, problems arise when using the data to compute changes in speed as a precursor to examining prosody.

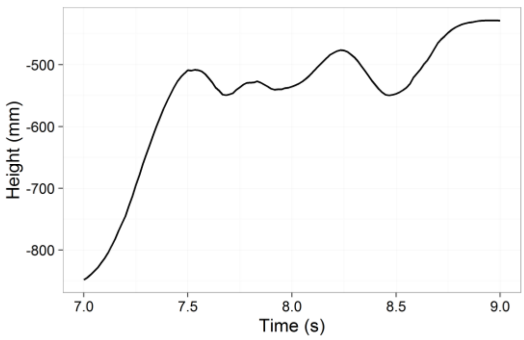


Figure 1: Height Information for a Right Wrist Marker.

Determining changes in speed is a two-step process. The first computes the speed from the marker's position data using a central difference approximation for the derivative:

$$s = \frac{dp}{dt} = \left| \frac{p_{i+1} - p_{i-1}}{2\Delta t} \right|$$

Figure 2 is a graph of the wrist marker's speed. The curve contains many small spikes which are due to the noise contained in the original position data.

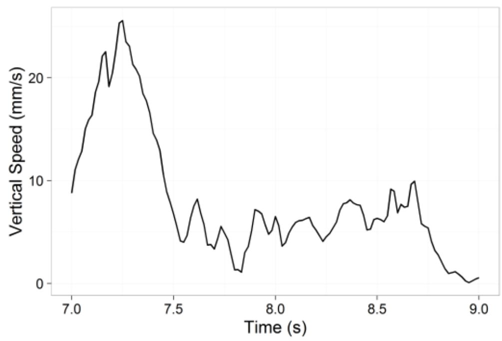


Figure 2: Speed of Right Wrist.

The second step computes the change in speed, which is essential for studying prosody:

$$\frac{ds}{dt} = \frac{s_{i+1} - s_{i-1}}{2\Delta t}$$

Figure 3 is a graph of the result. The spikes are even larger and dominate the curve. This jagged curve gives the impression of jerky motion, but the original position graph in Figure 1 reflects the smoothly flowing discourse of a fluent signer as confirmed in the original video.

The noise that was barely perceptible in Figure 1 has been magnified to the point where it is difficult to use visual inspection to identify any aspect in the prosodic structure of the utterance. From this example, it is clear that the motion capture data contains noise, but the question remains as to its origins and severity. Effective analysis requires its isolation and/or removal.

2. Fundamentals and Terminology of Signal Processing

This section presents a brief outline of the principles of signal processing used to clean a motion capture data stream. These principles are applicable to the analysis of any time series data, including motion capture. A more in-depth treatment can be found in (Smith 2011).

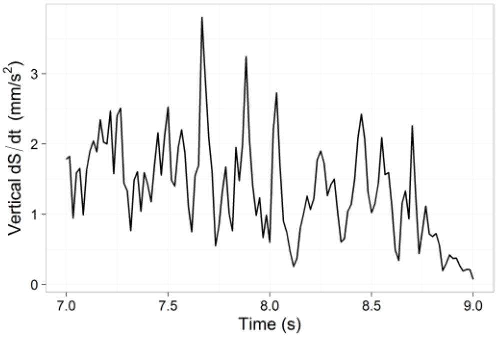


Figure 3: Change in Right Wrist Speed.

Several important concepts of signal processing can be analyzed from an idealized production of the word BICYCLE in American Sign Language (ASL). In this sign, the height of the right hand oscillates vertically in a regular manner similar to the idealized graph shown in Figure 4. Since the horizontal axis of this graph is time, this plot is said to be in the *time domain*.

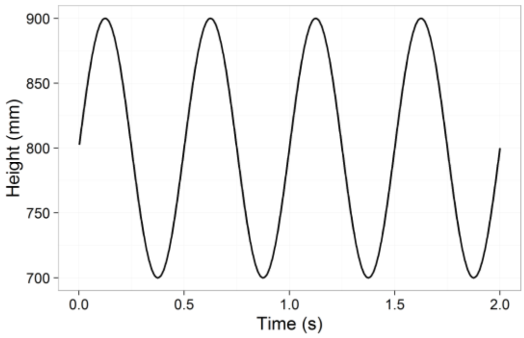


Figure 4: Height Data of a Wrist from an Idealized Production of BICYCLE.

The size of the oscillation is called the *amplitude* of the signal, whereas the speed at which the hand moves through the oscillation is its *frequency*. Amplitude is measured in units such as millimeters (mm), and frequency is measured in *cycles per second* also known as Hertz (Hz).

Unfortunately, the signal is rarely as simple as in Figure 4. Returning to Figure 1, the oscillations in the graph show variation in both their length and size. Thus, these oscillations change in both amplitude and frequency over the course of the phrase. To analyze more complicated signals, we need the Fourier transform (Duhamel and Vetterli 1990), which decomposes a signal into a collection of contributing pure oscillations. Figure 5 shows a density plot, analogous to a histogram, of all the oscillations present in the signal from Figure 1. This plot is called the signal's *spectrum* in the *frequency domain*, since it displays the strengths of the signal's oscillations at various frequencies, which are shown on the horizontal axis.

This spectrum was constructed with a Fourier transform on the original time-domain signal, and yields a list of amplitudes in the frequency domain, which we can then analyze and edit. For example, Figure 6 contains a plot of the signer's right wrist height while standing still with arms raised in a calibration posture.

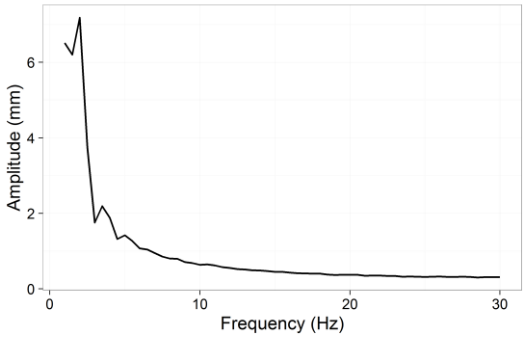


Figure 5: Frequency Spectrum of the Wrist Height During the First Two Seconds of the Phrase.

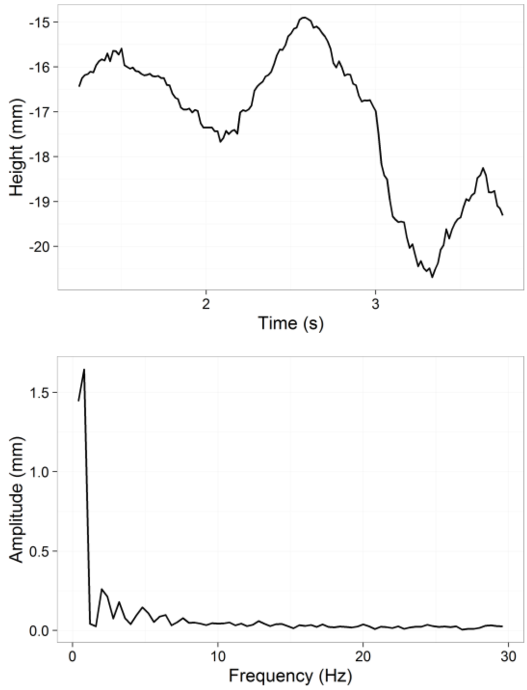


Figure 6: Time and Frequency Domain Plots of Signer with Arms Up.

An analysis of the right wrist height and its resulting spectrum yields one main low-frequency oscillation with a spread of smaller amplitudes at higher frequencies. These are fast, but tiny oscillations around a slow variation of the wrist height that occurs as the signer attempts to hold still.

Returning to the motion in Figure 5, we see a more complicated profile with a main high amplitude signal at low frequencies and then a smooth falloff in amplitude at higher frequencies. Their small amplitudes indicate that these fast oscillations contribute little to the signal. It is this *noise* that software needs to remove before meaningful analysis can be performed.

For our purposes, removing unwanted high frequencies will not alter the main signing signal. We do this by means of a *low-pass filter*, which sets all the frequency amplitudes above a certain threshold to zero. After the suppression of these amplitudes, we can recover the cleaned signal by inverting the Fourier transform, yielding a smoother trajectory for the wrist. The cleaned signal will rarely deviate from the original by more than a fraction of a millimeter. In our study, over 99% of the samples deviated by less than a millimeter.

3. Analyzing Noise in Sign Language Motion Corpora

This section discusses practical considerations for determining which frequencies are relevant to linguistic research and which can be safely considered as noise. Figure 7 contains a conceptual diagram of a spectrum for a coordinate value of a position marker in the frequency domain. The vertical axis is amplitude and the horizontal axis is frequency.

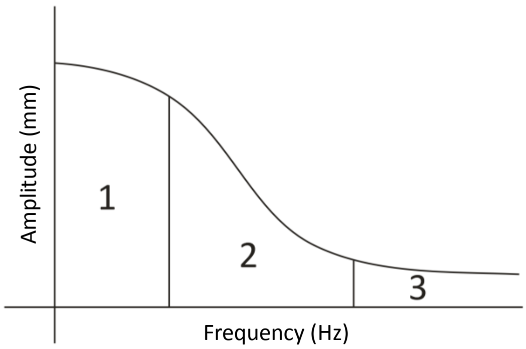


Figure 7: Conceptual Regions of Positional Data Graphed in the Frequency Domain.

The frequency spectrum in this diagram is divided into three sections which have different impacts on sign analysis. We begin with the region marked "3", representing frequencies above 12 Hz. According to (Marshall and Walsh 1956), the muscles in the human body cannot create oscillations faster than 10-12 Hz, and so the frequencies in this region can thus be seen as noise attributable to fluctuations in the recording technology. These frequencies can safely be eliminated before performing further analysis of the signal.

Frequencies slower than 10-12Hz, in regions 1 and 2, may be produced by human motion. However, not all such frequencies of motion have linguistic meaning for sign language. This can be clearly seen by looking at the types of motion that the human body produces in sign discourse and the oscillations of parts of a signer's body involved in such motion. On the slower end of the scale, oscillations on hip markers correspond to such linguistic processes as *role shift*. Due to the sheer mass involved in moving the human torso, these motions will have lower frequencies of no more than 0.5 Hz. In contrast, fingers being of much lower mass and smaller movements, are capable of higher frequencies, such as the motion displayed in fingerspelling or in internal movement such as trilling (WAIT, FINGERSPELL), but even here the cutoff is no more than 4 Hz as can be seen in analyses of finger spelling rates (Quinto-Pozos 2010).

So, the region in the diagram marked "1" contains the main low frequency movements generated by sign language production. The cutoff for this region will depend on a marker's placement, with lower frequencies for markers on the trunk of the body and higher frequencies at more distal markers. Table 1 gives a set of empirically-determined frequency cutoffs for intermediate markers. These limits are deliberately conservative to assure that no aspect of a human linguistic utterance is being compromised.

Joint	Frequency (Hz)
Hips	0.25
Waist	0.5
Upper spine	0.5
Neck	1.0
Shoulders	1.0
Elbow	2.0
Wrist	2.0

Table 1: Frequency Cut-offs for Selected Markers.

For linguistic analysis, we can clean the position data by converting it to the frequency domain, setting the amplitude of the frequencies in regions 2 and 3 to zero, and using the modified spectrum to reconstitute the marker's position in the time domain via an inverse Fourier transform. From the cleaned data, we proceed with the calculations for speed and speed change. The resulting graphs shown in Figure 8 do not exhibit the spikes seen in Figures 2 and 3.

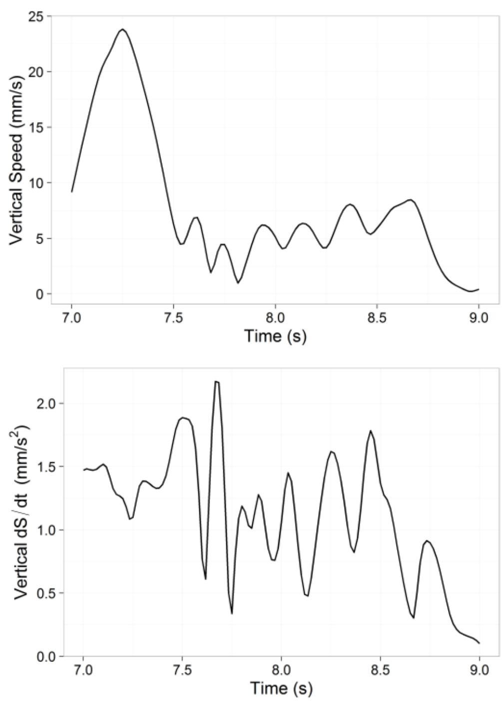


Figure 8: Speed and Change of Speed Computed with Cleaned Position Data.

4. A New Tool

To aid in isolating or removing noise from motion capture data, we created a software suite called SignCleaner to aid in the signal processing of motion capture data of signed utterances. The system accepts HTR, a common format of motion capture data (Parent, et al. 2009) and can accommodate any number of markers. The suite is available for download at http://tinyurl.com/jfysn2t and consists of two parts. The first part is a C# application that translates HTR data into a comma delimited (.csv) file compatible with the R statistical computation environment (R Core Team 2000). The second part is a collection of R scripts that perform the following:

- Removing noise (cleaning) using a Butterworth filter, based on a Fourier transform (Hong and Lancaster 2004). A Butterworth filter tapers the attenuation of the frequencies being removed for a highly smooth result. Researchers can adjust the frequency cutoffs to best accommodate their analyses.
- Computing speed and change of speed for each marker. Since these are scalar metrics, they lend themselves to easy visualization in the time domain.
- Visualizing the data to facilitate inspection for patterns or trends.
- Exporting the position, speed and speed change of markers as a CSV file, suitable for use in ELAN (Crasborn, et al. 2006).

The tool has been validated on a subset of the Wilbur corpus (Wilbur, et al. 2011), consisting of 58 markers with 9400 data points per marker. Figure 9 shows a screen shot of an ELAN session, showing a segment of the speed and change of speed of the right wrist sensor. Both measures are computed with the cleaned position data and the original, uncleaned data. The lighter curves in each track show the results from the original noisy position, whereas the darker curves are computed from the cleaned data.

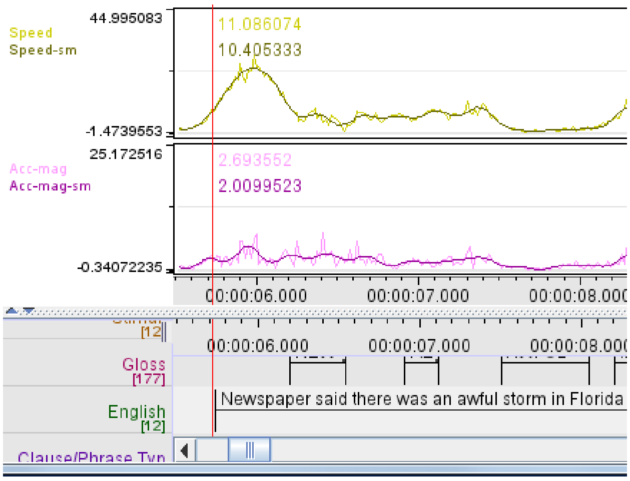


Figure 9: Elan Interface for Motion Plot Analysis.

5. A Novel Finding and its Application to Avatar Technology

Our discussion of Figure 7 did not consider the entire spectrum, so we return to it now. From the diagram, we know that we want to eliminate the frequencies in region 3 as they are noise introduced by the recording technology. Further we want to retain the frequencies in region 1 for linguistic analysis. This leaves region 2, which contains frequencies that are not of linguistic significance, but are none the less created by a human while producing signed utterances. From the perspective of linguistic analysis, this is noise, but from the perspective of avatar technology, this is valuable information for enlivening an avatar.

In order to create the illusion of life, avatars must continue to move, even when a signed discourse has concluded. A living human body is never completely still, even when at rest, and the human mind and visual sense are highly attuned to expect this dynamic. An avatar at rest needs to continually display subtle movements to avoid being perceived as a static image. This is a particular challenge for sparse-key animation systems (Perlin 1996)

In entertainment technology, two common techniques used to maintain the dynamics of an avatar are

- the manual adjustment of motion curves by an animation artist (Gleicher 1998), and
- the introduction of Perlin noise.

Since hand animation is time-consuming and expensive, Perlin noise is preferred because it can be automated (Perlin 1995). It can be tuned to a specific set of frequencies (Lagae, et al. 2010) and is therefore ideal for this situation. We can tune this type of noise so that it primarily contains frequencies in region 2, the enlivening frequencies, and these will be perceptible in the finished animation. Figure 10 shows the frequency spectrum for a version of Perlin noise tuned to roughly match the three regions of Figure 7.

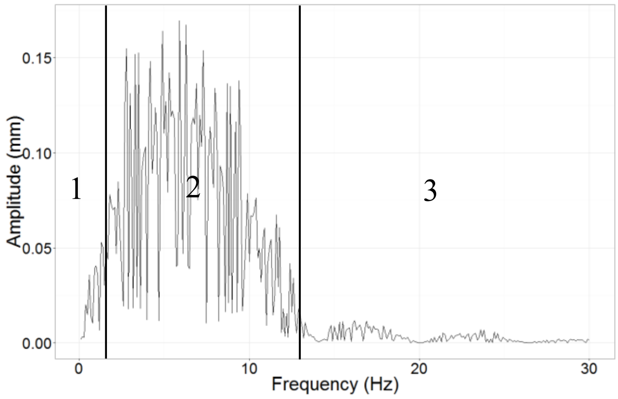


Figure 10: Spectrum of Perlin Noise.

The frequencies in this plot are essentially bounded on the right, and so there are very few high frequencies corresponding to region 3. In addition, the amplitudes of its low frequencies in region 1 are small enough so that the addition of this noise will not interfere with any intended animations such as a signed utterance. Since the range of frequencies is bounded on both the lower and upper ends, it corresponds nicely with enlivening region 2 of Figure 7.

Traditionally, Perlin noise is only applied in situations where the avatar has otherwise stopped moving, however an abrupt transition to Perlin noise is incompatible with the high fidelity motion required to make avatar signing easily legible. Attempts to gradually introduce Perlin noise do not improve the problem, and can introduce jarring discontinuities in the motion.

6. An Insight from Motion Capture Data

A *heatmap* facilitates further exploration the presence of noise in the motion capture data by visualizing the relationship of frequency and amplitude with time in the signal. Figure 11 displays a heatmap of the amplitude/frequency profile over an entire recording session computed using a sliding discrete Fourier transform (Jacobsen and Lyons 2003). In this visualization, the *x*-axis displays the frequency, the *y*-axis displays time, and the amplitude is displayed as a grayscale intensity with darker intensities representing higher amplitudes. The regions labeled in this figure correspond to regions in Figure 7. For frequencies in region 3 that are greater than 12Hz, the noise is nearly constant over the entire time range. This is to be expected since this noise does not come from human movement, but rather from the recording equipment itself.

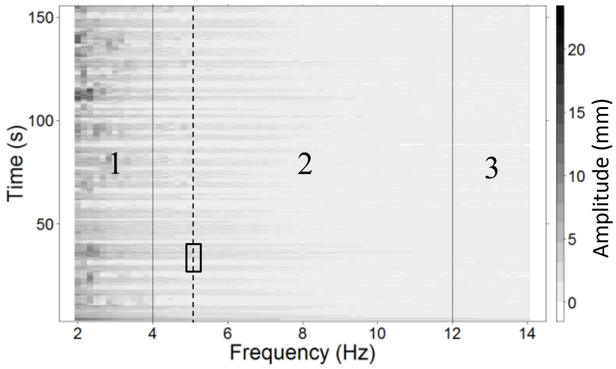


Figure 11: Heatmap of Amplitude vs Frequency and Time for Right Wrist Height.

The frequencies in region 2 are too high to warrant linguistic analysis, but are still produced by a human signer. The heatmap demonstrates that these frequencies are present throughout the entire discourse, whether the signer is producing utterances or is at rest. From a linguist's perspective, this is noise and can safely be ignored, but from an animator's perspective, region 2 frequencies are actually invaluable, as they can be used to enliven the avatar. These data inform us that these frequencies must be present whenever an avatar is signing or is at rest. Observers do not perceive these frequencies as noise during signing, since the frequencies of the signed utterances have comparatively higher amplitude. High-amplitude motions produced by signing overwhelm the subtle changes created by the lower amplitude frequencies from region 2.

To further investigate the relationship between noise and signing, we examine a representative clip of the height of the right wrist marker during two sentences which begin and end with the signer at rest. We will focus on a frequency of 5Hz which lies in the enlivening region of the heat map. A vertical slice of the heatmap at 5Hz, corresponding to the dotted line in Figure 11, can be plotted with time on the x-axis and the amplitude at 5Hz on the y-axis. Figure 12 shows the graph of the portion of this signal corresponding to the small rectangle in Figure 11. Active signing in this segment occurs between times 24 and 30 seconds. The signer is at rest at the onset and conclusion of the segment.

The conventional expectation would be that the amplitudes for this particular frequency should be lower while the person is signing. Yet in this example we find exactly the opposite. Counterintuitive as it is, the enlivening frequencies are not just present, but actually *increase* in amplitude in the center of this graph, during which the signer is actively producing utterances. So, when adding noise to enliven an avatar, we should not suppress or turn off that noise when the avatar is signing. Figures 11 and 12 thus provide additional evidence that we should apply these enlivening frequencies throughout an avatar's signing.

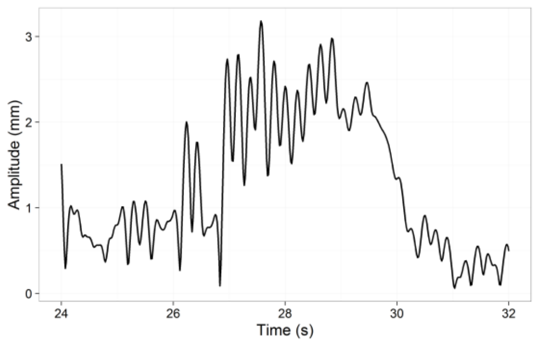


Figure 12: Amplitude of Wrist Height at 5Hz for Two Sentences.

7. Implementation

To add enlivening frequencies to the avatar, we apply Perlin noise generators to each joint using the frequency ranges dictated by region 2. The generators run continually, and independently, of any utterances produced by the avatar. The exception to this is the blinking action of the eyelids. Blinking is a discrete movement that must be controlled with a separate mechanism which is outside the scope of this paper (Baker and Padden 1978) (R. B. Wilbur 1994).

There is one additional consideration required when setting up the Perlin noise generators, as they also require knowledge of amplitude. This information is easily obtainable from the spectrum of each marker and is summarized in Table 2. Because our avatar requires angle data for its joint rotations, we use the fact that $sin(\theta) = \theta$ for small θ to estimate rotational data from positional data.

Perlin noise generators add a modest computational cost, but if the avatar is used in an environment where computing resources are limited, then implementing a single generator on the hips is an effective choice as the hips will transmit subtle motion, albeit coordinated, to the rest of the avatar's skeleton, even in the absence of noise on the other joints. (McDonald, Wolfe and Schnepp, et al. 2015).

Joint	Amplitude (degrees)
Hips	6.37×10^{-3}
Waist	4.78×10^{-3}
Upper spine	4.78×10^{-3}
Neck	2.39×10^{-3}
Shoulders	2.39×10^{-3}
Elbow	2.39×10^{-3}
Wrist	2.39×10^{-3}

Table 2: Amplitudes (noise strengths) for Perlin Noise.

8. Results

To test this approach, we applied Perlin noise generators to all the joints in the avatar's spinal column (hips, waist, upper spine, and neck) and arms (shoulder, elbow, and wrist). More distal joints were given noise with lower amplitudes and higher frequencies as indicated in Tables 1 and 2. The generators are active throughout the entire animation, regardless of whether the avatar is signing or not.

The reference http://tinyurl.com/zzl8btc is a link to a video demonstrating the effect. The video contains a side-by-side comparison of animations with and without Perlin noise generators. The animation on the left has no noise, while the one on the right has noise applied to all joints previously mentioned. When at rest, the figure on the left has the appearance of a static photograph, whereas the figure on the right continues moving subtly. The noise does not interfere with the portrayal of the signed utterances.

This approach is well accepted by test participants who view and rate our avatar's utterances for clarity and naturalness. In a developing a mathematical model for role shift as reported in (Schnepp, et al. 2013), Deaf participants fluent in ASL viewed and rated animations that incorporated this livening method. A majority of the participants rated clarity as either "clear" or "very clear" on a 5-point Likert scale. A follow-up study (McDonald, Wolfe and Moncrief, et al. 2016) yielded similar results. Clarity was a particularly important measure here, because it tested whether noise was interfering with the avatar's signing. The results indicate that applying noise to an avatar's joints, with frequencies and amplitudes appropriately tuned according to the results of the study of motion capture data, are effective in enlivening an avatar without impeding the avatar's ability to communicate.

9. Future work

We look forward to testing the scalability of SignCleaner by applying it on larger corpora. We also plan to use it for its original intended purpose of prosodic analysis. We will also add the ability to import other motion capture formats.

Acknowledgment

We extend our gratitude to Eleni Efthimiou for collegial discussions on the challenges and nature of the application of noise to an avatar system.

10. References

Baker, Charlotte, and Carol Padden. 1978. "Focusing on the Nonmanual Components of American Sign Language." In *Understanding Language through Sign Language Research.*, edited by Patricia Siple, 25-57. New York: Academic Press.

Crasborn, Onno, Han Sloetjes, Eric Auer, and Peter Wittenburg. 2006. "Combining video and numeric data in the analysis of sign languages within the ELAN annotation software." *Proc. LREC 2006 workshop on representation & processing of sign language.* Paris: ELRA. 82-87.

Duhamel, Pierre, and Martin Vetterli. 1990. "Fast Fourier transforms: a tutorial review and a state of the art." *Signal processing* 259-299.

Gleicher, Michael. 1998. "Retargetting motion to new characters." *Proceedings of the 25th annual conference on Computer graphics and interactive techniques* 33-42.

Hong, Jia-Shen G., and Michael J Lancaster. 2004. *Microstrip filters for RF/microwave applications.* Hoboken, NJ: John Wiley & Sons.

Jacobsen, Eric, and Richard Lyons. 2003. "The Sliding DFT." *Signal Processing Magazine, IEEE* 20 (2): 74-80.

Lagae, Ares, Sylvain Lefebvre, Rob Cook, Tony DeRose, George Drettakis, David S. Ebert, J. P. Lewis, Ken Perlin, and Matthias Zwicker. 2010. "A survey of procedural noise functions." *Computer Graphics Forum* 2579-2600.

Marshall, John, and E. Geoffrey Walsh. 1956. "Physiological Tremor." *Journal of Neurology, Neurosurgery, and Psychiatry* 260-267.

McDonald, John, Rosalee Wolfe, Jerry Schnepp, Julie Hochgesang, Diana Jamrozik, Marie Stumbo, Larwan Berke, Melissa Bialek, and Farah Thomas. 2015. "An Automated Technique for Real-Time Production of Lifelike Animations of American Sign Language." *Journal of Universal Access in the Information Society* 1-16.

McDonald, John, Rosalee Wolfe, Robyn Moncrief, and Souad Baowidan. 2016. " A computational model of role shift to support the synthesis of signed language." *Theoretical Issues of Sign Language Research (TISLR).* Melbourne, VIC.

Parent, Rick, D. S. Ebert, D. Gould, M. Gross, C. Kazmier, C. J. Lumsden, R. Keiser, et al. 2009. *Computer animation complete: all-in-one: learn motion capture, characteristic, point-based, and Maya winning techniques.* Burlington, VT: Morgan Kaufman.

Perlin, Ken. 1996. "A system for scripting interactive actors in virtual worlds." *Proceedings of ACM SIGGRAPH 96.* New Orleans: Association for Computing Machinery. 205-216.

Perlin, Ken. 1995. "Real time responsive animation with personality." *IEEE Transactions on Visualization and Computer Graphics* 5-15.

Quinto-Pozos, David. 2010. "Rates of fingerspelling in American Sign Language." *Theoretical Issues in Sign Language Research.* West Lafayette, IN. http://www.purdue.edu/tislr10/pdfs/Quinto-Pozos%20fingerspelling.pdf.

R Core Team. 2000. ftp://155.232.191.133/cran/doc/manuals/r-devel/R-lang.pdf. Accessed 2016-1-18. "R Language Definition."

Schnepp, Jerry, Rosalee Wolfe, John McDonald, and Jorge Toro. 2013. "Generating Co-occurring Facial Nonmanual Signals in Synthesized American Sign Language." *Eight International Conference on Computer Graphics Theory and Applications.* Barcelona. 407-416.

Smith, Steven. 2011. *The Scientist and Engineer's Guide to Digital Signal Processing.* Santa Clara, CA: California Technical Publishing.

Wilbur, Ronnie B. 1994. " Eyeblinks and ASL phrase structure." *Sign Language Studies* 221-240.

Wilbur, Ronnie B., and Evgenia Malaia. In Press. "A new technique for analyzing narrative prosodic effects in sign languages using motion capture technology." In *Linguistic foundations of narration in spoken and sign languages*, edited by A. Huebel & M. Steinbach (eds.). Amsterdam: John Benjamins.

Wilbur, Ronnie, Kadir Gokgoz, Robin Shay, and Aleix Martínez. 2011. "Mechanics and Issues in Making ASL Databases Available." *Building sign language corpora in North America.* Washington, DC. Accessed May 30, 2011. http://aslcorpus.blogspot.com/.

Using Sign Language Corpora as Bilingual Corpora for Data Mining: Contrastive Linguistics and Computer-Assisted Annotation

Laurence Meurant[1], Anthony Cleve[2], Onno Crasborn[3]
[1]F.R.S. - FNRS and University of Namur, [2]University of Namur, [3]Radboud University Nijmegen
[1,2]61, rue de Bruxelles, B-5000 Namur, Belgium; [3]PO Box 9103, NL-6500 HD Nijmegen, The Netherlands
laurence.meurant@unamur.be, anthony.cleve@unamur.be, o.crasborn@let.ru.nl

Abstract

More and more sign languages nowadays are now documented by large scale digital corpora. But exploiting sign language (SL) corpus data remains subject to the time consuming and expensive manual task of annotating. In this paper, we present an ongoing research that aims at testing a new approach to better mine SL data. It relies on the methodology of corpus-based contrastive linguistics, exploiting SL corpora as bilingual corpora. We present and illustrate the main improvements we foresee in developing such an approach: downstream, for the benefit of the linguistic description and the bilingual (signed - spoken) competence of teachers, learners and the users; and upstream, in order to enable the automatisation of the annotation process of sign language data. We also describe the methodology we are using to develop a concordancer able to turn SL corpora into searchable translation corpora, and to derive from it a tool support to annotation.

Keywords: Sign Languages, parallel corpora, annotation automatization

1. Introduction

As more and more sign languages nowadays, the French Belgian Sign Language (LSFB) is now documented by a large scale digital corpus (Meurant, 2015): the Corpus LSFB. This dataset includes around 150 hours of multi-camera recorded data, from which 12 hours are so far annotated with ID-Glosses (Johnston, 2010) (104,000 tokens, from which 98,200 fully lexicalized signs), and 2.5 hours translated into written French (2,400 sentences) and is supplemented by the metadata about the participants and the tasks. An online lexical database contains all the sign types glossed up to now, and serves as a dynamic external controlled vocabulary for the annotation process in ELAN. These data are made available online via a user-friendly web site. The French counterpart of the Corpus LSFB is now being collected: in the same setting and following the same protocol, pairs of French speaking informants are currently videorecorded. The collected data will be transcribed and translated into LSFB. When this work will be completed, we will for the first time benefit of a bidirectional translation corpus between a sign language (SL) and a spoken language (SpL). The Corpus NGT (Crasborn et al., 2008) has been an inspiring model for the Corpus LSFB. It includes NGT video data (72 hours), gloss annotations (150,000 tokens, 3,300 types), sentence-level translations into written Dutch (15 hours, 15,000 sentences), and the lexical database NGT Signbank (including translation equivalents and a detailed phonological description).

This kind of resource is not only essential to the linguistic description of sign languages, but it is also a potential wealth of information for pedagogic purposes, for the field of translation and interpretation studies and for the field of contrastive linguistics between signed and spoken languages. Exploiting corpus data remains subject to the time-consuming and expensive manual task of annotating, i.e. from the ID-Glossing to the analytic annotations. This slow process is unavoidable at this stage and crucial in enlarging the available data set that is needed to automate the annotation process in the near future.

In this paper, we present ongoing research that aims at testing a new approach to exploit SL data. This approach relies on the methodology of corpus-based contrastive linguistics. It exploits the fact that many SL corpora (including the LSFB and the NGT ones) do not only consist of video recordings of SL, but also glosses and translations into spoken language (SpL). In other words, our approach considers SL corpora as bilingual corpora. We present and illustrate the main improvements we foresee in developing the use of sign language corpora within a corpus-based contrastive methodology: downstream, for the benefit of the linguistic description and the bilingual (signed - spoken) competence of teachers, learners and the users; and upstream, in order to enable the automatisation of the annotation process of sign language data.

We first (Section 2.) present the major types of multilingual corpora used for the purpose of contrastive linguistics. Then we explain why SL data can be considered as translation corpora and we show how valuable the combination of translation corpora and comparable corpora would be in the fields of SL linguistics and of SL-SpL contrastive linguistics. We provide an overview of the possible uses of such bilingual data, not only to the benefit of linguists, but also of interpreters, translators, teachers and learners. In Section 3., we show the modelling of the different data sources we are using for the development of a concordancer between SL and SpL data, and we specifically detail the way they interact with one another. Section 4. describes the methodology we are using in order to develop the concordancer, from the challenging alignment of written texts and video recorded productions at the level of the word-sign, to the extraction of semantic equivalents in context, and the way these development are expected to automatically assist the annotation process.

2. Corpus-based contrastive analysis and Sign language data mining

2.1. Multilingual corpora

Beyond the domain of SL linguistics, the computer revolution also impacted the domain of contrastive linguistics in general by having enabled the development of multilingual corpora. Multilingual corpora, combined with alignment and search tools, are today acknowledged for their theoretical as well as practical importance in cross-linguistic studies and applications: they provide a rich basis of language correspondences in context that are able to provide new insights into the languages that are being compared (Altenberg, B. and Granger, 2002; Johansson, 2007). Multilingual corpora are the basis of all multilingual concordancers such as TransSearch (Bourdaillet et al., 2010) or Linguee (Linguee, 2015). Following the terminology of Altenberg, B. and Granger (2002), we will distinguish between translation corpora and comparable corpora, although both sometimes fall under the heading of "parallel corpora" used in a generic sense .

Translation corpora consist of original texts in one language and their translations into one or several other languages. They are unidirectional when the translation goes only in one direction, from the original language A to the target language B. If the translation goes in both directions, that is if each language is both source and target language, they are said bidirectional. Some translation corpora are aligned, which means that each unit of the original text (it can be a paragraph, a sentence, a phrase, or even a word, in the case of written texts) is linked to its corresponding unit in the other language. Such aligned translation corpora are also called "parallel corpora"[1]. The Hansard Corpus is a well known example of parallel bidirectional translation corpus. It has been the first one to be digitized and made available to linguists. It consists of parallel texts in English and Canadian French, drawn from official records of the proceedings of the Canadian Parliament.

Comparable corpora are made of texts in original language only, i.e. non translated ones, that share the same type, subject matter and communicative function. The gathered texts may be restricted to a specific domain (e.g. newspaper articles about football in English and French) or on the contrary they may represent a wide range of text types (e.g. balancing general news with economical, legal, medical, and political texts).

Each kind of multilingual data has its advantages and disadvantages. When using a translation corpus, one can rely on the semantic similarity of the texts in both languages: the objective, the discursive function, the register as well as the audience is typically the same in both version of the texts. But translated texts may always be suspected to reflect the transfers of features from the source language to the target language, or "translationese" (Gellerstam, 1996), and individual variations specific to the translators. On the contrary, the texts contained in a comparable corpus reflect the natural use of language, but it is sometimes difficult to know whether the compared texts are really comparable, for example in terms of register or discursive function. Therefore, the combination of both types appears to be particularly relevant since it eliminates or mitigates the disadvantages and strengthens the advantages of each type.

2.2. Sign language corpora as translation corpora

Due to the visual-gestural nature of SLs, most modern SL machine-readable corpora like the Corpus LSFB and the Corpus NGT are bilingual ones, since the videotaped data are accompanied by the written glosses of the signs and by the translation of the videos in written language. But as far as we know, this bilingual property of SL corpora has not been exploited yet for the development of contrastive linguistics. However, we see contrastive corpus linguistics between a signed and a spoken language as an effective solution to the current difficulty to detect interesting data amongst the sign language corpora.

Most SL corpora can at least be seen as unidirectional translation corpora, which provide a good basis for comparing how a specific meaning (retrieved from the SpL translation, in our cases in French or Dutch) is rendered in SL (LSFB or NGT respectively). If combined with alignment at the level of the sign and word, it would be an efficient means to extract aligned bilingual examples of words and signs in context. For example, a request on the word *même* ('same') within the French translations of the Corpus LSFB would provide the various signs used by the signers and that were translated by *même* (AUSSI, COMMUN, EGAL, MEME-AVANCER, MEME.MAFIA, MEME.REPETITION, MEME.Y, STABLE, but also MOI-MEME, PERSONNE.MOI, VOIR.MOI which have a reflexive meaning ('self') translated by *même*[2]. These signs would be presented in their context of use, that means within the video clip where they appear, and aligned with the corresponding contexts of occurrence in French. And conversely, a request on the sign AUSSI ('also, same') will provide the various words and word constructions used to translate the various tokens of this sign into French (such as *comme* 'as', *disons* 'let's say', *un genre de* 'sort of', *aussi* 'also', *et puis* 'and then'), according to their context of use. This type of information can be harvested from corpora to enrich the current lexical databases of the Corpus LSFB and the Corpus NGT with a classification of the meanings of the signs in context and their frequencies, which in turn will be used to assist the annotation process (see Section 4.).

By comparing sequences of signs or words using a concordancer, it will also be possible to search for the translation equivalents of non-lexical elements, as for example the equivalents of the passive forms of French, or the French and Dutch equivalents of the spatial left vs. right oppositions in LSFB or NGT, the ways LSFB or NGT expresses what is translated by prepositions into French or Dutch, or the ways partly-lexicalized signs (Johnston and Schembri, 2010) of LSFB or NGT are translated into French and Dutch, respectively. In particular, research on discourse can

[1] Henceforth, we will use 'parallel corpora' with this meaning.

[2] The signs corresponding to these glosses written in capitals can be seen on the lexicon part of the LSFB corpus website (`http://http://www.corpus-lsfb.be/lexique.php`)

greatly benefit from this methodology. For example, requests on French or Dutch discourse markers will provide examples in LSFB and in NGT that will illustrate how diverse the SL expressions of the equivalents of these markers are: Do SLs use discourse markers as equivalents, or other lexical and/or non-lexical resources, articulated manually and/or non-manually?
In their present state, both the LSFB and the NGT corpora (blocks A and B in figure 1) are unidirectional (relations 1 and 2 in figure 1): the SL original productions are translated into written French and Dutch respectively. We are currently testing the feasibility of building the counterpart of the LSFB corpus (block C in the figure), which means videorecording spoken French data and translating them into LSFB. In this way, we will be able to count on a bidirectional translation corpus between a SL and a SpL.

2.3. Towards comparable corpora

In building the French counterpart of the LSFB corpus, our data gain various additional dimensions, and especially the possibility to compare original LSFB productions and original French ones (relations 4 in figure 1). The French data are elicited in the same conditions as were the LSFB data, and following the same protocol. The informants are invited by pairs in the LSFB-Lab studio. A French-speaking moderator is leading them through the same tasks as the one used for the LSFB corpus.
The content of the tasks were minimally adapted to fit to the hearing and Belgian French culture of the informants, but the dialogic setting as well as the discourse genre of each task have been preserved, which make the French and the LSFB productions closely comparable.
Together, the LSFB corpus and its French counterpart (B and C in figure 1) provides a rich variety of possible comparisons (referred by the numbers of the arrows in the figure):

1. Comparison of original discourses in one language and their translation in the other one (relations 2 and 3);
2. Comparison of original discourses in both languages (bidirectional relation 4);
3. Comparison of original and translated texts in the same language (relations 5 and 6).

These three types of corpus-based comparisons have an heuristic power in the sense they offer the opportunity to discover features of the languages in contrast that could not be expected without the automatic comparison of large amounts of parallel data (Altenberg, B. and Granger, 2002; Gilquin, 2000). It is the reason why we are testing the feasibility of building such combination of translation and comparable corpora and its efficiency for the issue of corpus mining.
When it will be possible to link these bilingual corpora to data from LSFB and French learners, the *Constrastive Interlanguage Analysis* method (Granger, 1996) could be used to better understand the specific difficulties of LSFB signers learning French and of French speakers learning LSFB.

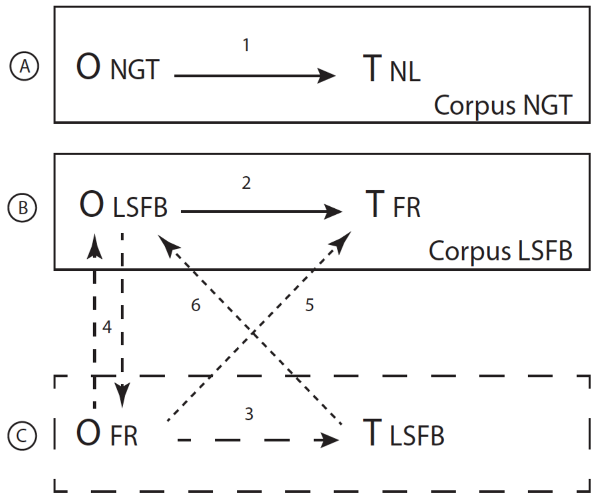


Figure 1: Corpora at our disposal (continuous line) and under construction (dotted line). O = original texts, T = translated texts. Both blocks A and B constitute translation corpora. Together, the original texts of B and C constitute comparable corpora.

2.4. Beyond the linguists' needs

Once these bilingual data gathered (the translation corpora or even the comparable ones) and the concordancer developed, linguists will take advantage of novel and effective means to detect interesting data within sign language corpora. But the resources that we are building for corpus-based contrastive analysis for the purpose of SL research also will benefit other kinds of users, ranging from interpreters, translation and interpreting trainers, and teachers for the deaf, to indeed all signing learners of French/Dutch and French/Dutch speaking learners of LSFB/NGT. A searchable database of aligned bilingual examples of language in use will constitute a useful resource for expanding one's knowledge of a second language and increasing one's level of bilingualism. It can be used to assist a wide range of tasks, among which the comprehension of LSFB/NGT or French/Dutch texts, the production of LSFB/NGT or French/Dutch texts, as well as the translation between LSFB/NGT and French/Dutch (in both directions).
When it comes especially to deaf pupils who learn a spoken language, these bilingual data can be seen as an efficient tool to support their learning and to foster their autonomy in the use of the spoken language, just as *TransSearch* or *Linguee* are supporting the speakers of one language who learn another one, at any level. For example, a deaf learner of French or Dutch may discover the variety of meaning of signs and above all learn to distinguish the various meanings with accuracy thanks to the signed equivalents at her/his disposal for each written example. She/he may also be helped in their use of some idiomatic features like gender of names, prepositions, *avoir/être* auxiliaries in French, etc.
Figure 2 shows a mock-up user interface of the tool that will be derived from the aligned bilingual data (in its LSFB-

French version). The terms between asterisks refer to the entities' names used in Figure 3.

3. Modelling the data resources

The combination of the Corpus LSFB and the Corpus NGT provides a relatively (i.e. for the present time) significant size of exploitable data which will be involved in the machine learning process underlying the development and the exploitation of our parallel concordancer.

Figure 3 provides a simplified "helicopter-view" of the various data artifacts that are involved in the creation of our multilingual and multimodal corpus-based concordancer. This model represents the main concepts involved in the data and tools, as well as their characteristics and relationships.

The figure shows the parallelism between the components available from the LSFB corpus (at the top), and the ones provided by the NGT corpus (at the bottom). The new data set under construction (in dotted line in Figure 1 and in grey in Figure 3) can be seen as a mirror of the existing Corpus LSFB: spoken French data (we foresee 40 hours of video), their transcription and translation into LSFB, as well as the annotation of the translations.

Each corpus consists of a set of videos (LSFB_VIDEO and NGT_VIDEO) where two signers achieved a task. Each video is identified by a unique ID, corresponding to its Unique Resource Identifier (uri), and is characterised by the duration of the video (Duration), and a brief description of the task (Task_Description).

Each corpus also includes a large set of signs (LSFB_SIGN and NGT_SIGN). For the LSFB corpus, this set of signs corresponds to the Lex-LSFB lexical database; for the NGT corpus it corresponds to the NGT Signbank. In both corpora each sign is characterised by a unique ID-gloss (ID_Gloss), and is linked to a set of keywords (FR_KEYWORD or NGT_KEYWORD) that represent (some of) the different possible meanings of the sign[3]. Note that the NGT Signbank also includes extra information about the signs, such as phonological descriptions, that are not depicted in Figure 3.

The occurrence of a given sign in a video is represented through an entity type SIGN_ANNOTATION. An annotation indicates the exact time period during which the sign appears in the video, in the form of a time interval (Begin and End). Note that when the same sign S occurs N times in the very same video V, there are N annotations linking S and V in the corpus, each with a distinct time interval. The annotation also records which of the two signers is the author of the sign, via attribute Turn.

As mentioned above, the corpus also provides, for a subset of the videos, the full French/Dutch translation (FR_TRANSLATION and NL_TRANSLATION) of the task. Each translation is made up of a set of translation fragments (FR_TRANSLATION_FRAGMENT and NL_TRANSLATION_FRAGMENT), that is a French/Dutch text fragment (Text) translating what is expressed in the respective sign language by one of the two signers (Turn) during time interval [Begin, End] of the video.

External tools and related data resources are also available. The CoBRA (Corpus Based Reading Assistant) tool (Deville et al., 2013) is based on bilingual corpora (Dutch-French and English-French) aligned at the level of the sentence, and allows the teachers to create labelled texts in Dutch (NL) or in English (EN) and French-speaking learners to be assisted in their reading by clicking on any word in order to know its meaning in its particular context of occurrence. CoBRA is based on a searchable concordancer, called the "Dico Corpus" tool, and on two bilingual dictionaries (FR-NL and FR-EN) called "DiCoBRA" that are (1) produced from a contrastive approach of the existing dictionaries of each language and (2) completed by the contrastive data provided by "Dico Corpus".

The CoBRA resources (CoBRA Corpus) currently include a global text corpus of over 30,000,000 words among which circa 15,000,000 French words, about 10,000,000 concordances (i.e. aligned bilingual examples), an English-French glossary of about 19,000 entries, and a Dutch-French glossary of about 20,000 entries. CoBRA's dictionary (DiCoBRA) includes circa 87,000 lemmas and 300,000 inflected forms of French.

4. Automated support for annotation

On the basis of the available data described above, we are currently developing a concordancer in order to exploit the LSFB and the NGT corpora as aligned (at the level of the sign and word) and searchable translation corpora (LSFB-French and NGT-Dutch). While doing this, we also aim at providing support to the annotation process of the lexical part of the signed data that have not been annotated to date. The methodology used is organised in three steps, the first one being in progress: building the alignment tool, extracting semantic equivalents from the annotated and translated data, and eventually developing the tool support for the annotation process.

4.1. Alignment

In order for the translation corpora to be exploited, they need to be aligned. This means that each unit from one language must be linked to its corresponding unit in the other language. Translation corpora of written texts can be aligned at the level of the paragraph, at the level of the sentence, or even at the level of the phrase or the word. The automatic alignment sentence by sentence is the most common. The matching between a sentence in the source language and a sentence in the target language is based on statistics exploiting information about typographical features (capitals and punctuation marks), the length of the sentence, and cognate words (Altenberg, B. and Granger (2002), p. 10).

In the case of SL data, the alignment cannot rely on any typographical feature. And, as has been extensively shown, the identification of sentences in SL remains a difficult task (Crasborn, O. (2007), Fenlon et al. (2007), Ormel and Crasborn (2012), Börstell et al. (2014), to name a few). Therefore, we decided to avoid investing in segmenting the data

[3] Within the NGT data, those meanings in context are retrieved from a specific tear from the annotations files where a word-level translation equivalent is created for every sign.

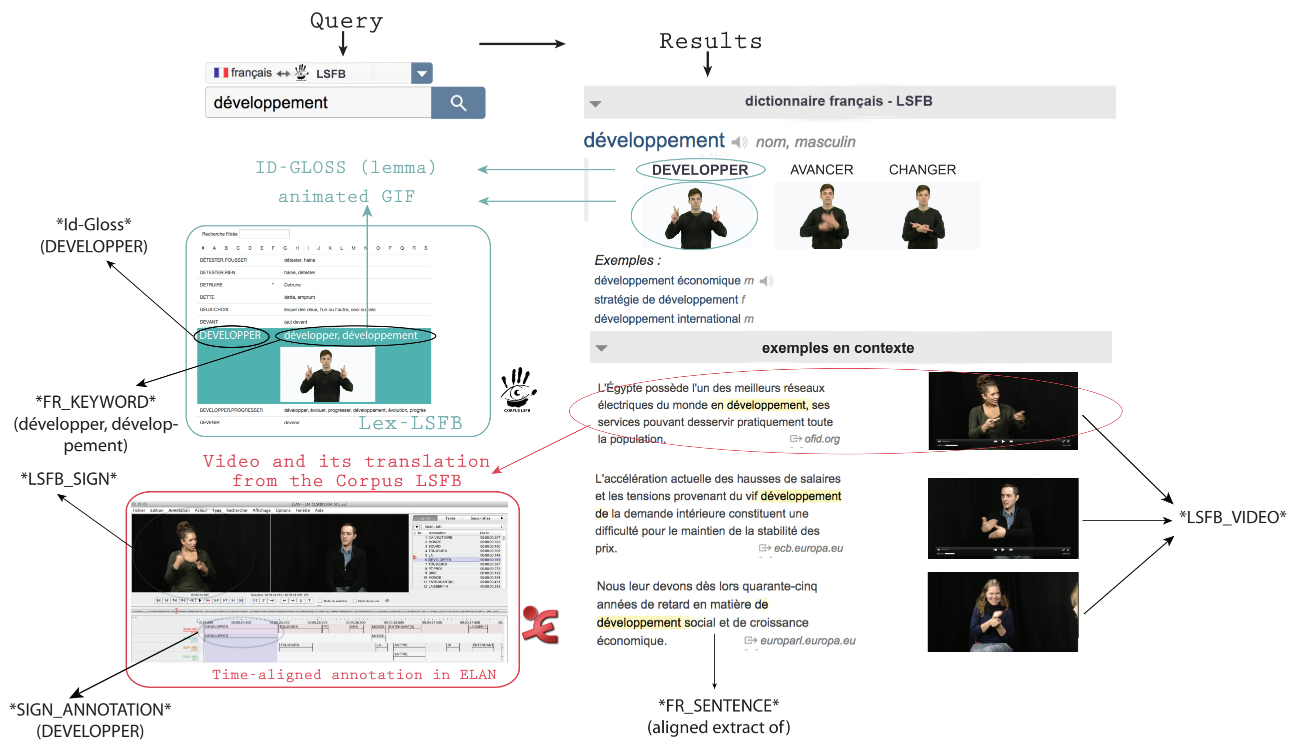


Figure 2: Model of the possible user interface of a bilingual tool (LSFB-French) derived from the parallel data. This figure is based on fictive examples and inspired by the Linguee user interface. The terms between asterisks refer to the entity names used in Figure 3

into sentences, and to establish the automatic alignment between the SL data (LSFB and NGT) and the written translation (French and Dutch respectively) based on other cues. We are currently working with a set of four kinds of information.

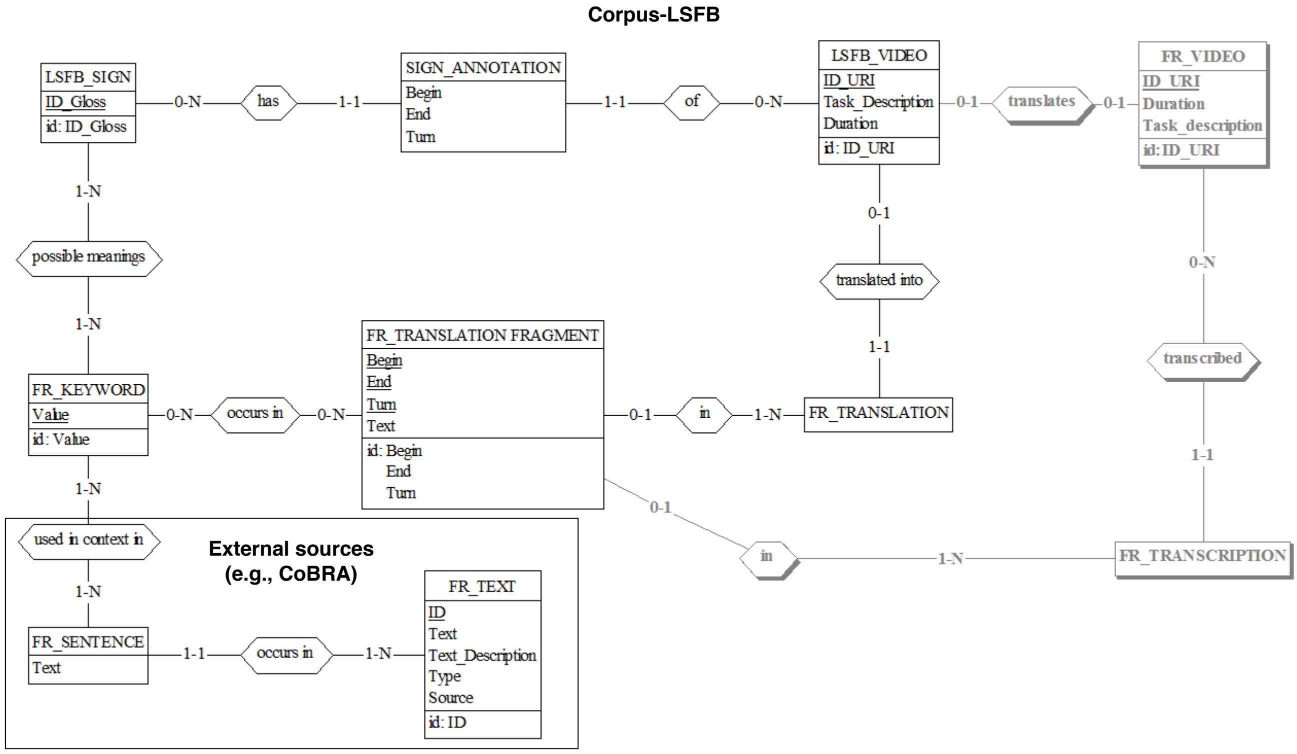
First of all, we can rely on the existing alignment created at the time of the translation process. Indeed, the peculiarity of our translation corpora consists in the fact that in both the LSFB and the NGT corpora, the translations have been encoded in ELAN, and thus time-aligned on the video data in the same way as the glosses are. In the LSFB corpus, the translators themselves were not asked to align their text to the video: the minimal unit they had to take into account was the turn. Afterwards, the alignment in ELAN was made by combining the segmentation of the translator into French sentences or paragraphs, the thematic coherence of the discourse, and finally the pragmatic display constraint of not making the translation segments too long to read in the website of the corpus. As for the NGT corpus, the translators directly entered their Dutch text in ELAN, aligning it at the level of the sentence-like unit in NGT. In any case, both corpora already provide a first alignment between the SL data and the written translations: in the LSFB corpus at the level of a paragraph-like group of French sentences, and in the NGT corpus at the level of the sentence. The issue lies in narrowing the scope of this existing alignment. Three other elements are exploited for this purpose.

Second, anchor signs and words are identified within the existing segments. A sign-word pair is considered as anchor when, in the available data, the sign and the word are strongly related or, in other words, have a high level of translation correspondence. Examples of such signs-word pairs are expected to be found amongst the numbers, the manually spelled words, or the colours signs, to name a few.

Third, the identification of the anchors pairs is supported by the semantic information provided manually during the annotation process. In the LSFB corpus, each entry of the lexical database has been provided with a list of possible meanings of the sign in French. In the NGT corpus, the meaning of each sign in context is specified for each sign within a dedicated tier of the ELAN file. This semantic content can be considered as starting bilingual lexicons for each pair of languages. In addition, this information constitute a first indicator of whether a sign is a good candidate to function as an anchor.

Fourth, we extracted for each gloss of the lexical databases a list of the preceding and following context of the sign. The result is a list of collocations for each entry, i.e. a list of the common sign combinations harvested from the data. Doing the same for each lemma of the translations, and linking the lists of each pair of contrasted languages (LSFB-French or NGT-Dutch), we expect to identify potential translation blocks, and thus refine the localisation of the semantic equivalents from the one language to the other, and finally improve the automatic alignment of the bilingual data.

The efficiency of the combination of these four resources for the alignment of the data will be tested pretty soon.



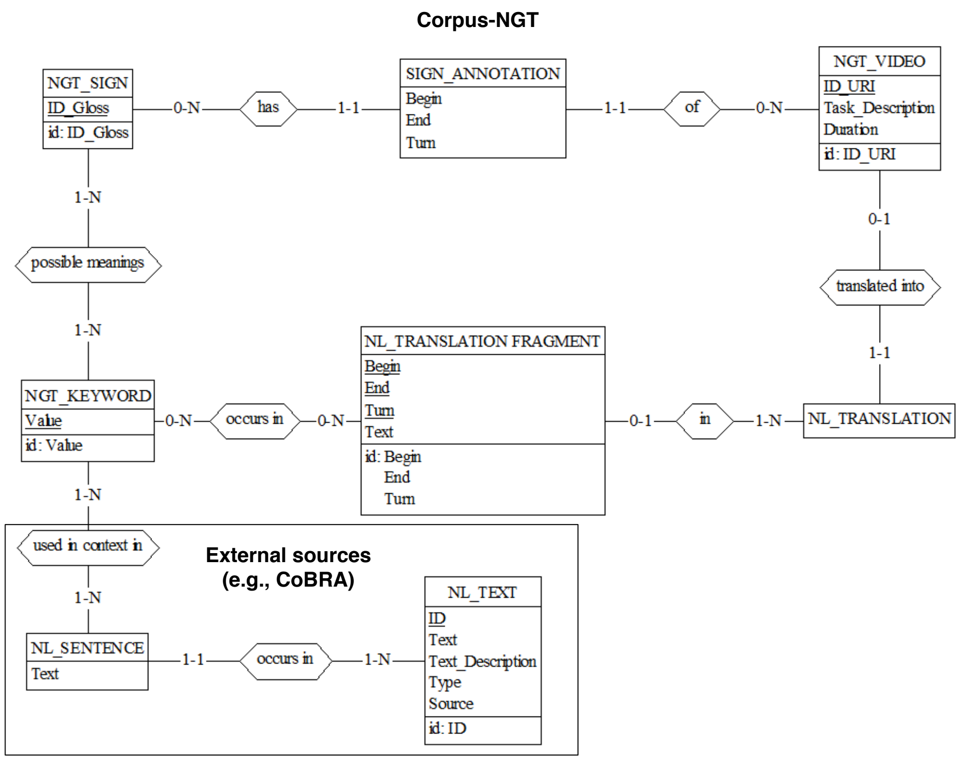


Figure 3: Entity-Relationship model of the data at our disposal. Black elements represent already available data. Light grey elements represent work in progress.

4.2. Extraction of semantic equivalents

Thanks to the alignment set up for the available data (104,000 tokens of LSFB and 2,400 sentences of French translation; 150,000 tokens of NGT and 14,000 sentences of Dutch translation), the second step will be to extract for each sign, starting with the more frequent ones, a series of parallel examples of use from the videotaped data and their written translation (see Section 2.2.). Among those examples of semantic equivalents, we will manually distinguish when the different meanings of a single sign in context occur. This task will be done with the help of the translations, but also with the help of the external data provided by the CoBRA (Corpus Based Reading Assistant) tool (Deville et al., 2013) (see Section 3.). CoBRA will be a source of additional examples of the French and Dutch words in use, and even more importantly, the DiCoBRA dictionaries that includes semantic information about French and Dutch words[4].

Harvesting this semantic information on the meanings of signs in context will supplement the starting data currently available in the lexical database of each corpus. In turn, the semantic equivalents and the information about the polysemy of the signs will be re-injected for the automatic alignment learning, together with the resources presented in Section 4.1. And, last but not least, the outcome of this second step is aimed to be exploited in order to provide an assistance the the annotators of SL data while annotating.

4.3. Towards a tool support to annotation

The task of the annotator should be assisted in different ways by the outputs of the preceding steps of the development of the concordancer. Having learned from the available aligned data, the program will be able to invite the annotator, when she/he chooses a gloss, to select a meaning amongst suggested options. Suggestions will also be improved by the collocations tables established as presented in Section 4.1. As the annotation progresses (within the phrase, then the clause or even the thematic chunk), taking into account the collocations of the annotations and of the selected meanings, the suggestions are expected to be more and more accurate. Eventually, we plan to investigate the efficiency of this "meaning in context" database and the associated list of anchor signs-words pairs and collocations for the purpose of suggesting annotations based on a previously translated text of the sign language data.

5. Perspectives

Now that more and more SL corpora are created, linguists face the challenge of mining interesting data amongst the large amount of the collected video files and the patiently accumulated annotations. This paper suggests directions towards the use of translated SL corpora as parallel corpora, and indicates how they could be exploited as a way to speed up the annotation process, but also as an insightful probe to get new insights into SLs and on the comparison between SLs and spoken languages in their written form. We propose to combine statistical and machine learning approaches to the manual work of annotators and translators. Our motivation for this bilingual approach to SL corpus linguistics eventually has the benefit of supporting the development of deaf learners and deaf users of written language, and will stimulate SL-SpL translation and interpreting studies.

6. Acknowledgements

Meurant and Cleve are supported by the University of Namur, within the framework of its research incentive program. Crasborn is supported by NWO grant 360.70.500 'Form-Meaning Units'.

7. Bibliographical References

Altenberg, B. et al. (Eds.). (2002). *Lexis in contrast: corpus-based approaches*. John Benjamins Publishing.

Börstell, C., Mesch, J., and Wallin, L. (2014). Segmenting the Swedish Sign Language corpus: On the possibilities of using visual cues as a basis for syntactic segmentation. In Crasborn, O. et al. (Eds.), *Beyond the manual channel. 6th Workshop on the Representation and Processing of Sign Languages*, pp. 7–10, Reykjavik. ELRA.

Bourdaillet, J., Huet, S., Langlais, P., and Lapalme, G. (2010). TransSearch: from a bilingual concordancer to a translation finder. *Machine Translation*, 24:241–271.

Crasborn, O. (Ed.). (2007). *Identifying sentences in signed languages*, volume 10-2 of *Special issue of Sign Language and Linguistics*. John Benjamins, Amsterdam; Philadelphia.

Deville, G., Dumortier, L., Meurisse, J.-R., and Miceli, M. (2013). Ressources lexicales: Contenu, construction, utilisation, évaluation. In Gala, N. et al. (Eds.), *Ressources lexicales pour l´aide à l´apprentissage des langues*, volume 30, pp. 291–312. John Benjamins.

Fenlon, J., Denmark, T., Campbell, R., and Woll, B. (2007). Seeing sentence boundaries. In Crasborn, O. (Ed.), *Special issue of Sign Language & Linguistics*, volume 10-2, pp. 177–200. John Benjamins, Amsterdam; Philadelphia.

Gellerstam, M. (1996). Translations as a source for cross-linguistic studies. *Lund studies in English*, 88:53–62.

Gilquin, G. (2000). The integrated contrastive model: Spicing up your data. *Languages in Contrast*, 3:95–123.

Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In Aijmer, K. et al. (Eds.), *Languages in Contrast. Text-based cross-linguistic studies*, volume 88 of *Lund Studies in English*, pp. 37–51. Lund University Press.

Johansson, S. (2007). Seeing through multilingual corpora. *Language and Computers*, 62:51–71.

Johnston, T. and Schembri, A. (2010). Variation, lexicalization and grammaticalization in signed languages. *Langage et société*, 131:19–35.

[4]This semantic information has been produced from a contrastive approach of the existing dictionaries of each language and completed by the contrastive data provided by the concordancer the authors called 'Dico Corpus'. In other words, DICoBRA corresponds, for French and Dutch words, to the outcome we aim for LSFB and NGT signs: a corpus-based dictionary of signs in context build on parallel data.

Johnston, T. (2010). From archive to corpus: transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, 15:106–131.

Linguee. (2015). `http://www.linguee.com`. [accessed 25-10-2015].

Ormel, E. and Crasborn, O. (2012). Prosodic correlates of sentences in signed languages: A Literature Review and Suggestions for New Types of Studies. *Sign Language Studies*, 12(2):109–145.

8. Language Resource References

Crasborn, O. and Zwitserlood, I. and Ros, J. (2008). *The Corpus NGT. An open access digital corpus of movies with annotations of Sign Language of the Netherlands*. Centre for Language Studies, Radboud University Nijmegen, ISLRN 175-346-174-413-3.

Meurant, L. (2015). *Corpus LSFB. First digital open access corpus of movies and annotations of French Belgian Sign Language (LSFB). Laboratoire de langue des signes de Belgique francophone (LSFB-Lab).* FRS-F.N.R.S et Université de Namur.

The French Belgian Sign Language Corpus
A User-Friendly Searchable Online Corpus

Laurence Meurant, Aurélie Sinte, Eric Bernagou
FRS-FNRS, University of Namur
Rue de Bruxelles, 61 - 5000 Namur - Belgium
laurence.meurant@unamur.be, aurelie.sinte@unamur.be, eric.bernagou@unamur.be

Abstract

This paper presents the first large-scale corpus of French Belgian Sign Language (LSFB) available via an open access website (www.corpus-lsfb.be). Visitors can search within the data and the metadata. Various tools allow the users to find sign language video clips by searching through the annotations and the lexical database, and to filter the data by signer, by region, by task or by keyword. The website includes a lexicon linked to an online LSFB dictionary.

Keywords: French Belgian Sign Language, searchable corpus, lexical database.

1. The LSFB corpus

1.1. The project

In Brussels and Wallonia, i.e. the French-speaking part of Belgium, significant advances have recently been made of the development of LSFB. It was officially recognised in 2003 by the Parliament of the Communauté française de Belgique. Since 2000, a bilingual (LSFB-French) education programme has been developed in Namur that includes deaf pupils within ordinary classes (Ghesquière et al., 2015; Ghesquière et Meurant, 2016). The first MA in Translation and Interpreting in LSFB-French opened in September 2014. And since the early 2000s, linguistic research has been conducted at the University of Namur. But as is the case for most sign languages, French Belgian Sign Language (LSFB) remains a less-resourced language. Until very recently, only small sets of recordings existed, most of which were private archives. A large-scale searchable corpus of videotaped data, documented by metadata about the signers and the tasks produced was direly needed.

In December 2015, the LSFB corpus website was launched, containing the results of the *Corpus LSFB* project conducted at the University of Namur between 2012 and 2015. The first aim of the project was to collect data for linguistic research on LSFB. However, we kept a close watch to ensure that the LSFB Corpus was also a useful tool for teachers, interpreters and students, and that it safeguards the cultural and linguistic heritage of the (French Belgian) Deaf Community. The corpus is available as an open access website[1] containing video data, annotations, translations and metadata. Several search options allow users to browse the data.

1.2. The setting and the technical equipment

The video recordings were collected in the fully equipped studio of the LSFB-Laboratory[2]. The participants were invited to come in pairs, and were guided by a deaf moderator. All were seated on chairs without armrests. Three *JVC Pro HD 3 CCD* cameras recorded the participants: one for an upper body view of each informant (Cam 1 and 2 in Figure 1), and one for a wide-shot of both of them (Cam 3 in Figure 1). Additionally, a *Sony DV Handycam* was used to record the moderator (Cam 4 in Figure 1). The positions of the participants and the cameras are illustrated in Figure 1.

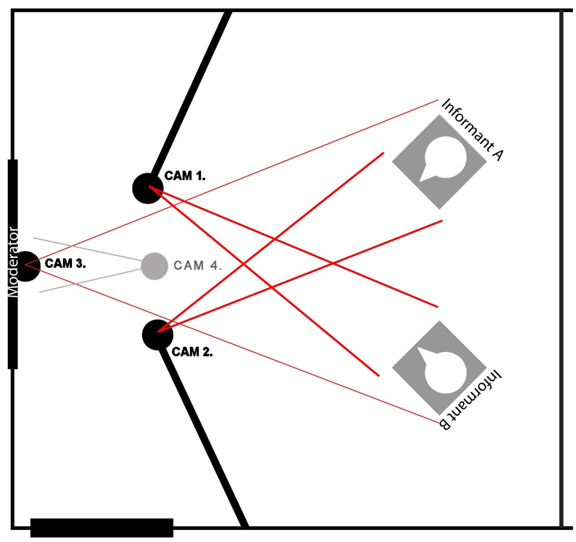


Figure 1: Positions of the participants and the cameras

1.3. The recording and the editing of the data

The data were recorded in full HD resolution (1920x1980 pixels), at 50 frames per second. Then they were edited and compressed with the software EDIUS. The edited files were exported in two sizes, both in .mp4 at 50 frames per second: in 1920x1080 pixels and in 720x576 pixels. These compression formats appeared to be the most convenient for using the videos in ELAN[3] either on PC or on Mac. The video files are cut and named according to the recording sessions (1 to 50), the tasks (1 to 19) and the camera shots (B: upper body shots, L: wide shot, M: moderator shot). In

[1] https://www.corpus-lsfb.be

[2] https://www.unamur.be/lettres/romanes/lsfb-lab

[3] http://tla.mpi.nl/tools/tla-tools/elan/

other words, one video file refers to a whole task performed by the two participants of a particular session.

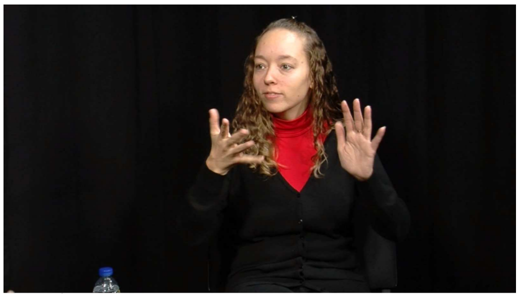
During each session and each task, the cameras recorded all the exchanges: the explanations and instructions given by the moderator before the task and at times during the task, for example if the participants asked a question or needed more explanations, the times when the participants look at a piece of material, memorise it and prepare their production, as well as the exchanges between the two participants performing the task. However, at the editing stage we separated the exchanges between the participants that we considered as the ‘answer’ part of the data from the rest of the productions, called the ‘questions’. For each task, the video files are organised in one question file (Q) and three synchronised answer files (A).

- The Q file shows the four shots at the same time, i.e. the two upper body views, the wide shot on both participants and the moderator view, corresponding to the ‘questions’ sequences. The Q file is comprised of the succession of initial instructions and explanations, followed by, in chronological order, all the sequences not considered part of the data, namely the moderator’s interventions and the participants preparations. The succession of these sequences is signalled by a change in the colour of the cross shape that separates the four shots on the screen. The colours are always used in the same order: black, red, yellow, green, orange, turquoise, blue and white.
- The A files contain the actual exchanges: the participants sign to each other while looking at each other. One A file corresponds to one camera shot: L is the wide shot, M is the moderator shot, S00A-B is the upper body shot showing participant 00A and S00B-B is the upper body shot showing participant 00B. When the exchanges between the participants are interrupted, these interruptions are signalled by a fully-coloured screen. The colours used and their order correspond to what has been established for the Q files. In this way, it is possible to link the different sequences of the Q file with their context in the A files of the same task.

Q file (with black cross between the 4 views)

A file, L shot

A file, S00A-B shot



A file, S00B-B shot

Figure 2: This figure shows the appearance of the question file and 3 answer files

1.4. The signers

The objective of the Corpus LSFB project was to collect a representative sample of the LSFB signs currently in use in Brussels and Wallonia. Yet due to the demographic features of deafness, only 5% of the signers are native signers, and amongst them only a small number have parents who are native signers themselves (Van Herreweghe and Vermeerbergen, 2012). A great number of the signers have acquired LSFB during the first years of their schooling in deaf schools; we considered them as near-native signers. For others, LSFB became their everyday language after age 7 or even during adolescence; we considered them as late signers. The LSFB corpus includes all these three profiles: 30% are native signers, 26% are near-native and 44% are late signers. Other variables related to the signers are represented in the data: the regional variants, the variants related to the gender and to the age of the signers. 57% of the signers are women and 43% are men. They represent four age groups: 18-25 (17%), 26-45 (49%), 46-65 (18%) and 66 + (16%), and they range from 18 to 95 years old.

100 signers participated in the data collection out of a total of an estimated number of 4,000 signers in the French-speaking part of Belgium. When it came to setting up the pairs of signers for each recording session, we used the following criteria in descending hierarchical order: the similarity in terms of linguistic profile, in terms of regional variant, in terms of age and in terms of gender.

Before starting the recordings, the moderator presented the issues related to their participation and the recording of their image to the participants. Then, they were invited to sign an informed consent form and to give their agreement for the use of their data for three purposes and the related type of distribution: for research, for education and training, and for the conservation of linguistic and cultural heritage. At this point, the participants were informed that they would be allowed to confirm or restrict their agreement afterwards, namely after having viewed their productions. Indeed, after the recording, each participant received a DVD with the video files in which they appeared. Each one was asked to confirm their agreement or to specify (with time codes) the clips they wanted to censure. Only a small proportion of the participants asked for some changes, and most accepted the open access distribution of their productions for the benefit of conserving LSFB heritage.

1.5. The tasks

The 100 signers were invited in pairs which means that, from 2013 to 2015, 50 sessions were recorded. Each session lasted between 4h and 6h, for a mean length of 3 hours of edited video per session. The signers were asked by a deaf moderator to perform 19 tasks: telling stories, memories and jokes; explaining maps, routes and pictures; explaining their name sign, their hobby; comparing, arguing, classifying objects and symbols; talking about sign language and the deaf community. The questions the participants were asked were in part inspired by corpus projects from other sign languages such as Australian Sign Language (Auslan), Sign Language of the Netherlands (NGT), German Sign Language (DGS) and Flemish Sign Language (VGT). Several tasks are quite similar to those projects in order to make comparative studies possible. For example, tasks number 2 and 3 (see Figure 1.5.) are widespread among the other Sign Language corpora, but in particular in the VGT and the DGS corpora. Task number 8 contains productions based on the same material (a village map) as tasks from the VGT and the DGS corpora. Tasks number 11 and 12 include productions from the famous ‘Horse’ and ‘Frog’ stories.

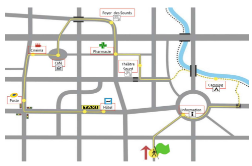
The tasks cover various genres such as narratives, explanations, descriptions, argumentations and discussions. The signers are invited to talk about the deaf community, in order to document some of its specific features: the first encounter of a deaf adult, important family celebrations, school life, relationships with hearing people etc., and about a variety of non-deaf issues and topics, in order to provide a wide range of lexicon. Figure 1.5. describes the topic of each task and figure 1.5. provides an overview of some pictures used in order to support the dialogues.

1.6. Annotations and translations

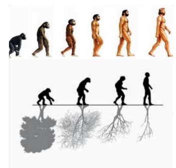
One ELAN annotation file (.eaf) has been created for each task and linked to the four synchronised video answer files, namely to the four available views on the participants (the three illustrated in figure 2, and the moderator shot). The annotation process was carried out by deaf annotators in ELAN (Sinte et al., 2015), on the basis of the Auslan Corpus Annotation Guidelines (Johnston 2015). More than 12 hours of videos are now annotated, sign by sign and hand by hand[4], with ID-glosses (Johnston 2010). This means that all the phonological and morphological variants of a lexeme have been subsumed under the same gloss or lemma. At the time of writing this paper, 98,200 tokens out of the 104,000 annotated tokens are fully lexicalised signs. In this first step of the annotation process, the partially lexicalised signs have only been identified as such by the label ‘DS’ which stands for ‘depicting sign’ (Johnston 2015). In the ongoing second phase of annotation, a short semantic description of the partially lexicalised signs is

1	Information to complete metadata files (age, school, family, etc.) (Because of confidentiality, this task is not available on the website.)
2	Explaining the name sign of both signers
3	Telling a childhood memory
4	Explaining the benefits and disadvantages of being deaf or hearing
5	Explaining what’signing well’means
6	Talking about the influence of emotions on sign language
7	Describing a procedure such as assembly instruction of a piece of furniture or a recipe
8	Describing a map or a route from any given starting point to a destination
9	Explaining a picture (what is special about it or what issues does it raise)
10	Arguing about polemic or shocking topics (general subjects such as verbal abuse, anorexia or gay marriage)
11	Telling a short story: joke, comic strip or short cartoon
12	Telling a long story: Where are you frog? or Paperman (cartoon)
13	Playing a role-playing game: ’Imagine you meet a minister and you have to convince him of... (deaf community topics)’
14	Talking about LSFB variations:’Do you easilyunderstand young/old signers, interpreters, signers from other regions? ’What are the differences?’
15	Talking about a hobby or a job, the material used, the way to proceed, rules, etc.
16	Description of drawn faces
17	Classifying pictures and explaining criteria
18	Explaining differences and similarities of objects and tools
19	Conclusion: discussing the activities of the day, the tasks, etc. with the moderator

Figure 3: Descriptions of the tasks

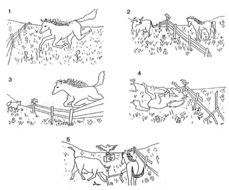


Task 8



Task 9

Task 10



Task 11

Figure 4: Sample of material

[4]Separate annotation tiers have been created for the right hand and for the left hand of each signer.

added within each annotation. The annotation files (.eaf) are linked and time aligned to the corpus website so that the annotations are directly visible by any user. Figure 5 shows the website interface of the video viewers and the annotations.

Among the 12 hours of annotated data, 2.5 hours have been translated so far (2,400 sentences) by a multidisciplinary team. Interpreters, linguists, bilingual teachers and deaf people took part in a workshop we organised at the University of Namur, from March to July 2015. In collaboration with Alain Bacci (Interpretis, Toulouse), who was invited as an expert for the whole workshop[5], we established the main keystones for the translation project. The translations are target oriented: the text is produced in the most natural French possible, ensuring the French lexicon and syntactic structure reflects the influence of LSFB as little as possible. The oral features that characterise the LSFB semi-spontaneous conversations have been translated into French text in the same way as magazines do when transcribing an interview. The translation process includes three main stages. First, the translators work alone, regularly requesting the help of a deaf LSFB expert in order to clarify any point of uncertainty. Second, a referee checks the French translation and suggests corrections or modifications to the translator if needed. And third, after the modifications have been inserted, the text is validated. To date, all the available translations have been validated by Alain Bacci. Within the Corpus LSFB website, the translations may be shown at the same time as the video is playing.

Apart from this basic annotation of the data, i.e. the ID-glosses and the French translations, a great number of annotations are currently being added, mostly in connection with ongoing PhD theses that are using the data. These annotations include:

- a segmentation into Basic Discourse Units (Degand and Simon, 2005; Gabarró-López and Meurant, this volume);
- the identification and qualification of buoys and discourse markers (Gabarró-López, 2015; Gabarró-López and Meurant, 2016);
- the identification of fluency and disfluency markers (Gotz, 2013) as pauses and holds, palm-up signs, repetitions, eye gaze directions, etc. (Notarrigo and Meurant, 2016);
- the identification of phonetic phenomena such as weak hand lowering in symmetrical signs and lowering of forehand signs (Paligot et al., 2016; Paligot and Meurant, 2016);
- the identification and qualification of rephrasing structures and their relation to the original phrasing (Meurant and Sinte, 2016).

These annotations will be made available for researchers on the website within the ELAN annotation files of the concerned videos.

2. The lexical database

At the beginning of the annotation process in ELAN, the annotators needed an online lexical database to gather the ID-glosses encountered in the data, to associate each ID-gloss with a video of the sign, and to share them with the other annotators of the team. A specific web tool was developed in order to solve this issue: the Lex-LSFB.

2.1. Lex-LSFB

The tool was developed in the PHP language and the MySql system, both being very popular and designed for web applications. Thanks to the Lex-LSFB tool, the annotators could (and still do) add or edit glosses in the database and add meta-information about each entry, such as an animated gif file which shows the sign in a video-like form for correct identification, and the possible phonological variants of the sign.

In order to link the web application to ELAN, we only had to export the lexical database as an XML file with the same features as the 'external controlled vocabulary' files processed by ELAN (the structure was copied from a controlled vocabulary created in ELAN). The exported file is referred to by an URL which is copied in the ad hoc field of the 'external controlled vocabulary' in ELAN. The controlled vocabulary is updated at every start of ELAN. If the .ecv file is updated, users must restart ELAN so that the program takes the last update into account. This tool makes the annotation work easier for the annotation team. We can also guess that it significantly improved the quality and the reliability of the annotation work.

This lexical tool is now available on the corpus website to any visitor without any hard installation and without any system requirement. It takes the form of a searchable directory containing all the glosses used in the annotation files, accompanied by keywords, i.e. possible French translations of the sign and by a video-like view of the sign in isolation.

2.2. Link to the LSFB dictionary

The web interface also includes, when possible, a link between each lexical entry and the corresponding sign within the external online dictionary of LSFB developed by the LSFB Association[6] in collaboration with the LSFB-Laboratory of the University of Namur. For each sign, the dictionary provides a definition and examples in LSFB, etymological information when available, as well as information about the regional distribution of the sign, links towards regional variants, homonyms and synonyms. The collaboration between the Corpus LSFB project and the dictionary aims to gradually supply the dictionary, currently based on limited monological data, with the information extracted from the dialogical and wider data of the

[5] Because of the lack of trained LSFB-French (the first MA started in September 2014, see section 1.1), we choose to work in collaboration with a trained of French Sign Language-French translator, who si also a trainer.

[6] http://dicto.lsfb.be/dico

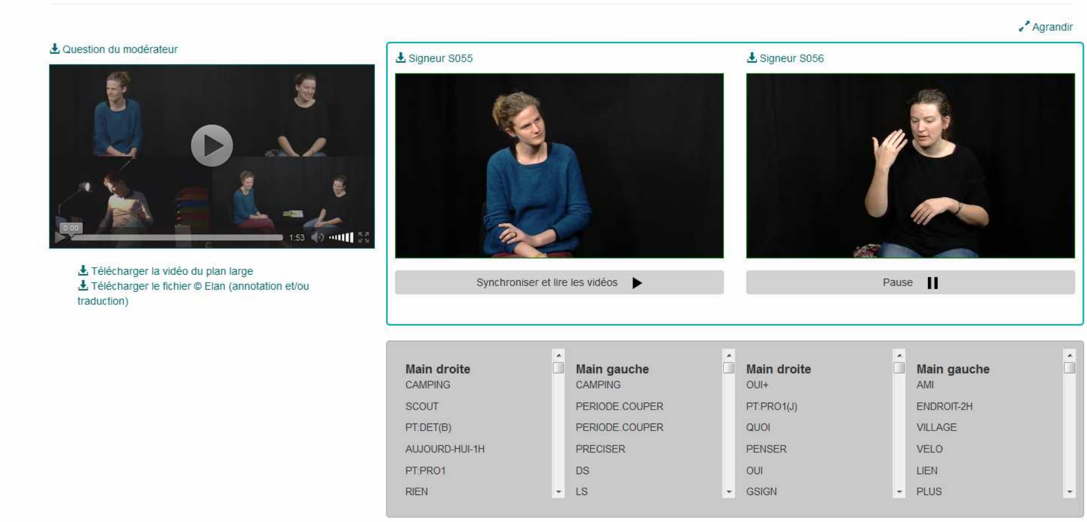


Figure 5: On the left: question file (with the 4 shots combined). On the right: single shot for each signer, as it appears on the website

LSFB corpus, presumably more representative of the language in use. Figure 6 and 7 illustrate the link between the lexical database of the corpus and the online external dictionary.

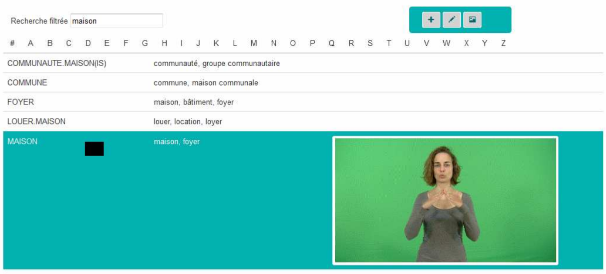


Figure 6: Lexical database and icon (a small black book) symbolising the link to the external dictionary

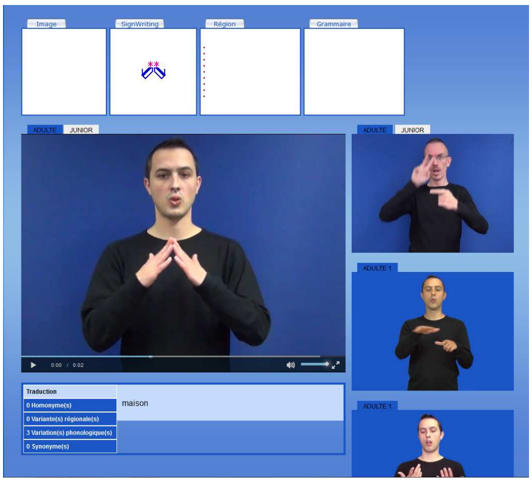


Figure 7: Online external LSFB dictionary

3. The website

The corpus LSFB website (www.lsfb-corpus.be) provides access to the whole content of the corpus. This includes the data (4 videos for each of the 19 tasks and for the 50 recording sessions), the metadata about signers (gender, age, profile and LSFB variant used) and the tasks (description and elicitation materials), the annotations and the translations. The videos showing each signer from each pair are synchronised online. All videos and .eaf ELAN files can be downloaded. The annotations and the translations can be displayed in real time while the videos are playing.

3.1. Licensing

The conditions of use of the LSFB Corpus website follow the BY-NC-SA Creative Commons conditions (http://creativecommons.org/licenses/by-nc-sa/4.0/legalcode). Which means:

1. Attribution (BY) - Any use of the corpus has to refer to the website www.corpus-lsfb.be and have to cite “Meurant, L. 2015. Corpus LSFB. First digital open access corpus of movies and annotations of French Belgian Sign Language (LSFB). LSFB-Lab, University of Namur. URL: http://www.corpus-lsfb.be”

2. Non commercial (NC) - The website is only for non-commercial uses.

3. Share alike (SA) - Any use has to be performed under the same sharing conditions (using the same Creative Commons licence).

By these conditions, we aim to encourage the use of the data by researchers, professionals and the general public.

3.2. Search options

Users can choose the way they want to browse and query the data and metadata. The ‘Corpus’ tab offers three kinds of entries (search, free consultation, consultation by signer)

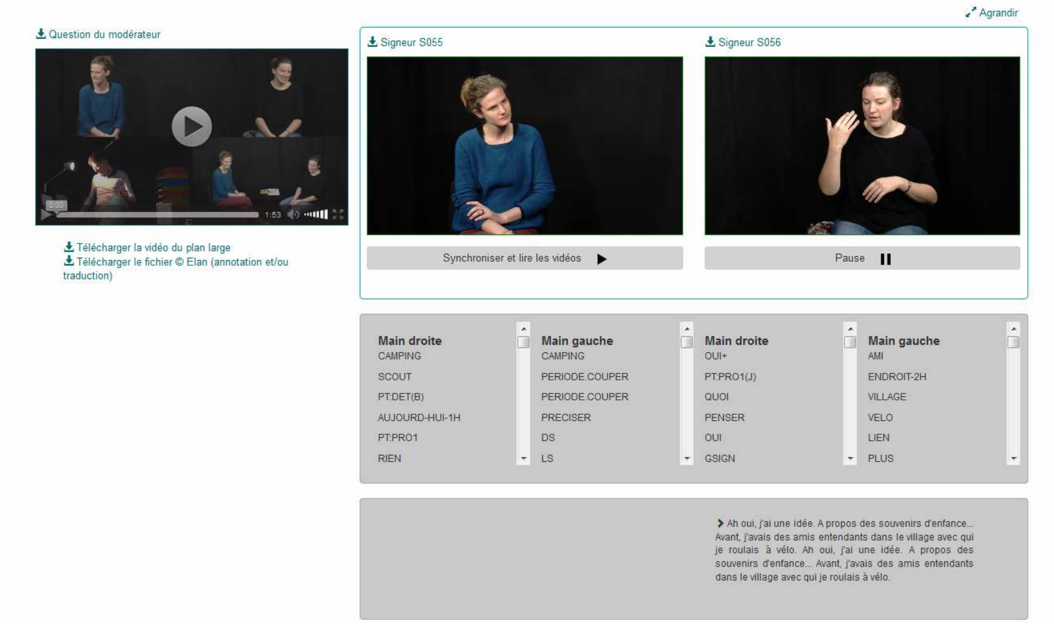


Figure 8: On the left: question file (with the 4 shots combined). On the right: single shot for each signer, as it appears on the website with annotations and translation.

and a demo that explains the general content of the website. The 'search' option leads the users to a page with six different ways to find videos (see figure 9):

Figure 9: The search page of the website, showing the various ways to search through the data.

1. Signer's profile – The search by signers profile includes 4 different filters that can be combined together: gender, age, linguistic profile (native, near-native or late learners) and regional variant.

2. Regional variants – A map of Belgium allows the users to choose one region. The results give all the sessions in which signers from this region appear.

3. Descriptors – Five filters enable a query to be made regarding the content of the videos, which means the subjects covered by the discussions as well as some linguistic features. Three descriptors are related to deaf culture: school, humour and name-sign. Three descriptors concern grammatical features of the language: classifiers, use of space and iconicity. Then, three descriptors are available to search videos containing discussions about the relationship of the signers to their language: the impact of emotions on their language, the topic of the norm, and the topic of language variations. Seven descriptors relate to the topic of the discussion: childhood, family, stories, hobbies, societal issues, deaf issues and memories. Finally, eight different discourse genres are available as descriptors as well: argumentation, conversation, description, discussion, explanation, fictive narrative, life story and procedure.

4. ID-gloss or keyword – It is also possible to find video clips by searching through the ID-glosses and the content of the lexical database. When querying for a gloss, the users obtain all the videos (and the time codes) in which the gloss appears.

5. Visual material – Pretty soon, it will be possible to find videos by means of the material used for the tasks.

6. The last choice enables only the list of videos that have been annotated and/or translated so far to be displayed.

3.3. Personalised session

Users can choose simply to visit the website or to create an account. The profile of the account varies in line with the user's profession: public, professional (teacher, interpreter) or researcher. Without an account, a visitor to the site cannot access the videos. The public profile allows users to see the data (apart from the videos censored by the participants concerned) but not the metadata. The professional profile enables the user to view data (apart from the videos censored by the participants concerned) and some metadata; and finally, the researcher profile provides access to all data and metadata.

Each registered user accesses the corpus on their own session. This allows them to add personal comments (linked to a specific video), to tag videos as favorites and to choose some parameters such as the automatic display of annotations and/or translation, the main colour of the website, etc. At the time of writing this paper, the website is available in French and in LSFB (see Figure 10). All the videos are made available for people with Usher syndrome. In July, the interface will also be available in English and International Signs. Three months after the launch of the website, 237 accounts have been created: 163 with a public profile, 56 with a professional profile (LSFB teachers, interpreters and interpreting trainers) and 5 with a researcher profile. The 13 remaining accounts belongs to the LSFB-Laboratory team members and the administrators of the web site.

Figure 10: Bilingual (LSFB and French) interface of the website.

4. Conclusion

Collecting and sharing the LSFB corpus led us to develop new tools that are expected to facilitate the study of LSFB and to foster the use of corpus data by teachers, interpreters and students. Three of them are worth mentioning, because of their innovative status within the field.

1. The online lexical database linked to ELAN facilitates the annotation process.

2. The annotations and the translations, time aligned with the videos, are a real asset for user-friendly web browsing of the LSFB corpus.

3. The numerous search options for browsing the data (videos), the metadata (signers and tasks) and the annotations make the LSFB corpus a very useful tool for corpus mining.

Of course, the availability of the data opens new perspectives in the development of corpus linguistic research on LSFB. The major illustration of this change is the three PhD theses in preparation on LSFB discourse: on fluency and disfluency markers, on the impact of genre variation on phonetic variation and on buoys and discourse markers. But such data also open the possibility to conduct cross-linguistic studies , and in particular to investigate whether VGT and LSFB, considered as one 'Belgian Sign Language' until very recently, have evolved as two variants of the same original language, or as two different languages.

In the short term, we consider the LSFB corpus as the first step for building parallel corpora to be used with the methodology of corpus-based contrastive linguistics between LSFB and French. For this purpose, we are currently collecting spoken French data on the basis of the same tasks as the ones used for the LSFB corpus and in the same conditions (e.g. in a studio, with pairs of participants, a moderator, and 4 cameras). These next steps aim at answering the various issues faced by all deaf learners of French, and in particular the ones registred in the bilingual educational program in Namur, as well as their teachers, on the one hand, and by the interpreters and the interpreting trainers, on the other hand, in terms of comparison between LSFB and French.

5. Acknowledgements

The LSFB corpus and the website are the result of the *Corpus LSFB* project funded by a three-year Incentive Grant for Scientific Research (M.I.S. F.4505.12) awarded by the F.R.S.-FNRS to Laurence Meurant. We acknowledge all the signers who participated to this project.

6. Bibliographical References

Crasborn, O. and Zwitserlood, I. (2008). The corpus NGT: an online corpus for professionals and laymen. In O. Crasborn, et al., editors, *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages*, pages 44–49. ELDA.

Crasborn, O. and Zwitserlood, I. and Ros, J. (2008). *The Corpus NGT. An open access digital corpus of movies with annotations of Sign Language of the Netherlands*. Centre for Language Studies, Radboud University Nijmegen, ISLRN 175-346-174-413-3.

Degand, L. and Simon, A. C. (2005). Minimal Discourse Units: Can we define them, and why should we? In *Proceedings of SEM-05. Connectors, discourse framing and discourse structure: from corpus-based and experimental analyses to discourse theories*, pages 65–74, Biarritz.

Gabarró-López, S., Meurant, L., and Barberá.Altimira, G. (2016). *Digging into buoys: their use across genres and their status in signed discourse*. Poster presented at TISLR 12 - 12th International Conference on Theoretical Issues in Sign Language Research, Melbourne, Australie.

Gabarró-López, S. (2016). *When "same" means something different: the sign SAME as a discourse relational device in two sign languages*. Paper presented at 3rd International Conference on Linguistic and Psycholinguistic Approaches to Text Structuring (LPTS 2016), Valencia, Espagne.

Ghesquière, M. and Meurant, L. (2016). L'envers de la broderie. une pédagogie bilingue langue des signes de belgique francophone - français. *Glottopol*, 27:95–111.

Ghesquière, M., de Halleux, C., and Meurant, L. (2015). Bilingual education by immersion in Namur, Belgium: Principles and pedagogic issues. In *Proceedings of the 22nd International Congress on the Education of the Deaf*.

Gotz, S. (2013). Fluency in native and nonnative english speech. *Studies in Corpus Linguistics*, 53.

Trevor Johnston. (2004-2007). *Corpus of grammar and discourse strategies of deaf native users of Auslan (Australian Sign Language)*. Macquarie University.

Johnston, T. (2010). From archive to corpus: transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, 15:106–131.

Johnston, T. (2015). Auslan Corpus Annotation Guidelines. Guidelines, Macquarie University; La Trobe University, Sydney; Melbourne.

Meurant, L. and Sinte, A. (2013). Towards a corpus of French Belgian Sign Language (LSFB) discourses. In C. Bolly et al., editors, *Text-Structuring: Across the Line of Speech and Writing Variation*, pages 199–212.

Meurant, L. and Sinte, A. (2016). La reformulation en langue des signes de Belgique francophone (LSFB). analyse dans un corpus de trois types de discours:narration, explication et conversation. *L'Information grammaticale*, 149:33–45.

Meurant, L., Cleve, A., and Crasborn, O. (2016). Using sign language corpora as bilingual corpora for data mining: Contrastive linguistics and computer-assisted annotation. volume *Proceedings of the 7th workshop on the Representation and Processing of Sign Languages:Corpus Mining : LREC 2016*.

Meurant, L. (2015). *Corpus LSFB. First digital open access corpus of movies and annotations of French Belgian Sign Language (LSFB). Laboratoire de langue des signes de Belgique francophone (LSFB-Lab)*. FRS-F.N.R.S et Université de Namur.

Notarrigo, I. and Meurant, L. (2015). *Markers of (dis)fluency across signers' profiles in French Belgian Sign Language (LSFB). A comparative analysis between Native, Near-Native and Late Signers*. Poster presented at 2nd International Conference on Sign Language Acquisition, Amsterdam, Pays-Bas.

Paligot, A. and Meurant, L. (2016). *Weak Hand Lowering across signing styles of French Belgian Sign Language (LSFB)*. Poster presented at TISLR 12 - 12th International Conference on Theoretical Issues in Sign Language Research, Melbourne, Australia.

Paligot, A., Van.Der.Kooij, E., Crasborn, O., and Bank, R. (2016). *Weak Drop in Context*. Paper presented at TISLR 12 - 12th International Conference on Theoretical Issues in Sign Language Research, Melbourne, Australia.

Prillwitz, S. (2008). DGS Corpus Project: Development of a Corpus-Based Electronic dictionary of German Sign Language/German. In *LREC 2008 Workshop Proceedings*, volume 23.

Christian Rathmann. (2015). *DGS Corpus Online*.

Sinte, A., De.Clerck, C., Fonzé, S., Sanchez, S., and Meurant, L. (2015). *Corpus LSFB (French Belgian Sign Language). Current annotation conventions compared with the "Digging into Signs"'s suggestions*. Poster presented at Digging into Signs Workshop: Developing Annotation Standards for Sign Language Corpora.

Van.Herreweghe, M. and Vermeerbergen, M. (2012). Data collection. In R. Pfau M. Steinbach et al., editors, *Sign Language. An International Handbook. Handbooks of Linguistics and Communication Science (HSK) series*, volume 37, pages 1023–1045. Mouton de Gruyter.

Sign Classification in Sign Language Corpora with Deep Neural Networks

Lionel Pigou, Mieke Van Herreweghe, Joni Dambre
Ghent University
{lionel.pigou, mieke.vanherreweghe, joni.dambre}@ugent.be

Abstract

Automatic and unconstrained sign language recognition (SLR) in image sequences remains a challenging problem. The variety of signers, backgrounds, sign executions and signer positions makes the development of SLR systems very challenging. Current methods try to alleviate this complexity by extracting engineered features to detect hand shapes, hand trajectories and facial expressions as an intermediate step for SLR. Our goal is to approach SLR based on feature learning rather than feature engineering. We tackle SLR using the recent advances in the domain of deep learning with deep neural networks. The problem is approached by classifying isolated signs from the Corpus VGT (Flemish Sign Language Corpus) and the Corpus NGT (Dutch Sign Language Corpus). Furthermore, we investigate cross-domain feature learning to boost the performance to cope with the fewer Corpus VGT annotations.

Keywords: sign language recognition, deep learning, neural networks

1. Introduction

SLR systems have many different use cases: corpus annotation, in hospitals, as a personal sign language learning assistant or translating daily conversations between signers and non-signers to name a few. Unfortunately, unconstrained SLR remains a big challenge. Sign language uses multiple communication channels in parallel with high visible intra-sign and low inter-sign variability compared to common classification tasks. In addition, publicly available annotated corpora are scarce and not intended for building classifiers in the first place.

A common approach in SLR is to get around the high dimensionality of image-based data by engineering features to detect joint trajectories (Charles et al., 2013), facial expressions (Liu et al., 2014) and hand shapes (Ong and Bowden, 2004) as an intermediate step. Data gloves (Oz and Leu, 2011), colored gloves (Wang and Popović, 2009) or depth cameras (Chai et al., 2013) are often deployed in order to obtain a reasonable identification accuracy.

In recent years, deep neural networks achieve state-of-the-art performance in many research domains including image classification (Szegedy et al., 2014), speech recognition (Graves et al., 2013) and human pose estimation (Pfister et al., 2014). The deep learning models that we use in this work are based on convolutional neural networks (CNNs) (Lecun et al., 1998). A CNN is a model with many parameters that are adjusted iteratively using optimization algorithms (= *learning*) and a large amount of annotated data.

In previous work (Pigou et al., 2015), we showed that deep neural networks are very successful for gesture recognition and gesture spotting in spatiotemporal data. Our developed system is able to recognize 20 different Italian gestures (i.e., emblems). We achieved a classification accuracy of 97.23% in the *Chalearn 2014 Looking At People* gesture spotting challenge (Escalera et al., 2014). This gives us an indication that deep neural networks can be useful for SLR.

In this work, the problem is approached by classifying isolated signs from the Corpus VGT (Van Herreweghe et al., 2015), the Flemish Sign Language Corpus, and the Corpus NGT (Crasborn et al., 2008; Crasborn and Zwitserlood, 2008), the Dutch Sign Language Corpus. Furthermore, we investigate cross-domain feature learning to boost the performance to cope with the fewer Corpus VGT annotations.

2. Methodology

2.1. Data

The two corpora used to explore SLR (Corpus VGT and Corpus NGT) have similar camera setups and use very similar gloss annotation rules with identical software (ELAN). Both corpora consist of Deaf signers that perform tasks such as retelling comic strips, discuss an event and debating on chosen topics. For each corpus, the 100 most frequently used signs are extracted together with their gloss. The data is split into three sets: 70% training set, 20% test set and 10% validation set. The training set is used to optimize the

Figure 1: A sample from the Corpus VGT (Ghent University), filmed from three viewpoints.

Figure 2: A sample from the Corpus NGT (Radboud University Nijmegen), filmed from two viewpoints.

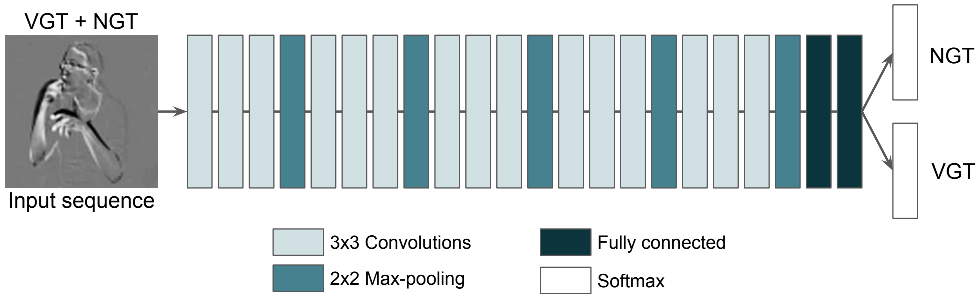


Figure 3: The architecture overview of the deep neural network used in this work. All layers are shared among corpora, except for the softmax classifier. This will boost the performance for the Corpus VGT, as it learns better features using the Corpus NGT with more annotations.

neural networks, the validation set is used for evaluation during training and the test set is used to evaluate the final models.

The Corpus VGT (Figure 1) uses Flemish Sign Language. The project started in Juli 2012 and ended in November 2015 at Ghent University, in collaboration with the Linguistics Group VGT of KU Leuven Campus Antwerp, and promoted by Prof. Dr. Mieke Van Herreweghe (Ghent University) and Prof. Dr. Myriam Vermeerbergen (KU Leuven Campus Antwerp). The corpus contains 140 hours of video and a small fraction is annotated. After cleaning the data, we extracted a total of 12599 video-gloss pairs from 53 different Deaf signers.

The Corpus NGT (Figure 2) contains Deaf signers using Dutch Sign Language from the Netherlands. This project was executed by the sign language group at the Radboud University Nijmegen. Every narrative or discussion fragment forms a clip of its own, with more than 2000 clips. We extracted a total of 55224 video-gloss pairs from 78 different Deaf signers.

As Figure 4 shows, there is a class imbalance for both corpora. This means that accuracy measures will be highly skewed. For example, only predicting the most common sign (which is "ME") for every sample across the whole dataset already results in 30.9% and 11.2% accuracy for the Corpus NGT and the Corpus VGT respectively.

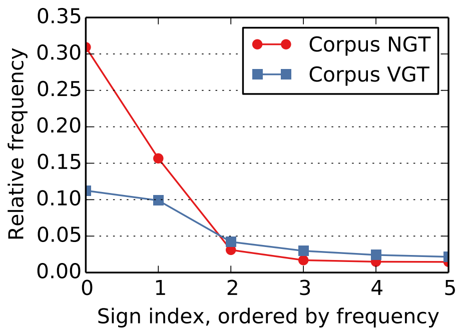


Figure 4: The relative frequency for the five most common signs in both corpora. The class imbalance is significant in both corpora, but is especially prevalent for the Corpus NGT.

2.2. Convolutional Neural Network (CNN)

CNNs are models that allow to learn a hierarchy of layered features instead of manually extracting them. They are among the most successful techniques in deep learning, a domain in machine learning that has proven to be very successful at recognizing patterns in high dimensional data such as images, videos and audio. These artificial networks are inspired by the visual cortex of the human brain. The neurons in a CNN will connect to a local region of the image, called a receptive field. This is accomplished by performing discrete convolutions on the image with filter values as trainable weights, which are optimized using the *gradient descent* algorithm. A second important building block in a CNN is a pooling scheme, where only the interesting information of the feature maps is pooled together.

These base operations are performed in multiple layers as illustrated in Figure 3. This architecture is inspired by (Simonyan and Zisserman, 2014). Three convolutional layers are stacked before performing max-pooling (only the maximum activation of each region remains) on non-overlapping 2x2 spatial regions. The input image sequence consists of 8 frames of size 128x128. Each frame is subtracted from the previous frame to remove static information. These frames are rotated, shifted and stretched randomly during training to artificially increase the amount of data in order to learn more generalized features. This technique is called *data augmentation*.

3. Results

3.1. Corpus NGT

The resulting model, with the highest score on the validation set, is illustrated in Figure 3 (without the VGT branch). The shorthand notation of the full architecture is as follows: C^3_{32}-P-C^3_{64}-P-C^3_{128}-P-C^3_{256}-P-C^3_{512}-P-D_{2048}-D_{2048}-S, where C^a_b denotes a stacked convolutional layers with b feature maps and 3x3 filters, P a max-pooling layer with 2x2 pooling regions, D_c a fully connected layer with c units and S a softmax classifier.

The top-N accuracy is a measure indicating the probability that the correct answer is within the model's N best guesses. The top-N accuracies of the test set for the Corpus NGT are depicted in Figure 5. The CNN achieves a top-1, top-3 and top-5 accuracy of 56.2%, 75.7% and 82.1% respectively for 100 signs. This is especially interesting for automatic corpus

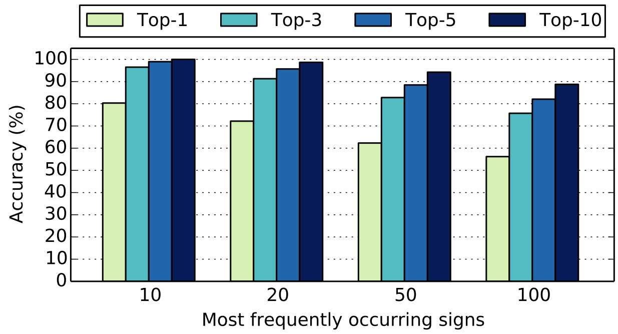


Figure 5: **Corpus NGT** top-N accuracies. A measure indicating the probability of the correct answer being within the model's N best guesses.

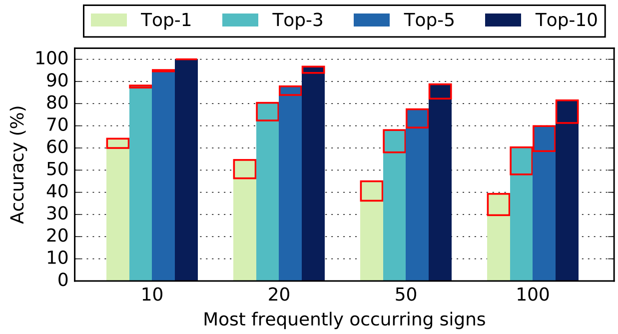


Figure 7: **Corpus VGT** top-N accuracies with cross-domain learned features. The red outline shows the improvement compared to the accuracies without cross-domain learning.

annotation, where providing a list with the N best guesses is appropriate.

As mentioned above, we have to keep in mind the class imbalance. The confusion matrix shows the fraction of true positives for each class (each sign) on the diagonal. It also tells us which classes it gets confused with. To have a better insight into the model's performance, we show the confusion matrix in Figure 6. Not surprisingly, almost all classes get confused with frequently occurring ones. The CNN learned to bet on common glosses when it is unsure about a certain input, because more often than not it will get rewarded for that. Other misclassification is due to signs that are hard to distinguish from each other.

3.2. Corpus VGT

To cope with the smaller amount of annotations for the Corpus VGT compared to the Corpus NGT, we train a shared model on both corpora (Figure 3). This cross-domain learning is a form of *transfer learning*, where the knowledge of one or more domains (in this case the Corpus NGT) is useful for other domains. Our motivation is that the learned features for both domains should be similar, except for the softmax classifier. All sign languages have similar visual features: they consist of hand, arm, face and body expressions. We hope to capture these generic building blocks in order to boost the performance for the Corpus VGT.

In Figure 7, the top-N accuracies are shown. It achieves a top-1, top-3 and top-5 accuracy of 39.3%, 60.3% and 69.9% respectively for 100 signs. To show the improvement using the cross-domain learning, the sensitivity (true positive rate) increase for each class is depicted in Figure 9. We clearly see a significant improvement for most signs, but a few classes are negatively affected by it. The resulting confusion matrix is shown in Figure 8. The errors are more spread out than the ones for the Corpus NGT, because the class imbalance is less prevalent.

4. Conclusion and Future Work

We show that CNNs are capable of learning features from image sequences across linguistic sign language corpora.

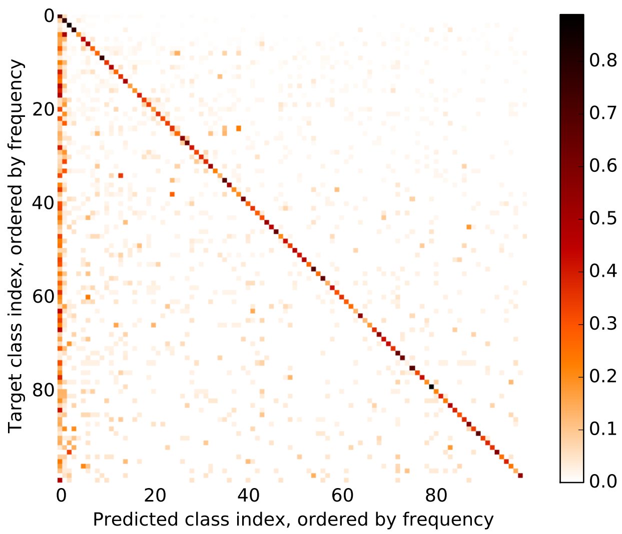


Figure 6: **Corpus NGT** confusion matrix indicating the classification performance of the deep neural network.

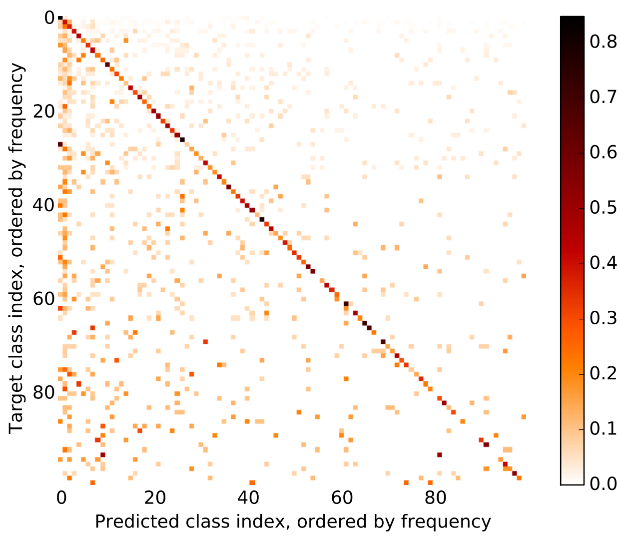


Figure 8: **Corpus VGT** confusion matrix with cross-domain learned features.

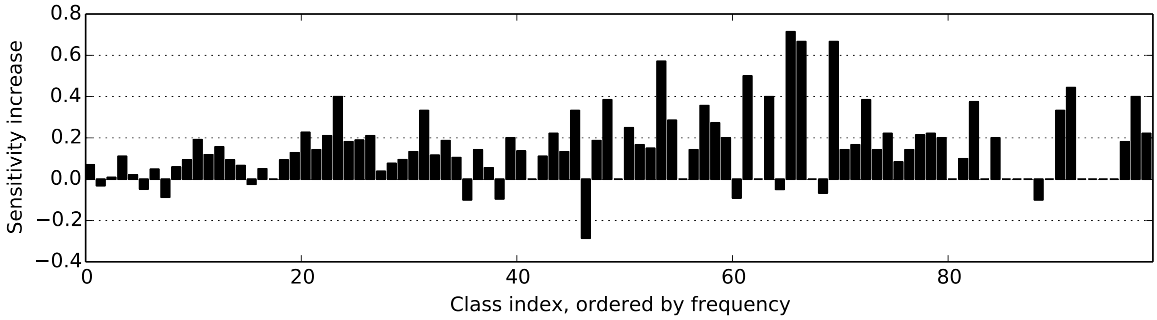


Figure 9: **Corpus VGT** sensitivity (true positive rate) increase compared to the model without cross-domain feature learning, depicted for each sign. Some signs are negatively affected by it. Further research will be required to determine the reason.

Our models achieve an accuracy of 39.3% with the Corpus VGT and 56.2% with the Corpus NGT for the 100 most common signs. We also show that the knowledge learned from the Corpus NGT can be passed on to boost the performance of the Corpus VGT.
Given the high dimensionality of video, the fact that these corpora are not tailored for machine learning and the fast and subtle movements of Deaf signers, deep neural networks show potential to build upon for SLR. The need for manual feature engineering, specialized hardware or other constraints decreases with more available corpora, advancements in unsupervised learning (learning from data without annotations) and language modeling.

5. Acknowledgments

We would like to thank NVIDIA Corporation for the donation of a GPU used for this research. The research leading to these results has received funding from the Agency for Innovation by Science and Technology in Flanders (IWT).

6. Bibliographical References

Chai, X., Li, G., Lin, Y., Xu, Z., Tang, Y., Chen, X., and Zhou, M. (2013). Sign language recognition and translation with kinect. In *IEEE Conf. on AFGR*.

Charles, J., Pfister, T., Everingham, M., and Zisserman, A. (2013). Automatic and efficient human pose estimation for sign language videos. *International Journal of Computer Vision*, pages 1–21.

Crasborn, O. A. and Zwitserlood, I. (2008). The corpus ngt: an online corpus for professionals and laymen. In *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages (LREC)*, pages 44–49. ELDA.

Crasborn, O., Zwitserlood, I., and Ros, J. (2008). The Corpus NGT. A digital open access corpus of movies and annotations of Sign Language of the Netherlands. *Centre for Language Studies, Radboud Universiteit Nijmegen. http://www.ru.nl/corpusngtukgp/.*

Escalera, S., Baró, X., Gonzàlez, J., Bautista, M. A., Madadi, M., Reyes, M., Ponce, V., Escalante, H. J., Shotton, J., and Guyon, I. (2014). Chalearn looking at people challenge 2014: Dataset and results. In *ECCV workshop*.

Graves, A., Mohamed, A.-r., and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 6645–6649. IEEE.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11).

Liu, J., Liu, B., Zhang, S., Yang, F., Yang, P., Metaxas, D. N., and Neidle, C. (2014). Non-manual grammatical marker recognition based on multi-scale, spatio-temporal analysis of head pose and facial expressions. *Image and Vision Computing*, 32(10):671–681.

Ong, E.-J. and Bowden, R. (2004). A boosted classifier tree for hand shape detection. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 889–894. IEEE.

Oz, C. and Leu, M. C. (2011). American sign language word recognition with a sensory glove using artificial neural networks. *Engineering Applications of Artificial Intelligence*, 24(7):1204–1213.

Pfister, T., Simonyan, K., Charles, J., and Zisserman, A. (2014). Deep convolutional neural networks for efficient pose estimation in gesture videos. *Asian Conference on Computer Vision (ACCV).*

Pigou, L., Oord, A. v. d., Dieleman, S., Van Herreweghe, M., and Dambre, J. (2015). Beyond temporal pooling: Recurrence and temporal convolutions for gesture recognition in video. *arXiv preprint arXiv:1506.01911*.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2014). Going deeper with convolutions. *arXiv preprint arXiv:1409.4842*.

Van Herreweghe, M., Vermeerbergen, M., Demey, E., De Durpel, H., H., N., and Verstraete, S. (2015). Het Corpus VGT. Een digitaal open access corpus van videos and annotaties van Vlaamse Gebarentaal, ontwikkeld aan de Universiteit Gent ism KU Leuven. www.corpusvgt.be.

Wang, R. Y. and Popović, J. (2009). Real-time hand-tracking with a color glove. *ACM transactions on graphics (TOG)*, 28(3):63.

Creating Corpora of Finland's Sign Languages

Juhana Salonen, Ritva Takkinen, Anna Puupponen, Henri Nieminen, Outi Pippuri
Department of Languages (Sign Language Centre), University of Jyväskylä, Finland
P.O. Box 35, FI-40014 University of Jyväskylä, Finland
E-mail: {juhana.salonen, ritva.a.takkinen, anna.puupponen, outi.pippuri}@jyu.fi & henri.i.nieminen@student.jyu.fi

Abstract

This paper discusses the process of creating corpora of the sign languages used in Finland, Finnish Sign Language (FinSL) and Finland-Swedish Sign Language (FinSSL). It describes the process of getting informants and data, editing and storing the data, the general principles of annotation, and the creation of a web-based lexical database, the FinSL Signbank, developed on the basis of the NGT Signbank, which is a branch of the Auslan Signbank. The corpus project of Finland's Sign Languages (CFINSL) started in 2014 at the Sign Language Centre of the University of Jyväskylä. Its aim is to collect conversations and narrations from 80 FinSL users and 20 FinSSL users who are living in different parts of Finland. The participants are filmed in signing sessions led by a native signer in the Audio-visual Research Centre at the University of Jyväskylä. The edited material is stored in the storage service provided by the CSC – IT Center for Science, and the metadata will be saved into CMDI metadata. Every informant is asked to sign a consent form where they state for what kinds of purposes their signing can be used. The corpus data are annotated using the ELAN tool. At the moment, annotations are created on the levels of glosses and translation.

Keywords: sign language corpus, Finnish Sign Language, Finland-Swedish Sign Language, annotation, metadata, Signbank

1. Background

In Finland there are two official sign languages, Finnish Sign Language (FinSL) and Finland-Swedish Sign language (FinSSL). FinSL is used mainly by deaf people who come from Finnish-speaking families and have attended Finnish deaf schools. The estimated number of deaf FinSL users is 4000–5000 and of hearing native signers (mainly codas) and second language users approximately 6000–9000[1]. FinSSL, on the other hand, is used mainly in the coastal areas of Finland among those deaf people whose family background is Swedish speaking. The number of deaf FinSSL users is now estimated at approximately 90, most of them over 55 years of age (Soininen, 2016). The creation of corpora will enable us to conduct wider, deeper, more diverse and more reliable research, on which we will be able to construct a comprehensive dictionary and a descriptive grammar of these two languages. Creating the corpus especially for FinSSL is crucial as the number of users is very small and includes mainly elderly people. It is essential that the documentation of the language takes place at once.

The corpus project was piloted at the Sign Language Centre of the University of Jyväskylä in 2013. In spring 2014 the four-year (2014–2018) CFINSL[2] project began, its aim to document both FinSL and FinSSL. The documentation will serve both linguistic (vocabulary, structure, language use, variation) and cultural (topics related to the deaf community) purposes as well as teaching. We aim to collect conversations and narrations from 80 FinSL users and 20 FinSSL users who are living in different parts of Finland.

[1] *http://www.kuurojenliitto.fi/fi/viittomakielet/viittomakielet-ja-viittomakieliset#.VrBp5E1f3L8*

[2] *https://www.jyu.fi/hum/laitokset/kielet/oppiaineet_kls/viittomakieli/tutkimus/menossa-olevat-projektit/suomen-viittomakielten-korpusprojekti*

2. Procedure

2.1 Collecting the data

In the project we collect data from participants in different parts of Finland with the help of contact persons in the deaf clubs. The material is recorded in the Audio-visual Research Centre at the University of Jyväskylä. The material is recorded in a professional setting in order to produce high-quality video material, for example for the quantitative phonetic analysis of FinSL and FinSSL (with e.g. computer-vision based technologies). We have tried to ensure a wide range of regional variation by recruiting participants from seven different parts of Finland (see the map in Figure 1).

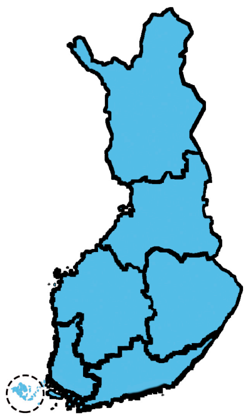


Figure 1: A map showing the areas where participants are recruited.

Normally both of the participants in a dialogue setting are from the same area in order to preserve and make clear any regional variation. Also the age variety is taken into account: we aim to get as even a distribution as possible across a range of ages: 18–29, 30–39, 40–54, 55–69 and 70–.

Participants are invited to a signing session led by a native signer. They are asked to perform seven language tasks, all of which are carried out in a dialogue setting. Tasks 1–2 and 6–7 are discussions, while tasks 3–5 are semi-interactive monologues. The tasks involve 1) introductions, 2) a discussion of work or hobbies, 3) narrating about cartoon strips (Ferd'nand), 4) narrating about a video, 5) narrating a story from a picture book (The Snowman, and Frog, where are you?), 6) discussing a topic related to the deaf world, and 7) free discussion (e.g. on travelling, TV-programmes, sports). Since some of the elicitation materials have also been used when collecting corpus material in other sign languages (e.g. Nishio et al., 2010; Mesch, 2015), the data will allow cross-linguistic comparison.

The video recording takes place in a studio of the Audio-visual Research Centre (see Figure 2). Before the recording session the instructor, a native signer, has a discussion with the two participants and explains what will happen in the signing situation. During the recording the instructor and the participants are present in the studio and the technicians are in a separate control room (see Figure 3). The instructor gives the participants instructions before each task. During the tasks he is available if more information is needed but otherwise he leaves the participants to discuss freely.

In the first task the participants take it in turn to introduce themselves. The other participant can ask for more information if he/she wants to. Task 2, telling about work or hobbies, is also signed by each signer in turn but discussion is free during each turn. Narrations about cartoons, videos and picture books (tasks 3, 4 and 5) are individual narrations, and discussion may take place afterwards. Tasks 6 and 7 are free dialogues and include a discussion about the deaf world and a free discussion. The length of the sessions is between one and a half and two hours.

Figure 2: The studio

Figure 3: The control room

The video recording takes place at the studio with seven Panasonic video cameras (3 x AG-HPX371E, 1 x AW-HE120KE, 3 x AG-HPX171E). Camera 1 records a general view of the situation, camera 2 records a complete picture of Signer B and camera 3 a complete picture of Signer A. Cameras 4 and 5 are angled towards the torso and face of Signers B and A, respectively. Camera 6 is angled towards the signers from directly above in order to get exact information of the movements of the head, body and hands on the sagittal plane. Camera 7 is directed towards the instructor in order to record the instructions given before and possibly during the tasks (see Figure 4). The HD films are saved in P2-disks (25–50 fps), stored in MXF format and compressed into low and high resolution MP4 files.

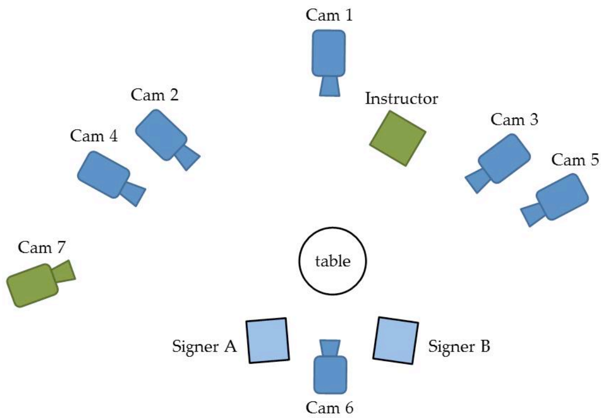


Figure 4: Recording in the studio: Camera setting

The edited material is stored in the storage service provided by the CSC – IT Center for Science[3], which is a state-owned company administered by the Finnish Ministry of Education and Culture. In addition, the annotated files will be stored in the Language Bank of Finland administered by FIN-CLARIN[4], which is part of the international CLARIN infrastructure. The data will be available for research and teaching purposes when permitted by the language informants.

2.2 Metadata

Metadata, in other words "data about data" (e.g. Burnard, 2014), are a crucial part of a corpus. Relevant metadata make the data accessible, and are appended to all media and annotation files. The metadata documented in the CFINSL project include information about the corpus itself (its name, language, the size of the corpus, distributor etc.) as well as about the participants (region, sex, age and education etc.), the content (the various language tasks and elicitation materials used), media (format and type), project (name, language, methodology) and

[3] CSC - IT Center for Science Ltd. maintains and develops the state-owned centralised IT infrastructure and uses it to provide nationwide IT services for research, libraries, archives, museums and culture as well as information, education and research. *https://www.csc.fi/csc*

[4] *https://kitwiki.csc.fi/twiki/bin/view/FinCLARIN/*

session (task name, participants, etc.). The metadata are currently documented in Excel, from which they will be converted into CMDI metadata (Component MetaData Infrastructure), a framework initiated and developed by CLARIN for the description and use of metadata. Searching the data can be done in ELAN[5] (Crasborn & Sloetjes, 2008), which is also the tool used for annotating the material.

2.3 Consent

The establishment of a corpus of sign language with open access is a sensitive issue because visual material is used. It is important to show the face of the language informant because the facial area carries a lot of grammatical and lexical information. Thus the informant cannot be made anonymous. It is therefore essential to carefully explain to the informant in both written form and in sign language that her/his signing will be available for research and later will be partly publicly available on the Internet. Every informant is required to sign a consent form where consent for different kinds of uses of her/his signing is sought separately, allowing every informant to decide for what purpose(s) he/she will permit his/her signing material to be used. On the consent form there are five different parts and participants must choose either the yes or the no option for each of them:

- Video material can be used for research purposes in the CFINSL project but publishing video clips or still images is prohibited
- Video material can be presented in public events (e.g. academic presentations and teaching)
- Still images can be taken from the video material for publications (electronic or paper)
- The whole video material can be published electronically e.g. in the Internet
- The name of the participant can be mentioned in publications

The informant will have the right to check her/his material, before its presentation or publication. Moreover, she/he can ask the administrator to remove her/his recorded video material from the corpus if she/he so wishes.

In addition, we will comply with the Personal Data Act (523/1999) as well as with the regulations set out by Office of the Data Protection Ombudsman concerning a Personal Data File, by creating a Description of File.

3. Annotation

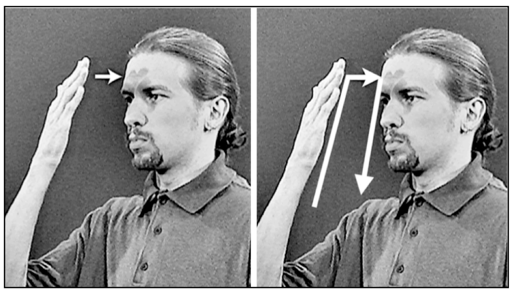
The ongoing process of annotating CFINSL data began during 2015. The corpus data are annotated using the ELAN tool, which enables time-aligned annotations to video media. The work started with the raw annotation of the narrative and discourse data of altogether 22 participants and approximately four hours of material. We started the annotation with a small amount of data with the aim of drawing up guidelines for creating annotation conventions for the corpus annotation. The work group in the annotation process consists of several annotators (most of them native signers) and translators (both native signers and native speakers). The annotation process was divided into three rounds:

1. The first round (raw annotation throughout 2015) was based on annotation with two tiers (a gloss tier and its comments)
2. The second round (during January-August 2016) is based on annotation with five tiers (a gloss tier of left/right hand, their comments, translation and comments on it)
3. The third round (starting in September 2016) will be a systematization of the annotation of the second round.

Our annotation work in the CFINSL project is based on four principles:

1. The length of the annotation cells is based on a view of the sign as a relatively long unit
2. Structural information is used in glosses to distinguish between forms with the same meaning (both phonetic and lexical variation)
3. Glosses are created for form-meaning pairs according to the contextual meaning of signs in discourse
4. Annotation is seen as a tool for future research and teaching

The first principle is related to Jantunen's (2013, 2015) understanding of a sign as a relatively long unit (Figure 5, see also the concept of broad segmentation in Hanke et al., 2012). This specification of the sign influences our annotation in that the annotation cells are presumed to be longer than the annotations done on the basis of the present mainstream view of a sign's length.



BLACK (short) BLACK (longer)

Figure 5: The length of the sign according to the mainstream view (short) and according to the view on which CFINSL annotation is based (longer). (Images from Jantunen, 2015: 117.)

In the present work we do not focus on exploring the borders (on- and off-sets) of the sign but rather annotate long cells, which are sure to capture the whole sign for

[5] *https://tla.mpi.nl/tools/tla-tools/elan/*

further investigation in the future. In practice this means that the annotated cell starts in the frame in which one of the parameters (usually the handshape, the orientation or the non-manual elements) of the sign is noticeable for the first time, and ends in the frame in which one of those parameters is noticeable for the last time.

The second principle emphasises the sign's phonological parameters. We use information about the four parameters of a sign (the handshape, location, movement and palm orientation) to code structural differences between signs with the same meaning. These differences may be free variation of only one parameter (phonetic variation) or differences between several parameters (lexical variation). With relation to the first option, at this stage of the annotation process we bring out equally all possible phonetic variants of a sign without combining them in the same 'family' as is done in the ID-gloss system (see Johnston & Schembri, 1999; Johnston, 2010; Cormier et al., 2012; Schembri et al., 2013). For example, we append information concerning the handshape, location, movement or orientation of the palm to a gloss, which helps us to distinguish the variants from each other (see Table 1 and 2).

HANDSHAPE	LOCATION
RUN(BB)	SKIN(cheek)
RUN(SS)	SKIN(back of a hand)

Table 1: Phonological parameters handshape and location differentiating between glosses for signs which differ in one parameter.

MOVEMENT	ORIENTATION
ARRANGE(sliding)	FINISHED(palms_down)
ARRANGE(bouncing)	FINISHED(palms_forward)
	FINISHED(palms_backward)

Table 2: Phonological parameters movement and orientation differentiating between glosses for signs which differ in one parameter.

We use information concerning the parameters of signs also when glossing different signs which have the same meaning (lexical variation). In this case we typically choose the most salient parameter for the gloss. E.g., Australia can be signed in at least three different ways in FinSL (see Figures 6–8)[6]. We have chosen the handshape of the signs as the most salient parameter to separate these signs from each other; otherwise the glosses are similar.

[6] Images for three signs meaning 'Australia' taken from KOTUS (2003).

Figure 6: AUSTRALIA(B)

Figure 7: AUSTRALIA(3)

Figure 8: AUSTRALIA(middle-finger)

The third principle concerns the semantics and grammatical phenomena of a sign. Every sign in the data is annotated (given a gloss) according to its meaning in context, although the same form might be used in another meaning in some other context. In this aspect our annotation is at this stage different from traditional ID-glosses (see e.g. Johnston, 2010). For example, the form demonstrated in Figure 9 can refer to the meanings 'everyday', 'jeans', 'countryside', 'sober' and 'redneck'. We annotate such form-meaning pairs in separate glosses, according to the contextual meaning of the occurrences.

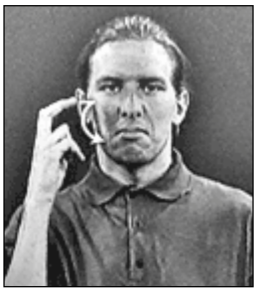


Figure 9: The manual form for form-meaning pairs glossed as EVERYDAY, JEANS, COUNTRYSIDE, SOBER and REDNECK (image from Suvi, the on-line dictionary of Finland's Sign Languages).

On the grammatical level, after the gloss we add codes indicating how the sign is modified morphologically. Grammatical codes (e.g. negation or descriptive utterances) are appended so that linguistic structures can be researched efficiently in the future. The annotation conventions for grammatical phenomena are currently being processed. We are considering using the symbol @ before coding different grammatical features after a gloss (see Wallin & Mesch, 2014). For instance:

- GLOSS@gesture
- GLOSS@depicting-sign
- GLOSS@repetition

During the annotation we have often used a comment tier in ELAN, into which we put different remarks about the modifications of a sign for subsequent work in coding these features in the glosses.

The last principle concerns an annotation as a tool. The aims of our glossing system are efficient search functions and machine readability, which are the same goals as for the ID-glosses used in the corpus work of several other sign languages. However, we are first creating glosses systematically with the help of the three previously mentioned principles, which will help us to build glosses as tools with different purposes in the future. It is important to remember that the glosses which are used in annotations have long-term effects on the research, teaching and learning of a sign language. We need first to test how well the glosses we have used serve different search processes; they must be as logical and usable as possible. We plan to arrange tests of the use of the corpus material in the contexts of teaching (pedagogical view) and research (linguistic view). We thereby hope to achieve a logical and usable glossing system which will serve as many aims as possible.

It is important to strive for consistency, usability and compatibility in the order of the glosses (see Keränen et al., 2016). The examples in Figure 10a demonstrate how a gloss can be a tool for the research and teaching of a sign language. It is much easier for researchers, teachers and students to search for a gloss (e.g. verbal KNOW) and its different structural and grammatical features by appending these features after a verbal of the same form (i.e. in this case the verbal KNOW). For instance, students can learn how the verbal KNOW can be modified in different ways (phonetic variation, prosody, negation), or how two completely different signs can have the meaning 'know' (lexical variation, TIETÄÄ-EI(55) and TIETÄÄ(repetition)). This aim corresponds with the ID-gloss system. If a gloss and its features were not in systematic order, according to a basic form, it would be much more difficult and messy to search for and find a certain gloss from a gloss list (see Figure 10b).

a	b
TIETÄÄ-EI(55) 'not know'	EI-TIETÄÄ(55)
TIETÄÄ-EI(BB) 'not know'	EI-TIETÄÄ(BB)
TIETÄÄ-PALJON 'to know a lot'	PALJON-TIETÄÄ
TIETÄÄ(loiva) 'to know (gentle)'	TIETÄÄ(loiva)
TIETÄÄ(toisto) 'to know (repetition)'	TIETÄÄ(toisto)

Figure 10: Examples of (a) an efficient search according to systematic annotation, (b) how unsystematic annotation may affect the search.

4. Creating a lexical database: Signbank

Our team has processed the glosses from the basic annotation work with the help of two lexical databases. Firstly, during the first round of annotation, we collected all the lexical glosses in Excel, as we did not yet have the FinSL Signbank in use. In this file we all commented on the existing glosses and then modified them as necessary, systematizing and confirming the glosses on the basis of the comments and the above-mentioned four principles of our linguistic concept.

Secondly, since May 2015 we have been working on the FinSL Signbank[7], a web-based lexical database used by researchers to store videos and relevant information about glosses. During the second round of annotation we are gradually transferring the lexicon from Excel to the FinSL Signbank and ultimately we plan to use only the FinSL Signbank. The FinSL Signbank has been developed on the basis of the NGT Signbank[8], which is a branch of the Auslan Signbank[9]. The source codes for these three versions of Signbank are all available on Github (https://github.com/Signbank). Some features of the NGT Signbank that are not necessary for our work at the moment were modified, hidden or deleted from the FinSL Signbank in order to match the current and future needs of the CFINSL project.

Our three main objectives for the use of Signbank in the CFINSL project are to allow two different research teams to upload their data sets, make the user interface translatable into multiple languages, and to be able to export glosses from Signbank to ELAN. With regard to the first objective, our current aim in the CFINSL project is to include the annotated glosses of both FinSL and FinSSL in two different dictionaries inside Signbank. In addition, co-operation between the CFINSL project and the corpus project of the Finnish Association of the Deaf[10] may result in three separate corpus lexicons within the FinSL Signbank. Work on this feature started in January 2016 and is currently in progress.

With regard to the second objective, the interface of the FinSL Signbank is now translatable into multiple languages. The process began during June 2015 and it was done with internationalization and localization features of Django[11], the web framework with which the

[7] *http://signbank.csc.fi*
[8] *http://signbank.science.ru.nl*
[9] *http://www.auslan.org.au*
[10] *http://www.kuurojenliitto.fi/en*
[11] *https://docs.djangoproject.com/en/1.8/topics/i18n/*

Auslan Signbank was built. These internationalization and localization features are needed in order to provide the interface in at least three languages: Finnish, Swedish and English.

Finally, in relation to the third objective, exporting glosses from the FinSL Signbank to ELAN works, but the feature needs further testing so that we can avoid possible problems in the future. In addition, we have added some new functions to the FinSL Signbank interface in order to help annotators' work in the lexical database. One of the new functions is the creation of colour codes for the glosses listed on the search page in Signbank. The listed glosses are automatically given a colour code according to whether the gloss entries include videos, are under evaluation, or have been approved by the administrators.

Signbank is an important tool for annotating new material efficiently and for observing coherent annotation conventions for the FinSL and FinSSL corpora. The use of the FinSL Signbank for annotation purposes began with multiple tests during autumn 2015. At the time of writing, we have begun transferring all the confirmed (i.e. our commonly accepted) glosses from the Excel lexicon into the FinSL Signbank. For the moment, the process of creating lexical entries in the FinSL Signbank follows the annotation conventions and principles described in Section 3 of the current paper. The description of the lexical entries begins with glosses and translation equivalents.

5. Conclusion

In this paper we have described the work of creating corpora of Finland's Sign Languages. This work, which is being carried out at the University of Jyväskylä, is still in its early stages. We have described the process of collecting the data, consents and metadata; the process of annotating the data and developing conventions for the annotation; and the process of building a web-based lexical database for the corpus lexicon. The CFINSL project will document and store both sign languages for present and future generations: the annotation conventions and lexical database will work as a tool for the research, teaching and learning of FinSL and FinSSL.

6. Bibliographical References

Burnard, L. (2005). Metadata for Corpus Work. In M. Wynne (Ed.) *Developing Linguistic Corpora: A guide to good practice. AHDS Guides to Good Practice.* Oxford: Oxbow Books, pp. 30–46.

Cormier, K., Fenlon, J., Johnston, T., Rentelis, R., Schembri, A., Rowley, K., Adam, R., & Woll, B. (2012). From corpus to lexical database to online dictionary: Issues in annotation of the BSL Corpus and the development of BSL SignBank. In *Proceedings of the 5th Workshop on the representation and processing of sign languages: Interactions between corpus and lexicon.* Paris: ELRA, pp. 7–12.

Crasborn, O. & Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In *Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages*. Paris: ELRA, pp. 39–43.

Hanke, T., Matthes, S., Regen, A. & Worseck, S. (2012). Where does a sign start and end? Segmentation of continuous signing. In *Proceedings of the 5th LREC Workshop on the Representation and Processing of Sign Languages.* Paris: ELRA, pp. 69–74.

Jantunen, T. (2013). Signs and transitions: Do they differ phonetically and does it matter? *Sign Language Studies* 13(2), pp. 211–237.

Jantunen, Tommi (2015). How long is the sign? *Linguistics* 53(1), pp. 93–124.

Johnston, T. & Schembri, A. (1999). On defining lexeme in a signed language. *Sign language & linguistics*, 2(2), pp. 115–185.

Johnston, T. (2010). From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics* 15(1), pp. 106–131.

Keränen, J., Syrjälä, H., Salonen, J. & Takkinen, R. (2016). The Usability of the Annotation. To appear in *Proceedings of the 7th Workshop on the Representation and Processing of Sign Languages: Corpus Mining*. Paris: ELRA.

KOTUS (2003). The minutes of a meeting of the Board of Sign Languages of the Institute for the Languages of Finland, 9 October 2003. http://www.kotus.fi/files/509/kokous23-090603.pdf

Mesch, J. (2015). Svensk teckenspråkskorpus – dess tillkomst och uppbyggnad [Building the corpus for Swedish Sign Language]. In *Forskning om teckenspråk XXIV* [Research on sign language XXIV]. Stockholm: Stockholm University, Department of Linguistics, pp. 1–25.

Nishio, R., Hong, S.-E., König, S., Konrad, R., Langer, G., Hanke, T., Rathmann, C. (2010). Elicitation methods in the DGS (German Sign Language) Corpus Project. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies.* Paris: ELRA, pp. 178–185.

Schembri, A., Fenlon, J., Rentelis, R., Reynolds, S. & Cormier, K. (2013). *Building the British Sign Language Corpus*. Language Documentation and Conservation 7, 136–154.

Soininen, Maria (2016). Selvitys suomenruotsalaisen viittomakielen kokonaistilanteesta. Selvityksiä ja ohjeita 2/2016. Oikeusministeriö. http://urn.fi/URN:ISBN:978-952-259-490-7

Wallin, L. & Mesch, J. (2014). *Annoteringskonventioner för teckenspråkstexter. Version 5.* Avdelningen för teckenspråk, Institutionen för lingvistik, Stockholms universitet.

7. Language Resource References

SUVI. The online dictionary of Finland's Sign Languages. The Finnish Association of the Deaf. http://suvi.viittomat.net

A Digital Moroccan Sign Language STEM Thesaurus

Abdelhadi Soudi[1] & Corinne Vinopol[2]
[1]CLC, ENSMR-Rabat, Morocco
[2]Institute for Disabilities Research and Training, Inc., United States
E-mail: asoudi@gmail.com, corinne@idrt.com

Abstract

This paper presents a gesture-based linguistic approach to assisting Moroccan Sign Language (MSL) users in understanding and appropriately using Science, Technology, Engineering and Mathematics (STEM) terminology by creating the first-ever digital MSL STEM Thesaurus. The thesaurus enables Deaf individuals to describe signs and obtain Standard Arabic word equivalents, concept graphics, and definitions in both MSL and Arabic. This is accomplished not only by providing words comparable to signs that they know, but also by providing other information (e.g., signed definitions) that helps differentiate Arabic word choices. The thesaurus is supported by a Concordancer for better illustration and disambiguation of STEM terms. The thesaurus will likely prove to be an invaluable tool that will enable children and adults who rely on MSL for communication, both deaf and otherwise communication impaired, to better understand and write knowledgeably and clearly on STEM topics, and pass standardized assessments.

Keywords: Moroccan Sign Language, Standard Arabic, STEM Thesaurus

1. Rationale and Background

A population that has been underserved in STEM literacy and under-represented in STEM careers is deaf individuals. There are two primary reasons: First, much of the formal scientific information is not in accessible formats. In fact, most scientific information currently available is audio and text-based and without interpretation into sign language. Second, few educators investigate and use research that points to instructional practices that yield best results for deaf students.

This work addresses the challenging research problem of meeting the educational needs of deaf people who are underserved in education, in general, and in STEM literacy, in particular. In Morocco, there is no secondary education for the Deaf. There is a severe lack of instructional resources: all scientific information is Arabic/French audio and text-based and without interpretation into Moroccan Sign Language (MSL). Second, no materials interpret STEM content into MSL, and likewise, there are no resources that interpret STEM from MSL to Standard Arabic. Many deaf people who rely on sign language for communication do not have good facility with Moroccan Arabic (a spoken language) and Modern Standard Arabic (a written and spoken language, used especially in the media and other professional settings). Since sign languages have no written representation as do oral languages, sign languages can only be represented via video, graphics, and animation. As a result, reading achievement scores of deaf individuals usually fall far short of those found among hearing children of comparable abilities. Studies have shown that the average Deaf adult has the literacy competency of a 10-year-old (Traxler, 2000).

Sign languages can only be described, animated, or videotaped. A few researchers have attempted to develop a notation system to describe individual signs.

Stokoe W.C. (1960, 1965) proposed a notation system for ASL, and Lynn Friedman (1977) provided a phonological analysis of ASL. Contrary to the popular belief, Stokoe realized that signs are not just whole entities, but are composed of smaller atomic units. He developed a transcription system based on sign components which he called "cheremes" and equated with the phonemes of spoken languages. Signs can be described by four cheremes: location, hand shape, motion, and orientation. A number of other writing systems have been developed for representing sign languages in written form. These include HamNoSys (the Hamburg Notational System) Thomas Hanke and Constanze Schmaling (1989) and SignWriting developed by Valerie Sutton (1974). These systems, however, are hardly used or recognized by deaf people or their service providers.

For deaf students, multimedia approaches have been found to enhance factual recall as compared to traditional lecture formats. The combined effects of clear signing, use of media, structured lesson material, and, especially, and interactivity have been found particularly important in terms of performance on post-tests. A study by Dowaliby and Lang (1999) showed that the combined use of signs, graphics, text, and adjunct information also resulted in statistically significant gains as compared to the control group (text only). The results of three different studies with Earth Science, Physical Science, and Chemistry conducted by Donald Steely at the Oregon Center for Applied Science (ORCAS), indicated that interactive multimedia and web-based curriculum materials yielded significantly greater knowledge gains for deaf students as compared to traditional classroom experiences. Lang and Steely (2003) found that well-designed, proven-efficacious science instructional programs for hearing students can be successfully adapted for use with deaf students by interspersing text and American Sign Language explanations with content animation and by providing additional practice on vocabulary and content graphic organizers. Diebold, T. J.

& Waldron, M. B. (1988) concluded that the use of highly pictorial content and simplified English text produced significantly higher pre- to post-test gain scores than formats with less pictorial content.

It is in this context that we have created a STEM thesaurus of MSL and a Concordancing software. This assistive technology will help offer equal access and opportunities to STEM education by providing instructional material.

2. Thesaurus description and functionality

The thesaurus enables Deaf users to:

1. Describe a sign by selecting its cheremes from picture menus;
2. Obtain a graphic and video clip of the sign described by the 8 chosen cheremes;
3. Or, obtain an array of signs that most closely match the user's chereme selections (the chereme version of spell-check);
4. Obtain a list of the Arabic words that can be represented by that sign;
5. Obtain concept graphics to help distinguish the Arabic options;
6. See definitions in Arabic (text) and MSL (video) of the Arabic word options, and
7. Identify word forms and their parts of speech.

The thesaurus creation has been done in two phases. To demonstrate feasibility in Phase I, we picked a small sample of STEM terms to see if we could get the thesaurus to work. It was not a statistically significant sample, just one to check functionality of the software. These signs were selected from the database of software previously developed by our research team (i.e., Sign Generator), which includes 3,000 MSL signs (in both graphic and video format) and 8,500 corresponding Arabic words, symbols, and numbers. To do this, we reordered the database by the sign graphic names. In this way, we could easily determine which signs share more than one Standard Arabic word equivalent. Signs that have more than one Arabic word were given preference for our sample. We then identified the 4 cheremes for each hand for each of these signs (i.e., location, hand shape, palm orientation, and motion for dominant and non-dominant hand) and add their corresponding codes into the database. Each of the 250 chosen signs were given 8 codes, 4 for the cheremes of the dominant hand and 4 for the cheremes of the non-dominant hand. That is, we identified all of the variables for each of the cheremes and then developed a coding system which identified each in the database. For example, if there are 44 hand shapes used in MSL, the "A" hand shape is given the code HS-1, the "B" hand shape HS-2, etc.

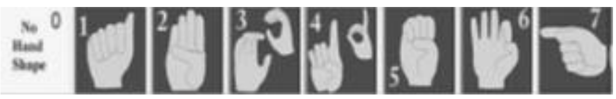


Figure 1: Example of hand shape chereme options and potential codes

Since our existing MSL database did not contain Standard Arabic and MSL definitions for the words that correspond to the 250 selected STEM signs, these were prepared and inputted. Voiceover was also provided for all of the definitions so that it can be appreciated by hearing people who do not have good facility with sign language (e.g., mainstream teachers who do not sign). Audio recording was done separately and then merged with the video before compression. Since the grammars of the two languages (i.e., Standard Arabic and MSL) are divergent, merging them required expertise in both languages (a multidisciplinary deaf and hearing research team).

As is shown in Figure 2, the MSL Thesaurus operates by having MSL users identify the four cheremes for each hand for the STEM sign for which they want to find Arabic equivalents by using drop-down pictorial menus. The program searches the database for the sign that most closely matches the chereme choices. If the cheremes selected do not exactly match how the sign is coded in the database, the program will provide options of signs that are described similarly (the chereme version of spell-check). Accurately described signs and sign options are displayed as graphics and videos. Once users verify their intended sign, the program will display, in addition to the sign graphic and video, the comparable Arabic word(s) and the Arabic word definition(s) in text, MSL definition(s) in video, concept graphic(s), and word forms. This will help deaf students discern, when writing on STEM topics, which Arabic word to use for their sign.

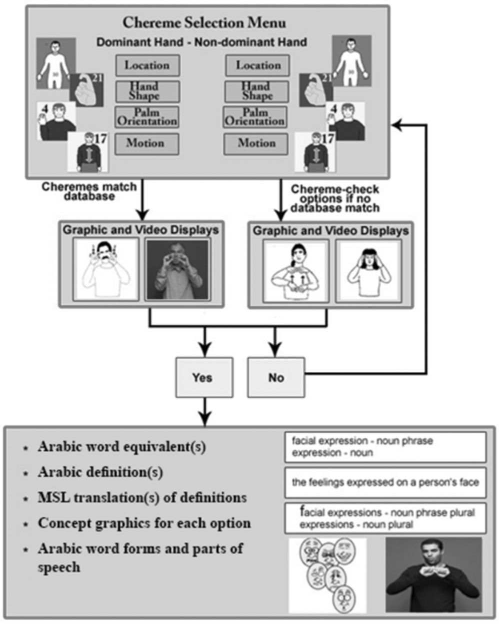


Figure 2: Software navigational functionality
(The sign for "facial expression" is used as an example)

Figure 3 below shows an example resulting from the choice of the 8 cheremes corresponding to the term "friction."

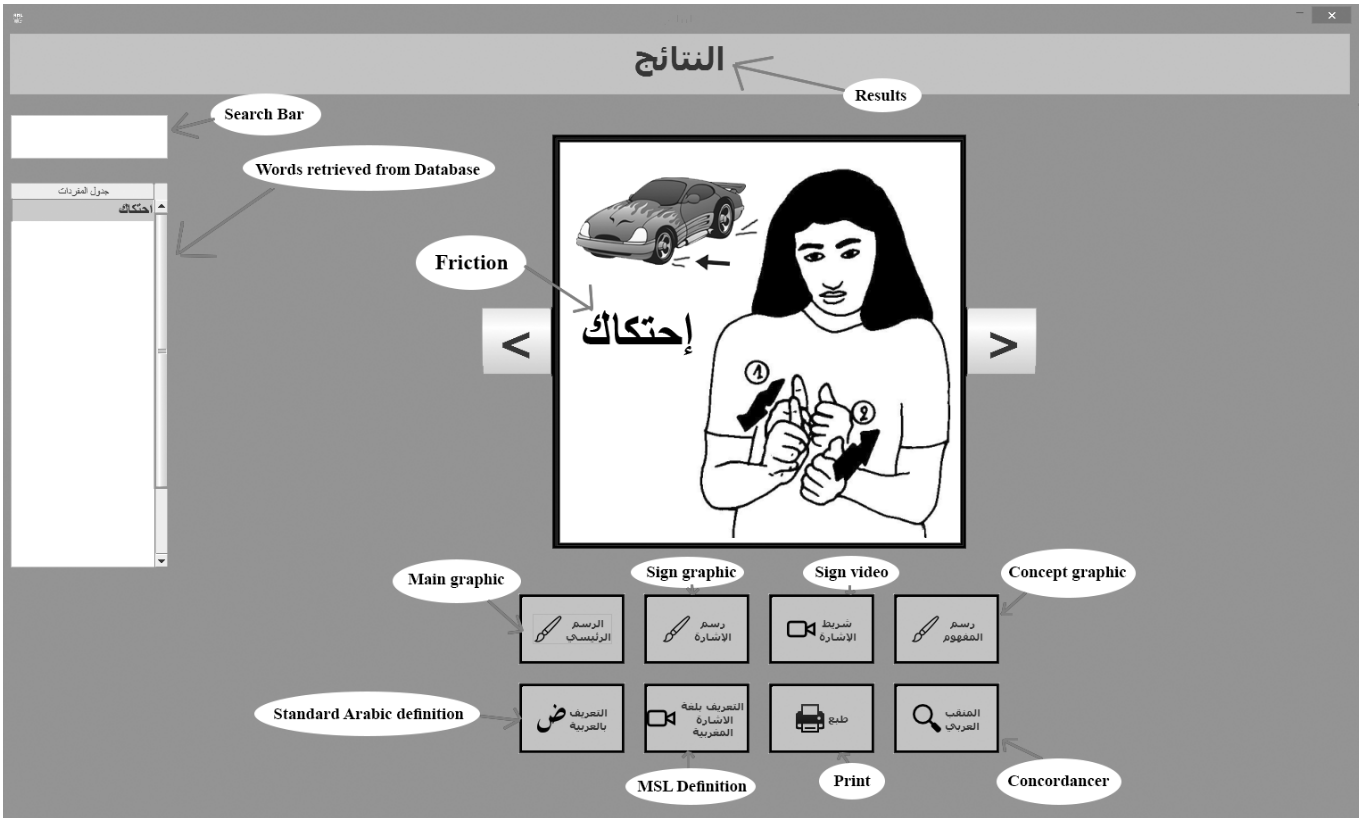


Figure 3: Example of a graphic sign (STEM term "friction") described by the 8 chosen cheremes

Users can choose a variety of output options for the depicted MSL sign (1. Main Graphic Sign which includes the corresponding Standard Arabic term, the concept and the Graphic sign, (2. MSL video clip of the sign, (3. MSL definition, Standard Arabic definition, the concept graphic). As can be seen in Figure 3, the thesaurus is also supported by a Concordancer for a better illustration and disambiguation of STEM terms. This tool provides a list of examples of a particular term or combination of terms, in its/their contexts drawn from a science corpus. By clicking on the Concordancer button, users can are invited to the Concordancer window and are provided a variety of options for searching examples of how the already selected term- "friction" in this case- is used. In order to enable users to search also for other possible inflected and derived forms of a STEM word, we have incorporated Arabic Morphological Analysis in the Concordancer. Arabic morphology/word formation represents a special type of morphological system. It is considered to be a non-concatenative morphology which depends on manipulating root letters in a non-concatenative manner, using different operations such as gemination and infixation. Arabic morphology requires infixation, prefixation and suffixation, giving rise to a large space of morphological variation. Stems are formed by a derivational combination of a root morpheme and a vowel melody; the two are arranged according to canonical patterns. For example, the Arabic stem katab (he wrote) is composed of the morpheme ktb (notion of writing) and the vowel melody morpheme 'a-a'. The two are coordinated according to the pattern CVCVC (C=consonant, V=vowel). This means that Arabic word structure is not built linearly as is the case in concatenative morphological systems. The language has a large degree of ambiguity in word senses, and further ambiguity attributable to a writing system that omits diacritics. (e.g., short vowels, consonant doubling, inflection marks). For example, "ktb" can correspond to kataba «he wrote », kutiba "was written" kutub "books", or 18 other forms). Accordingly, we used a tool that provides all the possible readings/analyses of an inputted word in Arabic. For such a task, we used Buckwater's Arabic Morphological Analyzer (BAMA) (Bucwalter 2002). In BAMA, the data consists primarily of three Arabic-English lexicon files: prefixes (299 entries), suffixes (618 entries), and stems (78, 839 entries)). The tool is based on a concatenative lexicon-driven approach. In (Soudi et al., 2007), we provide a detailed study of Arabic morphological issues.

As is shown in Figure 4, users can choose to see usage examples of the selected term in three ways: 1. examples showing exact match of the term (Figure 4.1), 2. morphological analysis: searching also for different word forms of the selected term. (Figure 4.2), 3. proximity: restricting the search by requiring contexts in which the selected term is adjacent to another specific term chosen by the user (Figure 4.3).

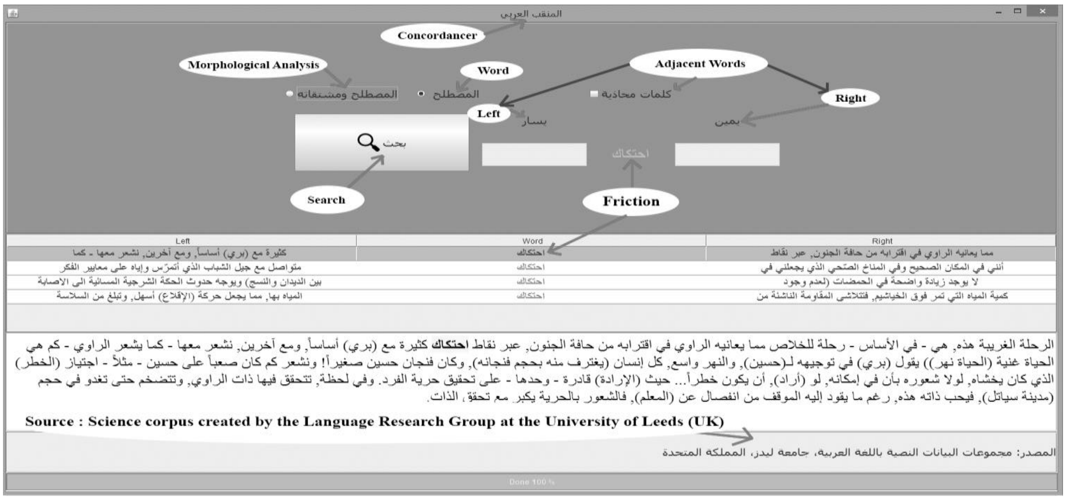


Figure 4.1: Concordancing with exact match search

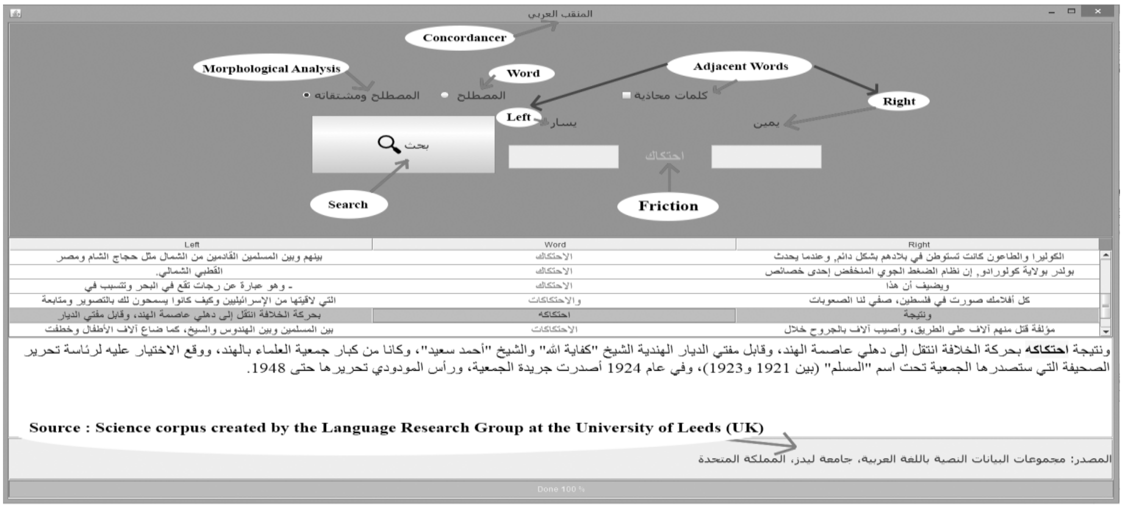


Figure 4.2: Concordancing with morphological analysis

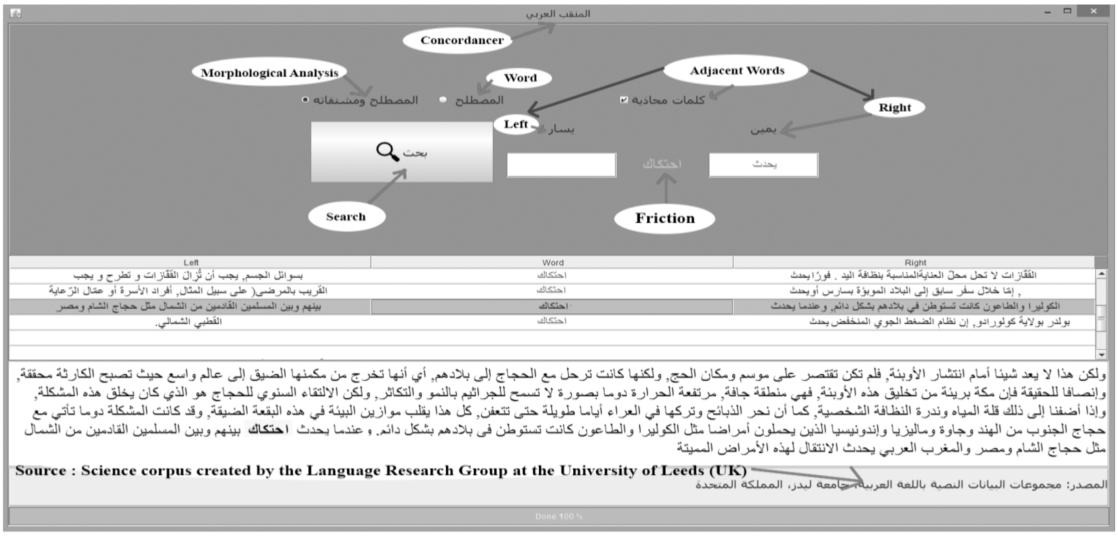


Figure 4.3: Concordancing with proximity

Figure 4: STEM term concordancing options

3. Evaluation of the Thesaurus and its Enrichment

Before proceeding to Phase II (Thesaurus enrichment), some preliminary feedback from typical users was necessary. Twenty deaf and hearing users (including Deaf educators) were invited to test the software in our Sign Language Lab. Ten users were asked to depict 100 STEM terms from four chereme menus to describe the intended sign. Each STEM term had to be depicted within some time frame. The users were able to quickly get exact graphic signs and video clips for 65 STEM terms. An analysis of the chereme variables selected by the users for the other unidentified 35 STEM terms shows that either an exact match is not found, or the user doesn't approve the returned sign and word. This helped us enhance the software so that in cases where the corresponding data is not found or the user does not approve it, the algorithm will return a set of signs that closely match the intended sign. Once the user has approved the sign, the Arabic word equivalent(s), definition(s) (in Standard Arabic and MSL) etc., are fetched from the database and displayed back to the user. In order for us to evaluate the navigational functionality of the software and search efficacy, the other 10 users were given the freedom to describe as many signs (corresponding to STEM terms) as possible by selecting their cheremes from picture menus.

Currently, more in-depth clinical and typical setting usability and efficacy evaluations are being addressed: develop lab observation protocol and usability protocol and arrange evaluation logistics.

The Thesaurus's improvement at the level of both data and navigational features is an ongoing process, and to date, 500 STEM signs are in the database.

4. Conclusion

In this paper, we have described a digital MSL STEM Thesaurus that enables MSL users to describe signs for STEM concepts that they know and use, and find Arabic word equivalents, parts of speech, definitions (in Arabic text and MSL video), and conceptual pictures to help disambiguate meanings. This assistive technology tool will help deaf and hard of hearing students to better understand the nuances of STEM terminology and foster improved written expression to respond to lessons and assessments of STEM content. This is accomplished not only by providing words comparable to signs that they know, but also by providing other information (e.g., signed definitions) that helps differentiate Arabic word choices.

5. Bibliographical References

Al-Sulaiti, L. and Atwell, E. (2006). The design of a corpus of contemporary Arabic. International Journal of Corpus Linguistics, vol. 11, pp. 135-171. [corpus freely available at http://www.comp.leeds.ac.uk /eric/latifa/research.htm].

Battison, R., 1978. Lexical Borrowing in American Sign Language. Linstok Press, Silver Spring, MD (Reprinted 2003, Linstok Press).

Diebold, T. J. and Waldron, M. B. (1988). Designing instructional formats: The effects of verbal and pictorial components on hearing-impaired students' comprehension of science concepts. American Annals of the Deaf, 133, 30-35.

Dowaliby, F., & Lang, H. (1999). Adjunct aids in instructional prose: a multimedia study with deaf college students. Journal of deaf studies and deaf education, 4(4), 270-82. doi:0.1093/deafed/4.4.270.

Friedman, L. A. (Ed.). (1977). Formational properties of American Sign Language. New York: Academic Press.

Lang, H. G., & Steely, D. (2003). Web-based science instruction for deaf students: What research says to the teacher. Instructional Science, 31,277–298.

Ramsey, C., & C. Padden (1998). Natives and newcomers: Gaining access to literacy in a classroom for deaf children. Anthropology and Education Quarterly. 29, 5-24

Singleton, J., Supalla, S., Litchfield, S., & Schley, S. (1998). From sign to word: Considering modality constraints in ASL/English bilingual education. Topics in Language Disorders, 18 (4).

Soudi (2013a). A Robust Technology facilitating interaction between the Hearing Culture and the Deaf Culture. In Proceedings of the session, Interculturalism, Migration and Multiculturalism of the conference "At the Crossroad between Africa, Asia and Europe: Challenges and Perspectives», Venice, Italy, 29 May-01 june 2013.

Soudi, A. Vinopol, C. and Er-rady, A. (2013b). Standard Arabic-to-Moroccan Sign Language Machine Translation: Does translation from a morphologically rich language reflect the interpreters' allocation of high cognitive resources? In Proceedings of The Fourth Conference of the Scandinavian Association for Language and Cognition (SALC IV), University of Eastern Finland, Finland, 12-14 June 2013.

Soudi A., Van den Bosch A. and Neumann G (2007). *Arabic Computational Morphology: knowledge-based and Empirical Methods*. Editors: Soudi A., Van den Bosch A. and Neumann G. In Kluwer/Springer's series on Text, Speech, and Language Technology (series editors Nancy Ide and Jean Veronis).

Stokoe William, Dorothy Casterline, and Carl Croneberg. (1965). A dictionary of American Sign Language. Washington, D.C., Gallaudet College.

Stokoe, William C., Jr. (1960). Sign language structure; an outline of the visual communication systems of the deaf. S.I.L.:O.P.8.

Strong, M., & Prinz, P. (1997). A study of the relationship between American Sign Language and English literacy. Journal of Deaf Studies and Deaf Education, 2(1), 37-46.

Sutton, Valerie. (1974). Sign Writing.

Thomas Hanke and Constanze Schmaling, (1989). HamNoSys, Version 3.0. University of Hamburg, http://www.sign-lang.uni-hamburg.de/Projects/HamNoSys.html.

Tim Buckwalter. (2002). Buckwalter Arabic Morphological Analyzer Version 1.0. Linguistic Data Consortium, catalog number LDC2002L49 and ISBN 1-58563-257-0. http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2002L49

Traxler, C.B. (2000). The Stanford Achievement Test, 9th edition: National norming and performance standards for deaf and hard of hearing students. Journal of Deaf Studies and Deaf Education, 5, 337–348.

Wilbur, R. and Adamo-Villani, N. (2006-2011). Software for Math Education for the Deaf, NSF award number: 0622900.

World Federation of the Deaf and Swedish National Association of the Deaf's report(2008): Global Education Pre-Planning Project on the Human Rights of Deaf People.

Online Concordancer for the Slovene Sign Language Corpus SIGNOR

Špela Vintar, Boštjan Jerko
University of Ljubljana
Faculty of Arts, Aškerčeva 2, Ljubljana
E-mail: spela.vintar@ff.uni-lj.si, bostjan.jerko@guest.arnes.si

Abstract

We present the first version of an online concordancing tool for the Slovene Sign Language SIGNOR corpus. The corpus search tool allows querying the SIGNOR annotated database by glosses and displays the hits in a keyword-in-context (KWIC) format, accompanied by frequency information, HamNoSys transcription and metadata. The main purpose of the tool is linguistic research, more specifically sign language lexicography, but also providing general public access to the corpus.

Keywords: SIGNOR, sign language corpus, online concordancing, corpus search tool

1. Introduction

Slovene Sign Language (SZJ) is the primary language of the Deaf community in Slovenia comprising between 1000 and 1500 users. Within a 3-year research project, a corpus of SZJ was compiled by collecting video samples from 80 informants, which were then transcribed and annotated at several levels of analysis[1]. In this paper we present an online concordancing tool which can be used to query the corpus annotations, explore sign frequencies and view signs within the authentic conversational context.

Online corpus interfaces for sign language corpora are scarce. We are familiar with searchable sign databases such as the Lexical Database of Sign Language in Klagenfurt[2], the Auslan Signbank[3] (Johnston 2001) and the BSL Signbank[4] (Cormier et al. 2012), and the open access online corpus of movies representing Dutch Sign Language (NGT) (Crasborn and Slöetjes 2014). Sign databases are inventories of signs which do not provide contextual information and cannot replace sign language corpora, where authentic conversations have been recorded and annotated. The Dutch NGT corpus is based on the ELAN corpus annotation workbench [5] for multimodal corpora, and the multi-tier search functionality is provided by the TROVA search engine. The main problem with representing sign language in an online querying environment is the potential complexity of queries; sign language corpora typically contain multi-layered annotations where different types of data (glosses, timecodes, audiovisual data, metadata etc.) overlap and are difficult to present in a user-friendly manner. Furthermore, existing query tools rely on corpus annotation workbenches (ELAN or iLex) which are not easily portable into a web environment and usually require dedicated browsers.

Our aim was to create a simple web interface where the corpus could be searched from any browser, however our tool currently does not support complex or multi-layer queries.

2. The SIGNOR Corpus

The compilation of the corpus started in 2011. Preliminary considerations involved issues of regional balance, the informants' competence in SZJ, text types, communicative settings, and elicitation techniques, as well as technical issues regarding the recording sessions and video processing. Having reviewed several related projects, our methodology of video session organization relied on Nishio et al. (2010), and the segmentation and annotation strategies were also mostly adapted from the German DGS project (Hanke et al. 2012; Konrad et al. 2012).

All of the recordings were converted into a common data format and stored on the project data server. For corpus annotation we used the iLex tool (Hanke and Storz 2008), which provides a flexible multiuser annotation environment and stores all signs, lexemes, and tokens in a database, thus facilitating consistency between annotators.

Annotation includes the following layers (Vintar et al. 2012, Vintar 2015):

- Tokenization. The video stream of signed dialogue is segmented into individual signs delimited by time codes.
- Glossing. The process of assigning each sign a lexical identifier is also referred to as lemmatization; in other words, each token is assigned a type.
- Mouthing. The voiceless or voiced articulations accompanying signs may constitute, reinforce, or alter their meaning.
- Meaning. Each sign is assigned its meaning in the given textual context.
- Compound meaning. Many signs are compositional or phrasal, and the meaning of such multisign units is annotated as a separate tier.

[1] http://www.lojze.si/signor/index.html
[2] http://ledasila.uni-klu.ac.at/TPM/
[3] http://www.auslan.org.au/dictionary/
[4] http://bslsignbank.ucl.ac.uk/dictionary/
[5] http://tla.mpi.nl/tools/tla-tools/elan/

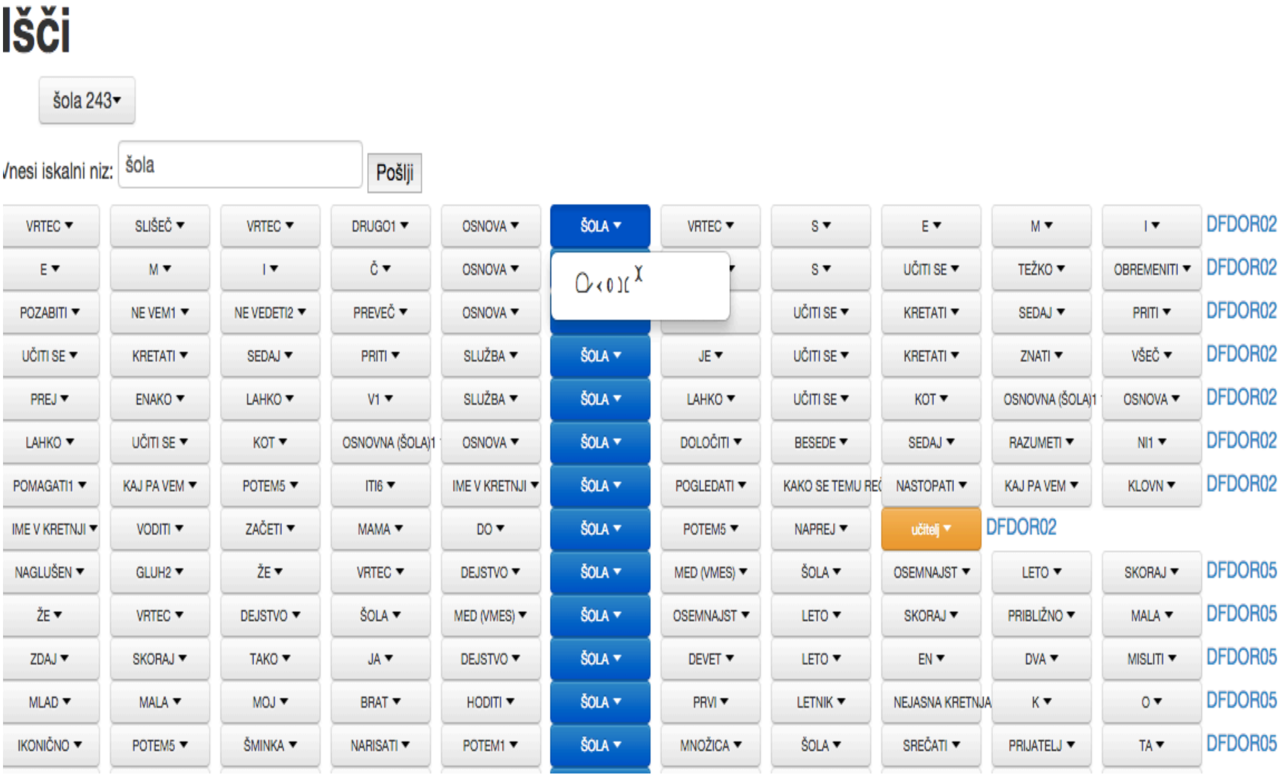


Figure 1: The SIGNOR search interface

- HamNoSys transcription (Schmaling and Hanke 2001). The graphical notation of signs helps distinguish sign variants and represents an important step for further processing or sign generation with animated agents.
- Segmentation into utterances. This step was performed on a section of the corpus comprising 3,000 utterances. Each utterance boundary is marked with a specific gloss indicating its form.

The overall length of recordings amounts to approximately 40 hours. The final size of the annotated corpus is 30,335 sign tokens and 2,976 sign types. Of the latter, 1,043 signs occur only once in our corpus. A lexical analysis of the corpus revealed that the frequencies of lexical categories roughly correspond to other sign language corpora (Vintar 2015). SZJ is rich in variants – up to 9 different variants have been found for the same sign, and it seems that variation occurs between different age groups, places of education and geographical regions.

3. The SIGNOR Concordancer

The aim of the search interface was to enable researchers, interpreters and SZJ users to explore signs in context and to compare the frequencies of various lexical items, including potential region- or age-related variants. The concordance line is composed of individual glosses, whereby compound signs are glossed with their complex meaning and marked in a different colour. For each concordance line the interface also displays the anonymized metadata upon click: Gender, Region, Level of Deafness, Education and Primary Hand of the informant.

As a main storage of all the data iLex uses PostgreSQL. For ease of access to the data we decided to access the database directly so we can automate the process of exporting annotations as needed. The concordancer uses MongoDB database for easier usage for online purposes. The data exported from PostgreSQL as CSV were imported to the concordancer database. The CSV files consist of an index of signs with associated data (start and end time codes, gloss, Hamnosys and compound meaning, if applicable). After that, a script is used to import all the data to MongoDB and another script to compute the frequencies of different signs.

The interface is simple and intuitive, providing a single search window to enter the query. The resulting concordance displays the search gloss in context in a keyword-in-context (KWIC) format, with a default window of +/-5 adjacent glosses. Compound meanings are written in lowercase and coloured orange so as to indicate that the sign is compositional. The frequency of the search gloss is displayed on top of the concordance window (Figure 2). A click on any gloss reveals the HamNoSys notation. If the user clicks the Hamnosys notation, an avatar is shown in a separate window signing the selected sign. We are currently using the avatar engine integrated into iLex (Figure 3).

Probably the most useful feature for the purposes of sign language lexicography, teaching or sociolinguistic research is the information on the frequency of sign variants. Thus, a query for AMERIKA ("America") will result in a KWIC display of all variants of the sign for AMERIKA, but the concordance can be filtered by

Figure 2: Sign variants

sign variants of which frequencies are displayed in a drop-down menu (Figure 2). Using filtering and the metadata links displayed for each concordance line, the user may draw conclusions on the distributional properties of each sign variant.

Figure 3: Sign animation generated from HamNoSys

4. Conclusion

This paper describes a simple online concordance tool for the SIGNOR corpus of Slovene Sign Language (SZJ). While advanced annotation tools such as ELAN or iLex allow for detailed and sophisticated queries of multimodal corpora, they are restricted to their own software environment and often too complex for the general public. Our purpose was to create an interface accessible to anyone, including Deaf people, sign language interpreters, teachers and students. It is also a good way of spreading the awareness about sign language among linguists and language policy makers. Our tool should be seen as work in progress as it has been developed within a very small nationally funded project, and the funding of future activities has not been secured yet.

Still, we plan to implement other features to better respond to the needs of potential users. One important future plan is to include the authentic video recordings into the online corpus, but currently we are still resolving legal issues related to data protection and have not obtained full permissions for the public release of all videos. Another improvement we plan is to include HamNoSys notations as a possible query type, so that users might have the possibility to search by signs or sign elements. Several technical improvements are also underway, including caching frequent searches for faster retrieval and optimizing for mobile access.

Acknowledgement

This work was partly funded by the Slovene Research Agency (ARRS) grant J6-4081, 2011-2014.

Bibliography

Cormier, Kearsy, Jordan Fenlon, Trevor Johnston, Ramas Rentelis, Adam Schembri, Katherine Rowley, Robert Adam, and Bencie Woll. 2012. From corpus to lexical database to online dictionary: Issues in annotation of the BSL Corpus and the development of BSL SignBank. *5th Workshop on the Representation of Sign Languages: Interactions between Corpus and Lexicon [workshop part of 8th International Conference on Language Resources and Evaluation*, Turkey, Istanbul LREC 2012. Paris: ELRA. pp. 7–12.

Crasborn, Onno, and Han Slöetjes. 2014. Improving the Exploitation of Linguistic Annotations in ELAN. In N. Calzolari, K. Choukri & et al. (Eds.), *Proceedings of LREC 2014, Ninth International Conference on Language Resources and Evaluation*. Paris: ELRA.

Hanke, T., and J. Storz. 2008. iLex: A Database Tool For Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In *Proceedings of the Language Resources and Evaluation Conference 2008, May 28–May 30.* Paris: ELRA.

Hanke, T., S. Matthes, A. Regen, and S. Worseck. 2012. Where Does a Sign Start and End? Segmentation of Continuous Signing. In *Proceedings of the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon Language Resources and Evaluation Conference (LREC), Istanbul,* ed. O. Crasborn, E. Efthimiou, E. Fotinea, T. Hanke, J. Kristoffersen, and J. Mesch, 69–74. Paris: ELRA.

Johnston, Trevor. 2001. The lexical database of Auslan (Australian Sign Language). *Sign Language & Linguistics* 4.1-2 (2001): 145-169.

Konrad, R., T. Hanke, S. König, G. Langer, S. Matthes, R. Nishio, and A. Regen. 2012. From Form to Function: A Database Approach to Handle Lexicon Building and Spotting Token Forms in Sign Languages. In *5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon Language Resources and Evaluation Conference (LREC), Istanbul,* ed. O. Crasborn, E. Efthimiou, E. Fotinea, T. Hanke, J. Kristoffersen, and J. Mesch, 87–94. Paris: ELRA.

Nishio, R, S.-E. Hong, S. König, R. Konrad, G. Langer, T. Hanke, and C. Rathmann. 2010. Elicitation

Methods in the DGS (German Sign Language) Corpus Project. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies,* 178–85. Paris: ELRA.

Schmaling, C., and T. Hanke. 2001. *HamNoSys 4.0*. http://www.sign-lang.uni-hamburg.de/Projekte/HamNoSys/HNS4.0/englisch/HNS4.pdf, accessed October 13, 2014.

Sloetjes, H., and Wittenburg, P. 2008. Annotation by category – ELAN and ISO DCR. In: *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008).* Paris: ELRA.

Vintar, Š., Jerko, B. and Kulovec, M. 2012. Compiling the Slovene Sign Language Corpus. In *5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon Language Resources and Evaluation Conference (LREC), Istanbul,* ed. O. Crasborn, E. Efthimiou, E. Fotinea, T. Hanke, J. Kristoffersen, and J. Mesch, 87–94. Paris: ELRA.

Vintar, Š. 2015. Lexical Properties of Slovene Sign Language: A Corpus-Based Study. *Sign Language Studies*, *15*(2), pp.182-201.