# Workshop Proceedings

# $5^{th}$ Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon

## Language Resources and Evaluation Conference (LREC)
## Istanbul, May 2012

## Editors and Workshop Organizers

| | |
|---|---|
| Onno Crasborn | Radboud University, Nijmegen NL |
| Eleni Efthimiou | Institute for Language and Speech Processing, Athens GR |
| Evita Fotinea | Institute for Language and Speech Processing, Athens GR |
| Thomas Hanke | Institute of German Sign Language, University of Hamburg, Hamburg DE |
| Jette Kristoffersen | Centre for Sign Language, University College Capital, Copenhagen DK |
| Johanna Mesch | Stockholm University, Stockholm SE |

## Workshop Programme Committee

| | |
|---|---|
| Richard Bowden | University of Surrey, Guildford GB |
| Penny Boyes Braem | Center for Sign Language Research, Basel CH |
| Annelies Braffort | LIMSI/CNRS, Orsay FR |
| Christophe Collet | IRIT, University of Toulouse, Toulouse FR |
| Helen Cooper | University of Surrey, Guildford GB |
| Kearsy Cormier | Deafness Cognition and Language Research Centre, London GB |
| Onno Crasborn | Radboud University, Nijmegen NL |
| Eleni Efthimiou | Institute for Language and Speech Processing, Athens GR |
| Evita Fotinea | Institute for Language and Speech Processing, Athens GR |
| John Glauert | University of East Anglia, Norwich GB |
| Thomas Hanke | Institute of German Sign Language, University of Hamburg, Hamburg DE |
| Alexis Heloir | German Research Centre for Artificial Intelligence, Saarbrücken DE |
| Jens Heßmann | University of Applied Sciences Magdeburg-Stendal, Magdeburg DE |
| Matt Huenerfauth | City University of New York, New York US |
| Trevor Johnston | Macquarie University, Sydney AU |
| Reiner Konrad | Institute of German Sign Language, University of Hamburg, Hamburg DE |
| Jette Kristoffersen | Centre for Sign Language, University College Capital, Copenhagen DK |
| Lorraine Leeson | Trinity College, Dublin IE |
| Petros Maragos | National Technical University, Athens GR |
| Johanna Mesch | Stockholm University, Stockholm SE |
| Carol Neidle | Boston University, Boston US |
| Christian Rathmann | Institute of German Sign Language, University of Hamburg, Hamburg DE |
| Adam Schembri | National Institute for Deaf Studies and Sign Language, La Trobe University, Melbourne AU |
| Meike Vaupel | University of Applied Sciences, Zwickau DE |

# Table of contents

# Author Index

# Editors' Preface

This collection of papers stems from the Fifth Workshop on the Representation and Processing of Sign Languages, held in May 2012 as a satellite to the Language Resources and Evaluation Conference in Istanbul. 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 third time, the workshop had sign language corpora as its main topic. This time, the focus was on the interaction between corpus and lexicon. More than half of the papers presented contribute to this topic. 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

The Editors

# SIGNSPEAK Project Tools: A way to improve the communication bridge between signer and hearing communities

**Javier Caminero[1], Mari Carmen Rodríguez-Gancedo[1], Álvaro Hernández-Trapote[2], Beatriz López-Mencía[2]**

[1] Telefónica R&D, Madrid, Spain

[2] Universidad Politécnica de Madrid, Spain

Email: fjcg@tid.es, mcrg@tid.es, alvaro@gaps.ssr.upm.es, beatriz@gaps.ssr.upm.es

**Abstract**

The SIGNSPEAK project is aimed at developing a novel scientific approach for improving the communication between signer and hearing communities. In this way, SIGNSPEAK technology captures the video information from the signer and converts it into text. To do that, SIGNSPEAK consortium has devoted great efforts to the creation and annotation of the RWTH-Phoenix corpus. Based on it, a multimodal processing of the captured video is carried out and the resultant sign sequence is translated into natural language. Afterwards, the intended message could be communicated to hearing-able people using a text-to-speech (TTS) engine. In the reverse way, speech from hearing-able people would be transformed into text using Automatic Speech Recognition (ASR) and then the text would be processed by virtual avatars able to compose the suitable sign sequence. In SIGNSPEAK project, scientific and usability approaches have been combined to go beyond the state-of-the-art and contributing to suppress barriers between signer and hearing communities. In this work, a special stress was put in the development of a prototype and also, in setting of the grounds for future real industrial applications.

**Keywords:** tool development, sign-language-to-text, user evaluation

## 1. Introduction

Communication for Deaf community is based on sign language since it is “the only language Deaf people can acquire effortlessly and spontaneously when given the right input” (Wheatley and Pabsch, 2010). Unfortunately, deaf and hard of hearing signers have serious limitations for communicating with people without no sign-language skills and thus, the integration into educational, social and work environments is not complete.

Although the mother tongue is defined as the first language that one has acquired, for the deaf community, it is more complex than that, then only a small percentage of deaf children acquire a sign language naturally and in similar stages as hearing children do with a spoken language.

Taking into account these peculiarities, we realize that deaf people usually find numerous barriers in communication. Some of these barriers include the presence of an operator (which may be seen as intrusive and do not represent parity with hearing people), slow communications connections and lack of awareness of how to communicate with people who are deaf or have speech difficulties. About these issues, recent studies (Market Research, 2011) reveal that deaf people and individuals with speech difficulties need freely-accessible services and equipment to ensure that their communication needs are totally fulfilled.

Having all these issues into consideration, SIGNSPEAK project[1] is aimed to provide deaf people a communication bridge between signers and hearing communities. Thus, a new vision-based technology for translating continuous sign language into text is being developed. For that purpose, it has been needed the creation of RWTH-Phoenix, a suitable video corpus for data-driven automatic sign language processing (Stein et. al., 2010). As a consequence of the automation of the services and applications provided by the SIGNSPEAK technology, users’ privacy feeling and their confidentiality in the communication process would be improved.

[1] http://www.signspeak.eu

## 2. SIGNSPEAK: establishing a new communication bridge

As it is showed in Figure 1, SIGNSPEAK technology captures the video information from the signer and converts it into text. In order to do that, a multimodal processing of the video is carried out and afterwards, the resultant sequence of signs is translated into natural language. Using a text-to-speech (TTS) engine, the intended message is communicated to people who are able to hear. In the reverse way the speech from hearing community is captured and translated into text (Automatic Speech Recognition-ASR). Then the text is used by virtual signers (avatars) which compose the suitable sequence of signs.

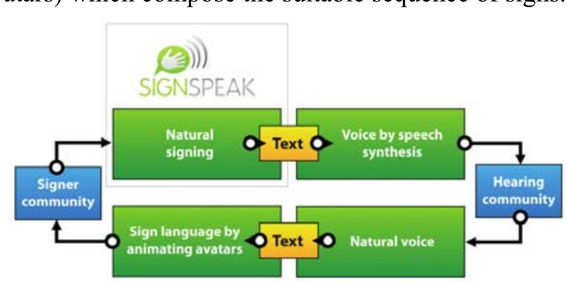



Figure 1. Communication bridge between Signer and Hearing communities

### 2.1 Text-to-Speech

A text-to-speech (TTS) synthesiser can be defined as a piece of software which transforms into speech any input sentence in text format (Dutoit, 1997). This functionality makes a TTS very useful for communication systems because it avoids pre-recording every sentence or words planned to be used in a service. There is a wide availability of products, i.e. Loquendo TTS, Nuance Vocalizer or Festival.

Despite of the great performance of the aforementioned systems, there are yet critics to the use of TTS for certain applications due to: pronunciation of new and rare words (Spiegel, 2003), prosody (Hirschberg, 2002) or limited availability in certain new languages.

### 2.2 Automatic Speech Recognition

The human voice is generated by the vibration of the vocal cords. The vibration of the cords moves the air and these variations of pressure arrive to the listener's ear. Then the pressure waves are transformed into a signal that is processed by the brain and properly interpreted. The acoustic features of this signal allow the listener to differentiate one sound from another, and that is what an Automatic Speech Recogniser (ASR) tries to accomplish.

Some of the most relevant actors in the development of this technology are: CMU Sphinx, RWTH ASR, Dragon Naturally Speaking or Microsoft Speech API.

However, the performance of an ASR system usually depends drastically on external factors (Acero, 1992): input level, additive background noise, channel distortion, etc.

### 2.3 Signing Avatars

Recently, the virtualization of everyday life and the gaming industry has promoted a great development of the virtual characters field. The improvement of several communication technologies as the automatic speech recognition or the text-to-speech engines makes it real to create virtual agents able to interact with users. The benefits are obvious: cheaper customer service and 24/7 availability. Furthermore, through this kind of interfaces, users could establish relationships close to those ones between humans (Reeves and Nass 1996).

Some applications which use avatars or a sort of them for representing Sign Language are Sign Smith Studio, Sys Consulting, ViSiCAST, eSIGN, DePaul ASL Project, SignSynth, TEAM, etc.

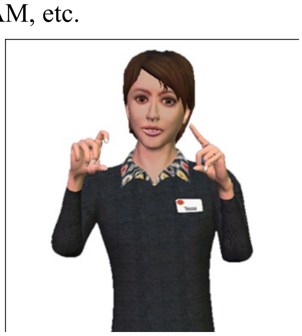


Figure 2. ViSiCAST signing avatar

### 2.4 Sign Language to Text

The most challenging technology included in the communication bridge proposed by SIGNSPEAK is the translation of Sign Language into text. That is, capture the movements, expressions and emotions of the signers; identify the signs from the extracted features, and then translate the sequence of them into natural language in order to obtain a message understandable by hearing users.

The means used to capture hand movements can be classified mainly in two groups: instrumented and video-based. For instrumented proposals, gloves are usually complemented with other devices, as accelerometers. It means users have to remain close to the radiant source, in the case of a wireless connection, or close and physically tethered to the computer in the case of a wired one. Furthermore, current glove technology is not intended for daily use; the gloves deteriorate quickly with extended use and output becomes increasingly noisy as they break down. In the other hand, this kind of solution uses to be more reliable, overall against ambient noise or other adverse background conditions. In video-based approaches, the signer avoids having attached to hers/his body any instrumentation. However, the working conditions should be controlled and the amount of data obtained, compared with instrumented systems, is lower.

SIGNSPEAK project wants to go beyond most of the limitations previously presented. The project follows a global planning approach to transfer the technology to the daily life of deaf community and its scope implies advances in several research fields and the need of taking into account the industrial perspective.

## 3. Main Technological Factors

The SIGNSPEAK project is intended to be a first step to achieve a sophisticated technology able to complete the communication bridge between hearing and deaf community. In this preliminary stage, the demands about the performance of the technology should be ambitious but bearing in the mind the possible problems which could arise in a realistic scenario. Thus, for a proper operation of the technologies involved in the communication bridge (SIGNSPEAK, TTS, ASR, avatars) different user and environmental factors and some technological limitations need to be considered. Next, we point out some of them, however for more detailed information refer to (Gancedo, Caminero and Van Kampen, 2011).

### 3.1 User factors

User factors are individual differences that include demographic variables and situational variables that account for differences attributable to circumstances such as experience and training (Agarwal and Prasad, 1998).

Some user factors which could be relevant to SIGNSPEAK technology might be:

- *Gender*. Research has shown that there are differences between men and women regarding the cognitive structures employed during the interaction with technology products (Venkatesh and Morris, 2000). For example regarding signing avatars, in (Bailenson and Yee, 2005) is showed how users prefer avatars which are similar to them and mimic their behavior. Thus, it could be suggested that signing avatars should mimic the signing style of users or even adopt the users' gender.

- *Experience with technology*. For automatic speech recognition (ASR), in (Karat et al., 2000) is described an experiment where the ASR performance is worst for novice users than for expert ones. Furthermore, the latter group of users is more effective carrying out the corrections when the system fails.
- *Age*. From the point of view of acceptance of technology, age is recognized as a key factor. Specially, senior users, who do not usually have great experience with technology and have age-related problems with cognitive abilities, face difficulties understanding and interacting with technological devices (Ziefle and Bay, 2008). On the contrary, older users are more inclined to accept technologies when the usefulness is clear and there is a good support of the system (tutorials, help system, etc.) (Arning and Ziefle, 2010).
- *Cultural background.* For ASR engines, the problematic issue is quite clear: the accent. In (Huang et al., 2001) the accent was identified as one of the principal components of speech variation.
- *Other factors*. Many more user factors could affect the acceptance of a new communication paradigm (i.e. SIGNSPEAK's communication bridge). For example the level of signers' expressiveness, the users' emotional state or the users' physiology.

### 3.2 Environmental factors

The conditions relative to the context where the interaction is performed are collectively called environmental factors. They include numerous variables as weather conditions (i.e. lighting), noise conditions (i.e. "the cocktail party effect") or location conditions (i.e. mobility, in-car scenario…).

In the case of TTS and ASR engines, arguably the most harmful effect is that posed by noisy environments. Regarding virtual signers, taking into account that deaf users should be looking with attention to the virtual agent, the cognitive load that the environment demands has to be taken into consideration. In order to illustrate this, let's imagine an application designed for interacting through a tactile interface and that uses at the same time a virtual signer for communicating the information. Then, it is necessary to set the message of the virtual signer in such a way that it does not coincide with any other visual message. For SIGNSPEAK, there are three main tasks related to the multimodal visual analysis: tracking of hand positions, facial analysis and body pose estimation. All of them need robust tracking algorithms, since they should avoid the effect of i.e. signing hands moving in front of the face, or signing hands crossing the other hands.

### 3.3 Resource-related factors

Among these factors, we can include the computational power or memory availability in the devices or quality of communication requirements. Thus, these resources have influence on the selection of a concrete technology, the use of a concrete device (i.e. a desktop environment vs. a portable device) or in a worst case scenario, a degradation of the user perceived quality.

In the case of SIGNSPEAK technology, very demanding requisites regarding computational power are needed. Its flow network implies several stages with certain complexity. Due to that, there is a delay factor of around 20 times compared to real-time (for example, the translation of 6 seconds of video will take around 2 minutes) for testing data coming from the same domain as the data used to train the system.

### 3.4 User perception and acceptance of the technology

User acceptance of a new technology does not depend exclusively on its technical functionality. User perception of a new technology is built from a set of psychological, social and contextual factors that are related to its use in everyday life applications. Some of these factors have been already mentioned in section 3.1 although complete models from different perspectives and at various levels have been developed (Venkatesh et. al., 2003).

Results of an expert survey performed in SIGNSPEAK project, regarding what aspects are more important for selecting technological products are showed in Figure 3. These results are presented through a bar graph showing three colours depending on the relevance for the user (red for low relevance, orange for mid-range relevance values and green for high relevance).

After analysing the results, we can see that the price of a product is not seen as a fundamental factor and apparently when the service provided by the technology is really useful, price is not very important. Logically this fact, that is applicable to any target population, gains importance for the deaf community since technology helps them break down very annoying communication barriers. At the other end, the more relevant factors are: usefulness, easiness of use and having the ultimate technology. Related with the abovementioned great need of technology products able to help the deaf community, it seems clear that usefulness is a key variable for choosing a device. This may indicate that cutting-edge technology has been associated to a perceived loss of reliability of the technology's performance (a factor which was not represented in the survey). And finally, due to the accessibility difficulties which traditionally deaf people have to face in the use of technology products, easiness of use is also lightly highlighted.

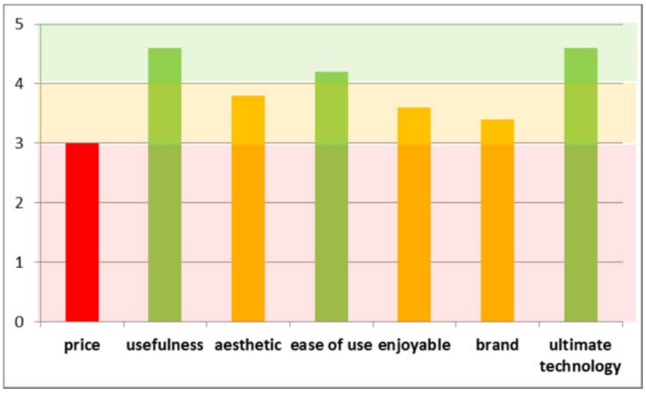



Figure 3. Bar graph about the most relevant factors in the selection of a technological product.

# 4. Application Scenarios

In order to select a relevant application scenario for applying SIGNSPEAK technology, an expert survey was performed. The creation of the questionnaire and the selection of the experts were made jointly between European Union of the Deaf (EUD) and Telefónica R&D (TID), both partners of SIGNSPEAK project.

In section 4.1 a review of the communication problems highlighted by these experts is presented. Later, in section 4.2, the experts' feelings about a set of possible application scenarios are listed.

## 4.1 Communication problems

The expert survey addresses the issue about the more unpleasant situations when deaf people have to communicate with non-signers.

Some of these situations are listed below:

- Telephoning hearing people through a relay service based on text, which is not their first language.
- Sending e-mails via text messages, instead of using their first language.
- Video films using sign language are often not subtitled and most of the hearing people cannot understand sign language.
- Accessing to public authorities/services (i.e. passport issuing service, banking, etc.) where most people cannot sign.
- Relay services almost never opens 24 hours a day.
- Hearing people cannot learn sign language without getting instructions in their own written or spoken language.

## 4.2 Scenario analysis

In the expert survey several scenarios were proposed. These scenarios, created in collaboration with EUD, take into account the communication needs of deaf community and the forecasted usefulness.

All of them have the following motivation story:

*"John and Mary are a deaf-hearing marriage and they have one child, Susan, who is 7 years old and she is also deaf. This family is bilingual; sign language and spoken language. They have hearing neighbours and family members who cannot sign very well."*

### 4.2.1 Sign language e-learning

This scenario is as follows:

"A neighbor girl of Susan is following a course for improving her sign language skills. For doing this course, pupils have to connect to the teacher through Internet (using a webcam). Then, pupils see the teacher in their monitors and the teacher can see all the pupils at their own homes. The teacher gives the lessons using sign-language and, thanks to SIGNSPEAK technology, text subtitles appear at the same time."

In this case, experts told us *"beginners learn better with signing videos without subtitles and then they can watch signing videos with the subtitles to see if they already understand sign language"*.

### 4.2.2 Answering machine

This scenario is as follows:

"John is in a congress and makes a video call to home. Nobody is at home, so he leaves a recorded video with his sign language message. The answering machine, through SIGNSPEAK technology, translates the sign language message into text. When Mary arrives home, she realizes there are several messages. As she is busy, she decides listen the messages while preparing the dinner. She listens to her husband's message through a voice synthesizer."

This service arouses a similar feeling as Sign Language e-learning, at least for those who can sign very well, since *"I would prefer to see him directly signing instead of hearing the voice synthesizer"* betting for the concept of a more realistic conversation. Therefore, for someone not able to sign well or at all, a service like this would be considered as a good idea

### 4.2.3 Play Sign Language

This scenario is as follows:

"Susan has a game console which includes a camera. She wants to play with her neighbour girl. They love to play an educative adventure game that makes you practice some sign language expressions. Using the video from the camera, SIGNSPEAK technology assesses the quality/correctness of the signs and the game gives Susan feedback about how to improve her sign language abilities. As the neighbour girl gets better, she moves forward the levels of the game. They improve their communication very well through playing the game."

Finally the game for practicing Sign Language was really welcomed since *"playing with sign language is the best way to learn it. If it is more formal as in the school, then children would get very bored."*

### 4.2.4 VideoSL mail

This scenario is as follows:

"Mary wants to send an email to several people. Some of them can hear while others cannot. She records a video signing and she sends it. SIGNSPEAK technology translates the sign language message into text and then it sends the email with the video and the text message to all the addressees."

VideoSL mail was considered as good and suitable to SIGNSPEAK technology. Additionally, it was detected as a possible application for learning: *"Hearing people would learn sign language by reading the text. Text and sign language should be next to each other in the system."*

## 5. The Prototype: VideoSL Mail

Since one of the main goals of SIGNSPEAK project was to analyse the industrial application of SIGNSPEAK technology in order to fully understand the possible implications of the integration of this technology, finally, based on the experts' opinions and the limitations of SIGNSPEAK technology (i.e. non-real-time processing), the VideoSL mail scenario was selected for the development of a prototype. This scenario was devised as employing a similar concept of use as Google Voice automatic voicemail transcription, helping a signer-hearing group of friends to socialise together without the need for interpreters.

The main advantages pursued with this prototype are:

- Text previewing of the video messages. This feature is particularly oriented to those deaf people comfortable with reading.
- Ability to search of information in video data.
- Allowing deaf people to express themselves using Sign Language.
- Making it possible that non-signers hearing people can understand a message expressed in Sign Language.

### 5.1 Architecture

Telefónica R&D has implemented a framework and user interface based upon many of the principles of cloud computing. This framework will provide a flexible communications infrastructure for developing SIGNSPEAK services. Cloud computation is defined as the provision of computing services over the Internet in a manner reminiscent of those of public commodities such as electricity or watering systems. Thus, it is a way to offload processing of data to places other than the user's system.

In Figure 4, the general architecture diagram devised for SIGNSPEAK service is shown. The devices communicate directly using a web-based interface such as a browser (in the case of a traditional PC) or using a mobile application that adapts the UIs to the particularities of a mobile device (that could be either a tablet or a smartphone). These applications communicate their translation requests (in the figure, this channel is marked with green arrows) via the web interface to the SIGNSPEAK servers. The web interface is a standard Web Service that accepts basic data such as the stream/location of the input video and some settings for the translation (e.g., addressee or timing constraints). The Job Scheduler is a module that gathers all the translation requests and generates a list of "translation jobs" to be executed. Finally, the jobs are sent to the SIGNSPEAK translation pipeline and the results are tored into the database, being available at request.

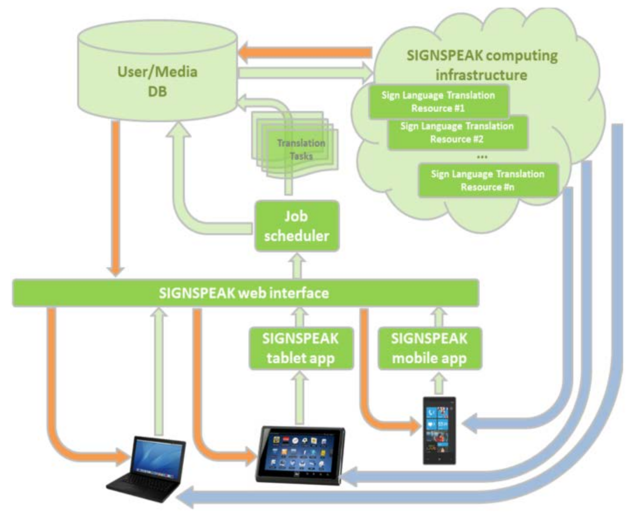



Figure 4. The SIGNSPEAK Cloud architecture

### 5.2 User interface

In order to simplify the portability of the interfaces between different devices and platforms, the user interface was implemented using HTML5 and JavaScript (making use of Sencha Touch library). Its main features are:

- Easy user interaction. Mails are presented in a vertical carousel, so users can use up/down swipe gestures to view the mails.
- Filtering capabilities. The search functionality is accessed through a text box and it makes possible the e-mail filtering based on the e-mail bodies or on the translations generated by SIGNSPEAK.
- Quick use feature. Users can select some videos as a sort of frequent replies and then attach them to their e-mails using a drag-and-drop paradigm.

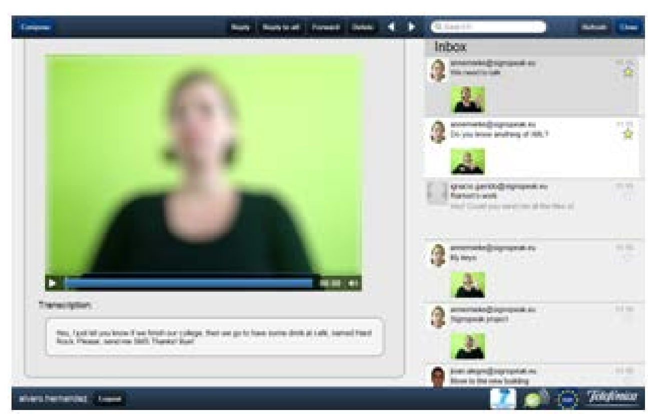



Figure 5. VideoSL mail user interface

### 5.3 User Experience evaluation

In collaboration with EUD a user evaluation has been carried out to gain insights about the suitable performing and acceptance of the VideoSL mail service.

The prototype was installed into a touch tablet device and a task-guided evaluation was carried out by 5 users. Once they had interacted with the application, they filled out a questionnaire. Some of the factors addressed by this preliminary evaluation are:

- Previous experience regarding email services and tablet PC devices.
- Likeability of the service.
- System performance.
- Usefulness of the service.
- Willingness to buy.
- Overall acceptance.

After gathering and interpreting users' feedback, the first results show a high acceptance and excitement about this system and how its daily life would be much easier thanks to the use of this technology.

## 6. Conclusions

In this paper, SIGNSPEAK project has been presented, focusing on one of its main challenges, i.e. how to improve the communication bridge between signer and hearing communities. Telefónica R&D as the main industrial partner of the project has addressed this challenge, firstly studying the main needs of potential users, and then creating an application prototype of a VideoSL email service, still without full functionality due to the limitations of the state-of-the-art technology for a real-time operation, but able to provide a similar User Experience to that than a real service would cause.

A preliminary user's feedback has been collected, showing how excited they are about this prototype, but also making us aware of the necessity of continuing the Research on this technology field.

## 7. Acknowledgements

This work received funding from the European Commission's Seventh Framework Program under grant agreement number 231424 (FP7-ICT-2007-3).

The authors express their gratitude to the European Union of the Deaf (EUD) for its cooperation in the framework of SIGNSPEAK project, providing first-hand useful information and access to real users and their needs.

## 8. References

Acero, A. (1992). Acoustical and Environmental Robustness in Automatic Speech Recognition. Kluwer Academic Publishers

Agarwal, R. and Prasad, J. (1998). A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology. Info. Sys. Research (INFORMS) 9: 204-215

Arning, K. and M. Ziefle (2010). Ask and you will receive: Training older adults to use a PDA in an active learning environment. International Journal of Mobile Human-Computer Interaction 2, no. 1: 21-47

Bailenson, J.N., and N. Yee (2005). Digital chameleons, Automatic Assimilation of Nonverbal Gestures in Immersive Virtual Environments. Psychological Science (SAGE Publications) 16, no. 10: 814

Dutoit, T (1997). An introduction to text-to-speech synthesis. Vol. 3. Springer

Hirschberg, J. (2002). Communication and prosody: Functional aspects of prosody. Speech Communication 36, no. 1-2: 31-43

Huang, C., Chen, T., Li, S., Chang, E. and Zhou, J. (2001). "Analysis of speaker variability." Citeseer. 1377-1380

Karat, J., Horn, D.B., Halverson, C.A. and Karat, C.M. (2000). Overcoming unusability: developing efficient strategies in speech recognition systems. ACM. 141-142

Market Research (2011). Opinion Leader. "OfCom Relay Services". UK competition and regulatory authority, Market Research

Reeves, B., and Nass, C. (1996). The media equation: how people treat computers, television, and new media like real people and places. Cambridge University Press New York, NY, USA

Rodríguez-Gancedo, M.C., Caminero, J., Van Kampen, A. (2011). Report about the study of the new communication bridge between signers and hearing Community. Signspeak Project, Deliverable D9.4. http://www.signspeak.eu/deliverables/SIGNSPEAK_D9.4_v1.0.pdf

Spiegel, M. F. (2003). "Proper Name Pronunciations for Speech Technology Applications." International Journal of Speech Technology (Springer Netherlands) 6: 419-427

Stein, D., Forster, J., Zelle, U., Dreuw, P. & Ney, H. (2010). RWTH-Phoenix: Analysis of the German Sign Language Corpus. 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, Malta, May

Venkatesh, V. and Morris, M. (2000). Why Don't Men Ever Stop to Ask for Directions? Gender, Social Influence, and Their Role in Technology Acceptance and Usage Behavior. MIS Quarterly (Management Information Systems Research Center, University of Minnesota) 24, no. 1: pp. 115-139

Venkatesh, V., Morris, M., Davis, G. & Davis, F. (2003). User Acceptance of Information Technology: Toward a Unified View, MIS Quarterly 27(3), pp. 425-478

Wheatley, M. and Pabsch, A. (2010). "Sign Language Legislation in the European Union". Edited by EUD. European Union of the Deaf

Ziefle, M. and Bay, S. (2008). Transgenerational designs in mobile technology. Handbook of Research on User Interface Design and Evaluation for Mobile Technology 1: 122-141

# From Corpus to Lexical Database to Online Dictionary: Issues in Annotation of the BSL Corpus and the Development of BSL SignBank

**Kearsy Cormier[1], Jordan Fenlon[2], Trevor Johnston[3], Ramas Rentelis, Adam Schembri[4], Katherine Rowley[1], Robert Adam[1] and Bencie Woll[1]**

University College London[1], Gallaudet University[2], Macquarie University[3], La Trobe University[4]
Correspondence to: Deafness, Cognition and Language Research Centre, University College London
49 Gordon Square, London, WC1H 0PD, United Kingdom
Email: k.cormier@ucl.ac.uk, jordan.fenlon@gallaudet.edu, trevor.johnston@mq.edu.au, hipas8@mac.com, a.schembri@latrobe.edu.au, kate.rowley@ucl.ac.uk, r.adam@ucl.ac.uk, b.woll@ucl.ac.uk

**Abstract**

One requirement of a sign language corpus is that it should be machine-readable, but only a systematic approach to annotation that involves lemmatisation of the sign language glosses can make this possible at the present time. Such lemmatisation involves grouping morphological and phonological variants together into a single lemma, so that all related variants of a sign can be identified and analysed as a single sign. This lemmatisation process is made more straightforward by the existence of a comprehensive lexical database, as in the case for Australian Sign Language (Auslan). When annotation of data collected as part of the British Sign Language (BSL) Corpus Project began, no such lexical database for BSL existed. Therefore, a lemmatised BSL lexical database was created concurrently during annotation of the BSL Corpus data. As part of ongoing work by the Deafness Cognition & Language Research Centre, this lexical database is being developed into an online BSL dictionary, BSL SignBank. This paper describes the adaptation of the Auslan lexical database into a BSL lexical database, and the current development of this lexical database into BSL SignBank.

**Keywords:** corpus, lexicon, dictionary, lemmatisation, British Sign Language, ID gloss

## 1. Introduction

A systematic approach to corpus annotation that involves lemmatisation of glosses is required to make a sign language corpus into a true linguistic corpus in the sense intended by McEnery and Wilson (1996) – i.e., a finite, accessible, representative set of language recordings that is machine-readable. Such lemmatisation involves not only grouping together morphological but also phonological variants into a single lemma, so that all related variants of a sign can be identified and analysed as a single sign. This lemmatisation process is made more straightforward by the existence of a comprehensive lexical database, as in the case for Australian Sign Language (Auslan) (Johnston, 2001). When lexical annotation of data collected as part of the British Sign Language (BSL) Corpus Project (Schembri, Fenlon, Rentelis, & Cormier, 2011) began in 2011, no such lexical database for BSL existed. Publicly available BSL dictionaries (e.g., Brien, 1992) focused on translation equivalents and were not lemmatised in a way which would allow *ID glossing*, i.e., type-token matching (Johnston, 2010). In order to lemmatise the data for the purposes of the BSL Lexical Frequency Study, a lexical database for BSL was created concurrently during annotation. As part of ongoing work by the Deafness Cognition and Language Research Centre (DCAL, 2011-2015), this lexical database is being developed into an online BSL dictionary, BSL SignBank. Here we describe the adaptation of the Auslan lexical database (Johnston, 2001) into a BSL lexical database, and the current development of this lexical database into an online BSL dictionary.

## 2. BSL Lexical Database (BLD)

When planning annotation of the BSL Corpus data, we began by taking advantage of the fact that a lexical database for Auslan (a sign language variety closely related to BSL which shares much of the same lexicon) already existed (Johnston, 2001). The Auslan lexical database (ALD) was initially created as an offline database, first in tabular format in Microsoft Word and then later HyperCard, then FoxPro, then FileMaker Pro. As of 2004, the Auslan lexical database additionally exists as an online dictionary as Auslan SignBank (http://www.auslan.org.au). The dictionary contains approximately 7000 entries (4000 of which are publicly viewable) and is organised in an order based on phonological parameters (Johnston, 2003). This ordering aids in identifying signs that are homonyms (or near homonyms) as signs that are formationally the same (or similar) end up as entries that are adjacent to each other, so that decisions about whether these signs are homonyms or not can be made more easily.

Because of the ease of manipulating an offline database in FileMaker Pro (e.g. adding/deleting/editing entries, searching, sorting), as opposed to a bespoke online database which requires a programmer for manipulation, we began by cloning the offline FileMaker Pro version of Johnston's Auslan lexical database in early 2011. This was the beginning of the BSL Lexical Database (BLD).

Annotators began lexical annotation of the BSL Corpus data by first searching BLD for keywords linked to the meaning of each BSL sign in the corpus video. If the sign already existed in BLD (i.e., if it was an Auslan sign that had been carried over into BLD), annotators ensured that the sign was coded as a BSL sign if it had not been already, and used that entry to annotate the sign in question (either with the Auslan ID gloss, or with a different ID gloss if needed). For BSL signs that were not in BLD already (i.e., they were not Auslan signs from ALD), annotators added entries for these BSL signs. New entries included very basic lexical attributes: ID gloss, movie clip, and keywords (English translation equivalents). The BSL Corpus team met weekly to discuss lemmatisation issues. (See §3.1 for more on lemmatisation.)

The BSL Lexical Frequency Study (LFS) (Cormier, Fenlon, Rentelis, & Schembri, 2011) was based on approximately 25,000 lemmatised sign tokens from the conversational data in the BSL Corpus, annotated using BLD. These 25,000 sign tokens represented 2506 signs, including 'partly-lexical' signs (e.g., pointing signs and classifier constructions) and 'non-lexical' signs (e.g., constructed action). (Annotations were carried out following Johnston's guidelines for the annotation of the Auslan corpus, [www.auslan.org.au/about/annotations/](www.auslan.org.au/about/annotations/).) Roughly 16,000 sign tokens from the LFS (representing roughly 1500 sign types) were lexical signs, and all of these signs are represented in BLD. Preliminary annotation of an additional 25,000 sign tokens from the conversation data and also concurrent ID glossing of sign tokens from the lexical elicitation task resulted in the inclusion of approximately 1800 sign types in BLD as of mid-2011.

## 3. From BLD to BSL SignBank

Although work on the LFS was completed with the end of the BSL Corpus Project in June 2011, further development of BLD continued, as part of DCAL's plan to create a corpus-based online dictionary and reference grammar (2011-2015).

The first step in adapting BLD into an online dictionary was to check form-meaning pairings between similar signs within the database. This initially entailed fitting the newly added signs (approximately 700 of the 1800 BSL lexical signs in BLD) into the numbering system outlined in Johnston (2003). This numbering system has signs ordered by the handshape of the dominant hand, following an order that roughly follows the order of numeral signs (and thus, the number of extended fingers) in Auslan from zero upwards. Within each handshape, one-handed signs are first, followed by signs made with two hands that have the same handshape (*double-handed signs* in Johnston's terminology), followed by signs with two hands that have different handshapes (*two-handed signs* in Johnston's terminology). Within this, signs were then ordered by primary location, from the top of the head downward. Ordering beyond these features (handedness, handshape, and location) then roughly followed a series of other phonological parameters (e.g. symmetry, orientation, location on non-dominant hand, and contact). However, as Johnston (2003:456) notes:

"The Auslan dictionaries only partially implemented the finer decision schema… because, in practice, discrimination beyond three or four levels within the decision schema has not been necessary in order to sequence most lexical signs. The reason for this is simply that the data contain few exemplars of more finely discriminated lexical signs. Indeed, even in those handshape sections that contain hundreds of distinctive signs, often no need arose to adhere to any strict sequencing beyond the major and minor features and secondary tabulation."

Attempting to add 700 BSL signs into this numbering system quickly proved to be problematic, particularly for dense phonological neighbourhoods. For example, Auslan and BSL both have many double-handed signs in neutral space with unmarked handshapes (e.g., with the 1 handshape or 5 handshape). Because there was no strict sequencing for Auslan signs via Johnston's (2003) system beyond the major parameters and a few minor parameters (because, as noted above, it was not needed for Auslan signs), it became difficult to find only one appropriate position within the numbering system where these signs belonged. After attempting to add in a few hundred BSL signs into the Johnston numbering system, we found that we ended up with several clusters of phonologically similar signs scattered throughout these dense phonological neighbourhoods, which made it increasingly difficult to find homonyms, near-homonyms, minimal pairs, and near minimal-pairs (which was meant to be one of the purposes of the numbering system in the first place – to easily identify these similar signs to check lemmatisation).

It became clear that the only way to check phonologically similar signs to ensure proper lemmatisation (e.g., that homonyms had been distinguished) was to code phonological information for each of the entries in the database first, on the assumption that these would represent tentative lemmata until proper lemmatisation could be done. There were several options for phonological coding of the lexical entries in BLD. One was to use a standard notation system like HamNoSys (the Hamburg Notation System). The Auslan lexical database contained HamNoSys transcriptions for each entry. However, HamNoSys is a phonetic transcription system, a much greater level of phonetic detail than was needed for organisation/sorting of the database. Furthermore, we needed the ability to search for/sort by various combinations of phonological parameters. HamNoSys transcriptions consist of a string of symbols, and sorting via parameters representing the symbols in the middle of the string would not have been straightforward. It is for this reason that the Auslan lexical database contains fields that redundantly encode information about the major phonological parameters for each Auslan entry (handedness, handshape and location). Thus the next step was coding for these major phonological attributes for the 1800 BSL signs from the BSL Corpus Project. Fields for other phonological parameters (e.g., movement) will be added after a first attempt at lemmatisation via searching/sorting, to see what kinds of parameters will be needed to distinguish signs at a detailed level.

Before such searching/sorting for lemmatisation purposes can take place, the database needs to contain a certain core vocabulary. If this is not the case, entries would need to be re-lemmatised after core vocabulary is added. It thus helps to try to ensure that core vocabulary is included before this process takes place. There is no easy way to systematically determine what "core" signs might be missing from BLD, which was based largely on spontaneous conversational data. However, the lexicon of BSL has been documented to a degree in previous dictionaries. The only such dictionary based on linguistic principles similar to those in the ALD is Brien (1992), which contains just under 1800 lexical entries. Thus, one way to ensure that the lexical database contained important core vocabulary was to check if signs in Brien (1992) were in BLD and if they were not, to add them to the database. Based on previous work by Johnston and Schembri (1999), we were aware that signs in Brien (1992) had not been systematically lemmatised, but the degree to which this was true quickly became apparent once we began including lexical items from the BSL/English dictionary in the BLD. Homonyms in Brien (1992) are typically combined into one entry[1], while signs that are clearly phonological variants are sometimes listed as separate variants for no apparent reason. Thus the process of including signs from Brien (1992) in the BLD required us to lemmatise and/or re-lemmatise those entries (e.g. by considering the relationship between the Brien (1992) signs and potential phonological/lexical variants that already existed in BLD).

### 3.1 Lemmatisation

Here we outline the principles and procedures that we used in lemmatising signs that were added to BLD as part of the Lexical Frequency Study under BSLCP, and subsequently in lemmatising (and re-lemmatising) signs from Brien (1992) into/with signs from BLD.

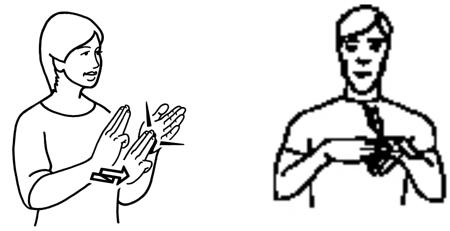
On a basic level, decisions about lemmatisation during annotation were made based on form and meaning. Two sign tokens A and B with the same form and the same meaning were considered to constitute a single lemma, with one ID gloss attributed to them. It is important to note that an ID gloss is not "the meaning" of the sign; it is simply a unique label given to a lexical item in order to aid in consistent identification of lexical items during annotation (Johnston, 2010). The meaning (via definitions) and/or English translation equivalents are stored in the lexical database. English mouthing was ignored for the purposes of lemmatisation, although of course mouthing can be used in determining some elements of meaning.

Lemmatisation involves not only grouping phonological variants but also morphological variants into a single lemma. Therefore, morphological modifications used in particular tokens such as directionality/agreement marking, number marking, aspect marking, etc were not used to distinguish lemmas.

Two sign tokens A and B with clearly different lexical meanings were considered to constitute two different lemmas, with a different ID gloss given to each one. This was the case regardless of whether the phonological forms were completely different, similar, or identical.

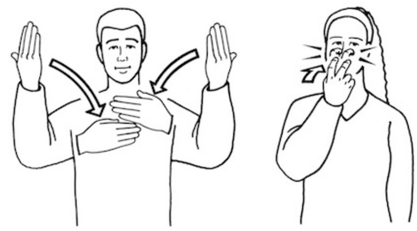
Beyond this basic level, there are various possibilities with similar/different forms and meanings. These are the primary criteria we considered:

*Phonological variants*. If sign tokens A and B differ in only one phonological parameter, and the meanings are the same or similar, then A and B are likely to be phonological variants of one lemma. For example, BSL MOTHER(M-hand) and MOTHER(B-hand), shown in Figures 1a and 1b, differ only in handshape and have the same meaning. These two phonological variants are both part of the lemma represented by the ID gloss MOTHER.



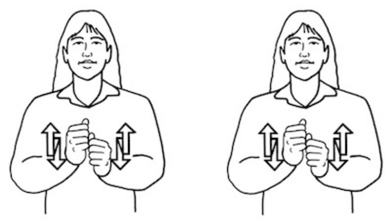
Figures 1a and 1b: Phonological variants of lexeme MOTHER: MOTHER(M-hand) and MOTHER(B-hand)

*Lexical variants*. If sign tokens A and B differ in more than one phonological parameter, and the meanings are the same or similar, then A and B may be lexical variants (separate lemmas). For example, BSL NIGHT1 is produced with two flat hands in neutral space, and NIGHT2 is produced with a bent-V handshape at the nose, as shown in Figure 2. These two lexical variants which have the same meaning (both have English translation equivalents of 'night', 'tonight', 'evening', 'dark') are distinguished in the ID gloss with numbers.



Figures 2a and 2b: Lexical variants NIGHT1 and NIGHT2

*Homonyms*. If sign tokens A and B differ in meaning but have the same phonological form, these forms are homonyms (separate lemmas). For example, both BSL BROTHER and MARCH-MONTH are produced with two A-hands in neutral space brushing against each other with alternating movement, as shown in Figure 3.

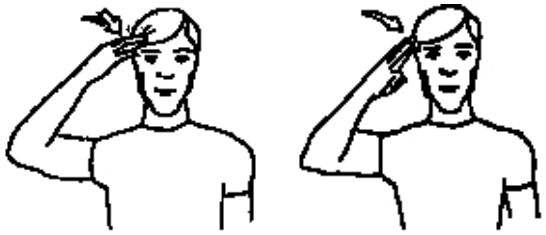


Figures 3a and 3b: Homonyms BROTHER and MARCH-MONTH

[1] The combination of homonyms into a single entry is actually not uncommon within lexicography, as distinguishing similar versus different meanings can be difficult even for spoken languages (Atkins & Rundell, 2008).

There were also additional criteria that were considered during lemmatisation beyond form and meaning:

*Association of variant with social factors*. Even if two variants A and B have the same meaning and differ only in one parameter, if one of the variants has a strong association with a particular social group (e.g. region, age, gender) or particular register (e.g. child-directed signing), this may be enough to lemmatise it separately. For instance, in addition to MOTHER as in Figure 1 above, there are other variants meaning 'mother' with similar handshapes to MOTHER (as seen in Figure 1 above) but produced at the forehead. However, these additional variants (shown below in Figure 4) were judged to constitute a separate lexeme from MOTHER since they are thought to be found in child-directed signing (i.e., they are associated with English translation equivalents 'mummy' and 'mum' in addition to 'mother'). In addition to this, the sign MOTHER is clearly a single manual letter sign (derived from two-handed fingerspelled M), whereas the relationship between MUM and the two-handed manual alphabet is less clear (we assume that the M-hand variant is a post-hoc initialisation of the original sign), as described below in §3.2.



Figures 4a and 4b: Lexeme MUM (two phonological variants, MUM(B-hand) and MUM(M-hand)

*Morphological differences in variants*. If variant A can take different morphological modifications compared to variant B (e.g. agreement/directionality, aspect marking, number marking), this may be enough to lemmatise them separately even if they are phonologically similar.

For each pair or set of sign tokens in question, all of the above criteria were considered when determining whether variants belonged to the same or different lemmas. Often these criteria compete with each other, and sometimes decisions have to be made on the basis of competing criteria that may be of equal importance. This means that it can be a considerable challenge maintaining consistency in principles of lemmatisation across all the data.

### 3.2 Citation form or headword status

Given a set of phonological variants, for the purposes of a lexical database and/or dictionary, one may want to ascribe headword (or citation form) status to one of these variants. This is not always necessary, as it is possible to have phonological variants listed in a lexical database with ID glosses that do not ascribe primary status to any single variant (e.g. with distinguishing phonological information as part of the ID gloss). However, because BLD had been created as part of a study on lexical frequency under BSLCP, it was only lexical variants that were important, not phonological variants. Thus phonological variants were not distinguished in the LFS annotations nor were they distinguished as separate entries in BLD. For each BLD entry with known phonological variants, one of those variants was chosen as the headword, or citation form – i.e. the form shown in the movie clip and the form for which phonological information is coded in BLD. Citation forms were decided based on these criteria:

*Frequency (or assumed frequency)*. Given two phonological variants A and B, the variant with the highest frequency, or assumed frequency if there is no frequency information available, or the variant that is most widely used/understood across all social groups, could be considered the citation form or headword.

*Phonological processes*. Given two phonological variants A and B, if there is a known phonological process that could explain the change from A to B, then variant A could be considered the citation form or headword. Such phonological processes include change of sign location to one closer to centre of the body (Lucas, Bayley, Rose, & Wulf, 2002; Schembri et al., 2009), change in phonological parameter from more complex/marked to less complex/marked value (Battison, 1974, 1978), or distalisation of a variant from use of joints closer to the body to use of joints further away from the body (Mirus, Rathmann, & Meier, 2001). For example, the sign TOMORROW may be produced with movement of the elbow joint, wrist joint, and/or joint at the large knuckle of the index finger. The most distalised variant uses primarily the large knuckle joint only. The citation form as shown in Figure 5 includes the use of the more proximal elbow joint.

Figure 5: Citation form for TOMORROW

*Iconicity*. Given two phonological variants A and B, if A is more iconic than B, then A could be considered the citation form or headword, on the assumption that iconic signs become more arbitrary over time (Frishberg, 1975; Klima & Bellugi, 1979).

*Nativisation processes*. If A and B are both lexical signs with some association with fingerspelling (e.g. via initialisation or fingerspelled loan), but A is closer to the fully fingerspelled word, then A could be considered the citation form or headword, following nativisation processes of fingerspelled forms (Brentari & Padden, 2001; Cormier, Schembri, & Tyrone, 2008). For example, MOTHER(M-hand) as shown in Figure 1 above is considered the citation form for the lemma MOTHER, because as noted above the M-hand variant is clearly a single manual letter sign derived from two-handed fingerspelled M .

*Prestige (or assumed prestige) status*. Given two phonological variants A and B, if variant A but not variant B is strongly associated with a social group that is known or assumed to carry prestige (e.g., region, native signer language background, etc), then variant A could be considered the citation form or headword.

*Listing in other dictionaries (e.g. Brien 1992)*. Given two phonological variants A and B, if variant A is listed in another national BSL dictionary, especially Brien (1992), then variant A could be considered the citation form.

As with lemmatisation, these criteria were considered together rather than in isolation, and each set of related variants is considered on a case-by-case basis. For example, although given the two phonological variants MUM(B) and MUM(M) shown in Figure 4 above differ in the same way that MOTHER(B) and MOTHER(M) differ (i.e., handshape), the citation form for MUM is considered to be MUM(B) rather than MUM(M), due to the (assumed) frequency of MUM(B) over MUM(M) and also the fact that the B-hand variant is less likely to have been derived directly from the fingerspelled letter M which is located on the non-dominant hand (as noted above). None of the criteria are given particular preference overall, although (assumed) prestige status and listing in other dictionaries are rarely considered unless none of the other criteria are useful in determining citation form or headword.

The challenges for determining citation form are similar to the challenges for lemmatisation as noted above. That is, criteria can compete with each other. For example, one-handed versus double-handed variants are complex. They could be explained via the phonological process of weak drop (Battison, 1974; Brentari, 1998) with the double-handed variant as citation form which can become one-handed. On the other hand, as Frishberg (1975) notes, one-handed signs can also become two-handed via a general process of signs tending towards symmetry, particularly for signs produced below the neck (outside the area of highest visual acuity), although Frishberg notes this also occurs with some signs above the neck as well. Thus phonological processes generally cannot be used to determine whether a one-handed or two-handed variant should be attributed headword/citation form status. Frequency (or assumed frequency) is often the main criterion for these decisions.

## 4. Conclusion

Here we have described the process of adapting an existing lexical database for Auslan into a lexical database for BSL for the purposes of a study on lexical frequency, and the subsequent adaptation of this BSL lexical database into an online dictionary, BSL SignBank. The primary issues involved in preparing the lexical database for launch as an online dictionary involve systematic decisions about lemmatisation (in the course of checking existing lexical entries and adding new ones from other dictionaries) and also decisions about citation form based on sets of phonological variants. We have outlined the primary criteria used in making these decisions. Such criteria are tentative and always evolving as further work on the lexical database continues. Once a core set of lexical items within BLD has been amassed and lemmatised, this will be converted into BSL SignBank online, the initial launch for which is planned for 2013. This will initially contain at least 2000 entries. Eventually we expect BSL SignBank to have a number of entries similar to Auslan SignBank (i.e., 4000). It is clear that an online dictionary allows for growth and development over time in a way that was previously not possible with print dictionaries.

## 5. Acknowledgements

Support for this research was provided by grants RES-620-28-6002, Deafness, Cognition and Language Research Centre (DCAL) and RES-062-23-0825, British Sign Language Corpus Project (BSLCP) from the Economic and Social Research Council (ESRC), UK, and grant DP1094572 (2010-2013) Australian Research Council. We thank Frances Elton, Ellen Foote, Indie Johal, and Rosemary Stamp for assisting in lemmatisation, coding, and filming of signs for the BSL Lexical Database.

## 6. References

Atkins, B. T. S., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Battison, R. (1974). Phonological deletion in American Sign Language. *Sign Language Studies, 5*, 1-19.

Battison, R. (1978). *Lexical borrowing in American Sign Language*. Silver Spring, MD: Linstock Press.

Brentari, D. (1998). *A prosodic model of sign language phonology*. Cambridge, MA: MIT Press.

Brentari, D., & Padden, C. A. (2001). Native and foreign vocabulary in American Sign Language: A lexicon with multiple origins. In D. Brentari (Ed.), *Foreign vocabulary: A cross-linguistic investigation of word formation* (pp. 87-119). Mahwah, NJ: Lawrence Erlbaum Associates.

Brien, D. (Ed.). (1992). *Dictionary of British Sign Language/English*. Boston: Faber & Faber.

Cormier, K., Fenlon, J., Rentelis, R., & Schembri, A. (2011). Lexical frequency in British Sign Language conversation: A corpus-based approach. In P. K. Austin, O. Bond, L. Marten & D. Nathan (Eds.), *Proceedings of the Conference on Language Documentation and Linguistic Theory 3*. London: School of Oriental and African Studies.

Cormier, K., Schembri, A., & Tyrone, M. E. (2008). One hand or two? Nativisation of fingerspelling in ASL and BANZSL. *Sign Language and Linguistics, 11*(1), 3-44.

Frishberg, N. (1975). Arbitrariness and iconicity: Historical change in American Sign Language. *Language, 51*, 696-719.

Johnston, T. (2001). The lexical database of Auslan (Australian Sign Language). *Sign Language & Linguistics, 4*(1/2), 145-169.

Johnston, T. (2003). Language standardization and signed language dictionaries. *Sign Language Studies, 3*(4),

431-468.
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.
Klima, E., & Bellugi, U. (1979). *The Signs of Language*. Cambridge, MA: Harvard University Press.
Lucas, C., Bayley, R., Rose, M., & Wulf, A. (2002). Location variation in American Sign Language. *Sign Language Studies, 2*(4), 407-440.
McEnery, T., & Wilson, A. (1996). *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.
Mirus, G., Rathmann, C., & Meier, R. P. (2001). Proximalization and distalization of sign movement in adult learners. In V. Dively, M. Metzger, S. Taub & A. M. Baer (Eds.), *Signed languages: Discoveries from international research* (pp. 103-119). Washington, DC: Gallaudet University Press.
Schembri, A., Fenlon, J., Rentelis, R., & Cormier, K. (2011). British Sign Language Corpus Project: A corpus of digital video data of British Sign Language 2008-2011 (First Edition). London: University College London. http://www.bslcorpusproject.org.
Schembri, A., McKee, D., McKee, R., Johnston, T., Goswell, D., & Pivac, S. (2009). Phonological variation and change in Australian and New Zealand Sign Languages: The location variable. *Language Variation and Change, 21*(2), 193-231.

# From corpus to lexicon: the creation of ID-glosses for the Corpus NGT

**Onno Crasborn, Anne de Meijer**
Centre for Language Studies, Radboud University Nijmegen
PO Box 9103, NL-6500 HD Nijmegen, The Netherlands
E-mail: o.crasborn@let.ru.nl, a.demeijer@let.ru.nl

**Abstract**

When glossing of the Corpus NGT started in 2007, there was no lexicon at our disposal to base ID-glosses on. Semantic labels were used without ensuring a constant relationship between sign form and gloss. This is currently being repaired by creating a lexicon from scratch alongside with the creation of new annotations. This substantial task is still in progress, but promises to lead to several new research avenues for the future. The current paper describes some of the choices that were made in the process, and specifies some of the glossing conventions that were used.

**Keywords:** sign language, corpus annotation, gloss, ID-glossing, lexicon, lexical database

## 1. Introduction

In the first release of the Corpus NGT in 2008, a set of 64,000 glosses for 163 sessions was included in the online Language Archive at the Max Planck Institute for Psycholinguistics.[1] Like the media files, the annotation files for most sessions have become publicly accessible. Providing an EAF file for every session, also the non-annotated ones, enables users to view the synchronised movies and any available annotations in their browser, using the ANNEX tool.[2] ANNEX allows for similar searches as ELAN, both in single files and within and across corpora.

As the glosses were created by a diverse group of mostly linguistically naïve signers that were insufficiently supervised and monitored, the resulting annotations were of variable quality. Moreover, for many aspects of the glossing, insufficient explicit guidelines were available. This paper describes the various steps that were taken to improve the glossing since then, including the present glossing conventions, working towards a second release of a larger set of annotation documents with ID-glosses later in 2012.

## 2. General issues in glossing signed interaction

Unlike the documentation of spoken languages, samples of sign language utterances are typically not glossed in the language itself, lacking a commonly used writing system or phonetic notation system. Occasionally, researchers have used HamNoSys for this purpose. More commonly, researchers create glosses in the writing system of a spoken language, whether it is the spoken language known to the deaf community in question or the language of the publication, typically being English. The choice of the spoken language word is typically not crucial, as long as it is a label that is semantically interpretable with respect to the sign by the target audience. By consequence, it may be variable across publications, and moreover it is uninformative about the form, the precise meaning, or the function of the signed word in that particular context.

While such a strategy is efficient for presenting example sign sentences in text documents, Johnston (2008) argues that it would be unwise to go about in such a loose way when annotating sign corpora. More precisely, it is key that all instances of the same sign lemma or the same full form are represented by the same word. In fact it does not matter what this word is, and it could well be a unique number. As long as a unique identifier is used, the form in context can be related to a lexicon. Johnston calls such glosses 'ID-glosses'. They primarily serve the purpose of providing a unique written identifier for every sign instance. In practice, both for annotation of new texts and for the interpretation of annotated texts, written words form the most practical solution for the identification problem, even though they may provide the false impression that the full semantics of a form in context is covered by the written word that forms the gloss.

Crasborn, Hulsbosch & Sloetjes (2012) describe a technical solution within the ELAN annotation software that in fact does use a numerical identifier for every gloss; it is this identifier that links a given annotation on a gloss tier to an external XML file that contains a list of lexical items. The surface form of this unique identifier that users see is still a text string. This in fact works not only for glosses, but for any tier in an annotation document for which a (external) controlled vocabulary can be defined.

## 3. Glossing of the Corpus NGT

### 3.1 Initial procedure

While it is clear that one cannot get around using ID-glosses in creating a machine-readable linguistic

[1] As the left and the right hand are both assigned a gloss annotation in the case of two-handed lexical items, the total estimated number of signs is 49,000, of which 15,000 are two-handed and 34,000 are one-handed.

[2] ANNEX can be opened from the corpus browser at http://corpus1.mpi.nl: in the contextual menu of an annotation document, an option appears to view the node or to perform an annotation content search.

corpus, actually using them for resources of a particular language is contingent on the existence of a lexicon that one can refer to with unique identifiers. As no lexicon was readily available to the annotators of the Corpus NGT to select glosses from and to add new glosses to, the original task for annotators was to create a translation of the form in context that appeared to be most fit to the core meaning of the sign. Thus, specific contextual meanings of the sign should not appear in the glosses. While it was recognised at the time (2007-2008) that some variation in the selection of glosses of any given form would ensue, our hope was that it would be relatively easy to take into account such variation when the corpus would be used for research later on. For example, when searching for a sign with a specific phonological form, the researcher would always be aware that different glosses for that form would have been used, and adapt his searching strategies accordingly. While there may be some value in this approach, it still requires a substantial amount of interpretation and action from researchers. We gradually acknowledged that this would never lead to a truly machine-readable corpus for the lexical level. As the signed word is such a basic unit that will be involved in nearly any linguistic or technological study, machine-readability is especially crucial at this level. We therefore decided to create a lexicon specifically for the Corpus NGT annotations.

## 3.2 A lexicon for NGT corpus annotations

The creation of the Auslan corpus (Johnston 2008b) started long after lexical resources for this language were developed by the same researcher (Johnston 1998, 2001). Thus, not only was there systematic knowledge of the Auslan lexicon, there was also a published resource from the same team that could form the basis for ID-glosses. For NGT, there is no open access reference lexicon. The existing lexical resources published by the Dutch Sign Centre are not available for research purposes, nor were they created as such. Different subsets have different origins, often created for educational purposes. The glosses that are used are targeted at easy use by laymen in a computer interface or paper dictionary, rather than at efficient computer processing. In addition, it is not unlikely that the selection of signs does not cover the lexicon that is used in the recordings of the Corpus NGT. Most crucially, this lexicon could not be expanded during the process of corpus annotations, simply because the workflow of the Dutch Sign Centre is quite different from that of the annotation of the Corpus NGT. For these various reasons, it was decided to start to compile a lexicon specifically for the Corpus NGT.

The lexicon started as a simple Excel sheet compiling ID-glosses for (regional or other) variants and a rough phonological description of each of them. This is currently being expanded to include all glosses, including semantic categories that do not have variant forms (see 3.3.1 below). To facilitate the selection of the correct gloss for a particular sign form, three fields were added. The first one contains other possible Dutch translations of the same sign form. A second column displays NGT homonyms, to point out that the same sign form has multiple glosses for distinct meanings of the sign. A third column contains related ID-glosses (by form or meaning) that may easily be confused with one another because of resemblance in form, meaning, and/or function; this information is especially useful for creating new annotations.

The added value of a corpus-based lexicon like this one is reflected by the column with Dutch translation variants. Information in this column is not just composed by making up possible Dutch translation variants of the gloss, but also contains translations actually used for that sign form, originally by annotators in the phase of intuitive glossing and currently by annotators who create annotations on the child tier 'Meaning' for a gloss. At this moment, we have not yet developed an automatic way of harvesting these meanings specified for glosses.

Currently, a phonological description has been created for every ID-gloss. The translation variants, homonym, and related glosses columns are used extensively. Further, multiple other columns for additional information are created. Whenever information is available, we specify the origin of a sign (a specific region in the Netherlands, derived from fingerspelling, a gesture, an ASL loan, etc.), the image a sign depicts can be described (COFFEE displays the image of grinding of coffee beans), mouthings or mouth gestures can be added, and observed or known phonetic-phonological variation can be specified, such as one-handed occurrences of a sign described as two-handed. As the lexicon gradually grows, we expect the use of these columns to also increase.

This Excel-based lexicon is soon to be converted to the lexical database LEXUS, probably with more structure and built-in links to related glosses. A video clip of a citation form of each entry will need to be added, as well as links to instances of the full form in context in the Corpus NGT. An area of concern is the ease of updating the lexicon once it is in LEXUS; this will no doubt be less efficient than in Excel.

## 3.3 Additional annotation conventions for glosses

In the following paragraphs, we briefly characterise the various annotation conventions pertaining to gloss annotations. They will appear in a more detailed form with further description in Crasborn & de Meijer (in prep.).

### 3.3.1 General form and labelling of variants

The general form of glosses is a single Dutch word written in capital letters. The word used for the gloss is the most neutral choice with respect to meaning and grammatical marking. To distinguish between signs with the same meaning, but different forms, alphabetical suffixes are used. For example, there are entries for HOND-A, HOND-B, and HOND-C, being three different signs that all mean 'dog'. Signs with the same form, but unrelated meanings (homonyms) each receive their own gloss.

### 3.3.2 Signs vs. gestures

The lexical or gestural status of some sign forms is not easily determined. We consider gestures to be communicative hand movements that either are also used by the community of hearing non-signing speakers of Dutch, and/or that do not have a form-meaning relationship that can be described, such as beat gestures. Emblematic gestures that can be lemmatised (i.e., that have a root form and a meaning or other communicative function that can be listed) are treated like lexical items and are marked in the lexicon as (possible) gestures. All other potential gestures are marked by a percentage character (%) on the gloss tier. In this way, (possible) gestures can easily be retrieved and inspected more closely should this be relevant to an investigation; alternatively, they can be left out in automatic processing of corpus data altogether.

### 3.3.3 Morphologically complex forms

Morphologically complex forms like classifier constructions or depicting signs cannot be annotated using an ID-gloss, due to their highly context dependent form and meaning. However, at least some of their components do have a constant form-meaning relationship that can be described. Classifiers are glossed by a three-partite combination of 1) movement, 2) type, and 3) handshape. Thus, the annotation consists of three consecutive codes, separated by an underscore '_'. MOVE_EC_1 for example is a classifier moving through space, representing an entity, that has an extended index finger as its handshape. It thus likely refers to a long and thin entity moving through space, possibly a person.

Each combination is listed in the lexicon, to facilitate data entry and avoid typos. Although some aspects of the form are described by the gloss, the meaning is left unspecified in the lexicon: there are no translation variants of the combinations. Therefore, for these glosses, the child tier dedicated to meaning always needs to be filled in, with a compact description. In the example above, this could be 'person moves forward', for instance. Signed constructions whose handshape and movement show the specific shape of a referent ('size and shape specifiers') are glossed in a similar manner.

### 3.3.4 The Modification tiers

Further modifications of the movement or other components of the constructions that we just discussed can equally be characterised on a child tier. For every gloss tier (one per hand), there is a 'Modification' tier that allows for a textual description of the modification. These tiers can be used for all types of signs, not just the morphologically complex ones. If the example form in section 3.3.3 would be modified by an arced movement expressing 'jumping forward', for instance, this would be encoded here, rather than by altering the MOVE component. The latter serves to distinguish movement through space from being at a specific location (AT), mere presence (BE), and action without a path movement (ACTION).

At present, we do not yet have specific annotation guidelines for the Modification tiers. We recognise that this would be beneficial at some point, distinguishing systematic recurrent modifications with a clear describable form from more idiosyncratic pantomimic modifications of (parts of) signs. Our strategy is to first let people intuitively use the tiers, and then after some time investigate what type of distinctions are created.

### 3.3.5 Some further conventions

Just as it is most practical to use words instead of unique numbers as glosses, for some categories of signs it is practical to use additional conventions regarding glossing. Although every unique form receives its own gloss, the conventions group together certain lexical or morphological categories to facilitate retrieval.

Examples of such further conventions:

- Hyphens (-) are used to separate multiple words representing a single gloss, whereas underscores (_) are used in glossing morphological complex forms (see section 3.3.3 above).
- Pointing signs start with the basic gloss PT; several types of pointing signs are specified in the lexicon.
- Compounds of two or more sequential parts are glossed by separate annotations for each part, and are marked on the meaning tier using '^'.
- Lexical negation is marked using the suffix -NIET 'not', so that the regular and negated forms are next to each other in various alphabetically sorted lists, such as in search results or in sorted presentations of the lexicon.
- Numbers are always glossed using digits.
- Name signs are preceded by an asterisk (*).
- Fingerspelling is marked by a hedge mark (#).
- Uncertainty on the part of the annotator is marked by a question mark (?) following the gloss; such glosses do not receive an ECV link (see Crasborn, Hulsbosch & Sloetjes, this volume).
- Double question marks (??) are used for unknown signs. These annotations should periodically be inspected by native signers other than the annotators in order to determine their nature.

### 3.3.6 A comparison with the Auslan annotation guidelines

As made explicit by Schembri & Crasborn (2010), it is desirable to work towards some kind of standardisation of annotation conventions for sign language corpora, in order to facilitate cross-linguistic research and to promote the use of published resources by other research groups. We have attempted to copy many of the published conventions for the Auslan corpus (Johnston, 2011).

Comparing the two sets of guidelines, the major correspondence is in the annotation of the basic gloss, based on ID-glosses linked to a lexicon. Other conventions, like listed in 3.3.4, show some minor differences relating to how the information is encoded. Whereas the annotations in the Corpus NGT mainly group together certain categories by using a single generic character, in the Auslan Corpus this information is mostly coded by additional information separated from the gloss by a colon or put between brackets. However, overall, the same type of information is annotated. We choose to relegate additional information about a sign to dependent tiers as much as possible (including Meaning, Modification, Handshape, and Location), so that there can be a link to the External Controlled Vocabulary for each form. We hope that separating the ID-gloss from additional information will facilitate automated processing of annotations, whether within ELAN or by creating scripts that work directly with the EAF files.

### 3.4 Discussion: conflicting principles

Ease of processing is thus an important consideration in these glossing conventions, complementing linguistic principles and sometimes conflicting with them. This is true for the very essence of the ID-gloss, a (combination of) spoken language word(s) that may not be semantically identical to the sign, but it also holds for the glossing of compounds, for instance. Our current lexicon assumes that every entry consists of a single sign syllable. For every syllable, all phonological features can be described in a uniform way, avoiding the complexity of multiple syllables that have different hand configuration or location properties, for instance. As NGT has very few compounds or other signs consisting of sequences different syllables (van der Kooij & Crasborn, 2008), there are not many signs for which the workaround for annotating compounds is problematic. But for all those that do exist, the properties of the component parts can be more easily processed. When calculating frequencies of handshapes, for instance, the handshapes of single glosses are now automatically taken into account, regardless of whether the handshape is part of a compound or not.

## 4. Applying ID-glosses to an already annotated corpus

In several rounds of revision, we are currently building the lexicon list that corresponds to the signs used in the Corpus NGT, agreeing upon the ID-glosses at the same time. We have gone through many stages in this tedious process, from spell-checking to the creation of specific conventions for name signs, for instance. The current situation of early 2012 is that about 80% of the more than 120,000 gloss annotations has a reference to the lexicon. These include the most frequent signs in the sessions that are glossed until now, as it is these that we started to assign ID-glosses to first. The remaining 20% consists of various categories, representing both known and unknown variation and known and unknown errors. Glosses referring to complex signs (depicting signs, modified lexical signs, pantomime) number about 10,000; their annotation as described in section 3.3.3 will still take quite some time. A much smaller set consists of signs that were unknown to the annotators at the time of first annotation, and are marked by double question marks; these will need to be inspected by one or more native signers. The largest proportion however, an estimated 20,000 glosses, are expected to consist of signs that are simply used infrequently. There may still occasionally be singleton glosses that refer to existing items in the lexicon, but we expect that most of them will be infrequent signs that have yet to be added to the lexicon. The lexicon currently counts 1,800 items; we expect it to grow to 4,500 by the time that these infrequent signs have all been inspected.

The lowest level of corrections, repairing typos and spelling mistakes, is slowly becoming less necessary now that a controlled vocabulary is used for the gloss tiers, making manually typed input less and less necessary (see Crasborn, Hulsbosch & Sloetjes, this volume).

A challenge that does persist and that is beyond the current round of establishing ID-glosses, is deciding whether all the variants and homonyms that have been created in the lexicon are actually independent lemmata or not. As we indicated above, some decisions on when to create a new lexical item were made to facilitate automated processing, but in other cases there was simply a lack of knowledge about (the uses or meanings of) a lexical item. It is here that the most difficult task lies as soon as a certain level of consistency is guaranteed, and it is a challenge for present users of the corpus to take the nature of the existing lexicon into account. We see this as a consequence of developing a lexicon and a corpus in tandem, and improving the nature of the lexicon will remain a rocky road for some years to come.

## 5. Expected developments

### 5.1 A second public release of the Corpus NGT annotation files

A second release of the Corpus NGT annotations will be made public in the Language Archive as soon as the unknown territory of 20,000 glosses has been inspected, hopefully before the end of 2012. The aim for this second release is not to definitively establish ID-glosses for all signs, but rather to make explicit which glosses still need closer inspection and should thus be treated with caution. This will include signs that have not been identified, but also complex constructions that have yet to be described in terms of component parts in the way outlined in section 3.3.3.

Upon that second release, the accompanying lexicon will be published in the online tool LEXUS as well as in the form of an external controlled vocabulary on a web server, linked to the gloss tiers. Moreover, the broader annotation conventions, including those for glosses summarised in this paper will be published in the form of a larger document (Crasborn & de Meijer, in prep.).

### 5.2 Development of lexicons

The RU lexicon will also be further enlarged during the annotation of other resources than the Corpus NGT, including an on-going data collection of longitudinal recordings of deaf parents with their children.

In the context of future research projects, we further hope to explore the option of the integration of lexicons of different signed languages within LEXUS. Moreover, we hope to create an English version of all the ID-glosses, and explore ways of switching between languages in ELAN for annotations like ID-glosses that should ideally be multilingual, much like ISOcat data categories may have multiple language sections. Generating ISOcat data categories for lexical items might in fact be a strategy to address this wish, and it may also facilitate multilingual lexica in the sense of a 'universal sign dictionary' ('universal SignBank', Trevor Johnston, pers. comm.): there could be a data category for a specific form that can have different meanings or functions in different languages.

## 6. Acknowledgements

The work described in this paper has been supported by ERC Starting Grant no. 210373 and NWO VIDI grant no. 276-70-012 awarded to Onno Crasborn, by EU grant 231424 'SignSpeak', and by the Centre for Language Studies of Radboud University Nijmegen. We thank Micha Hulsbosch for his help in automating tasks related to revising glosses. In alphabetical order, we thank our colleagues Richard Bank, Lianne van Dijken, Els van der Kooij, Yassine Nauta, Ellen Ormel, Johan Ros, Anna Sáfár, Merel van Zuilen, and Inge Zwitserlood for their contributions to the discussions on glossing conventions and the workflows that we followed. Moreover, we thank Karlijn Hermans for her substantial contribution to the correction of errors and the tagging of variants in the corpus.

## 7. References

Crasborn, O., Hulsbosch, M., & Sloetjes, H. (2012). Linking Corpus NGT annotations to a lexical database using open source tools ELAN and LEXUS. Paper presented at the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon.

Crasborn, O. & de Meijer, A. (in prep.) Annotation conventions for the Corpus NGT. Ms., Radboud University Nijmegen.

Johnston, T. (1998). Signs of Australia. A new dictionary of Auslan (Australian Sign Language) (the sign language of the Australian deaf community). North Rocks, NSW, Australia: North Rocks Press, Royal Institute for Deaf and Blind Children.

Johnston, T. (2001). The lexical database of Auslan (Australian sign language). Sign Language & Linguistics, 4(1/2), 145-170.

Johnston, T. (2008a). Corpus linguistics and signed languages: no lemmata, no corpus. In O. Crasborn, E. Efthimiou, T. Hanke, E. Thoutenhoofd, D. & I. Zwitserlood (Eds.), 5th Workshop on the Representation and Processing of Signed Languages: Construction and Exploitation of Sign Language Corpora (pp. 82-87). Paris: ELRA.

Johnston, T. 2008b. Corpus of grammar and discourse strategies of deaf native users of Auslan (Australian Sign Language), Endangered Languages Archive, SOAS, University of London. http://elar.soas.ac.uk/deposit/johnston2012auslan

Johnston, T. (2011). Auslan Corpus Annotation Guidelines. Unpublished manuscript, Macquairie University, Sydney. http://www.auslan.org.au/video/upload/attachments/AuslanCorpusAnnotationGuidelines30November2011.pdf

Kooij, E. van der, & Crasborn, O. (2008). Syllables and the word prosodic system in Sign Language of the Netherlands. Lingua, 118, 1307-1327.

# Linking Corpus NGT annotations to a lexical database using open source tools ELAN and LEXUS

**Onno Crasborn[1], Micha Hulsbosch[1,2], Han Sloetjes[2]**
[1] Centre for Language Studies, Radboud University Nijmegen
[2] Max Planck Institute for Psycholinguistics
[1] PO Box 9103, NL-6500 HD Nijmegen, The Netherlands
[2] PO Box 310, NL-6500 AH Nijmegen, The Netherlands
E-mail: o.crasborn@let.ru.nl, m.hulsbosch@let.ru.nl, han.sloetjes@mpi.nl

**Abstract**

This paper describes how we have made a first start with expanding the functionality of the ELAN annotation tool to create a bridge to a lexical database. A first lookup facility of an annotation in a LEXUS database is created, which generates a user-configurable selection of fields from that database, to be displayed in ELAN. In addition, an extension of the (open) controlled vocabularies that can be specified for tiers allows for the creation of very large vocabularies, such as lexical items in a language. Such an 'external controlled vocabulary' is an XML file that can be published on any web server and thus will be accessible to any interested party. Future development should allow for the vocabulary to be directly linked to a LEXUS database and thus also allow for access right management.

**Keywords:** sign language, annotation, Corpus NGT, lexicon, ELAN, LEXUS, controlled vocabulary

## 1. Introduction

Since the public release of the media files of the Corpus NGT[1] in December 2008, a subset of the data have been provided with gloss and translation annotations in the context of various research projects. During that process, it gradually became clear that glossing is not possible without a lexicon of ID-glosses (see Johnston 2008 for discussion). The creation of that lexicon is described by Crasborn & de Meijer (this volume). The present paper focuses on the facilitation of the process using a combination of the CLARIN standard tools ELAN[2] and LEXUS[3] created by the MPI for Psycholinguistics, a spreadsheet programme, and custom-made Perl scripts.

## 2. Displaying information from LEXUS in ELAN

As sign language glosses are always in the form of words from a spoken language, it is important that these glosses are consistently linked to lemmata or full forms in a lexicon. As the open source multimodal annotation tool ELAN does not have a lexicon function built in (as opposed to iLex[4], for instance), an effort was undertaken to create a bridge to the open source lexicon tool LEXUS. The first steps for this were taken in the CLARIN-NL project SignLinC (Crasborn, Hulsbosch, Sloetjes, Schermer & Harmsen, 2010), and this functionality has since been expanded.

For SignLinC, a lexicon tab in ELAN was created, in which properties of lexical items can be displayed after they have been entered in LEXUS. Figure 1 presents an example of two lexical items in the lexicon tab that both contain the contents of a selected gloss. They have been generated by a lookup in the SignPhon lexicon (which is used here just for demonstration purposes). The resulting hits are displayed with their hierarchical structure, so that the desired information can be quickly selected from a large list of properties. Figures 2-4 illustrate the configuration of this service in ELAN. The actual link to a LEXUS lexicon requires logging in to LEXUS, so that a list of accessible lexica is presented to the user.

In practice, this link works as long as the gloss of a sign is identical to the top-level field in LEXUS: an online lookup is done on the basis of the text string that is in the ELAN annotation. However, there is no such live link between LEXUS and the creation of new annotations. This would not be trivial to develop, in part because ELAN is a stand-alone programme while LEXUS currently is a web-based tool. To avoid the associated complexities, we have created an alternative solution, that ideally will be replaced by a further developed bridge between ELAN and LEXUS. It is MPI's intention to create a stand-alone version of LEXUS. This would facilitate the development of further interaction between ELAN and LEXUS.

## 3. Defining an external controlled vocabulary

Instead of a direct connection to LEXUS for the creation of new lexical annotations, an 'external controlled vocabulary' (ECV) can now be defined. The ECV file itself is a fairly simple XML file that needs to be published on a web server. It is added to a file in the same way as other controlled vocabularies (Figure 5), and the same options can be applied, including assigning a specific colour to a specific item. Like a regular CV, an ECV can be linked to a 'linguistic type' specification for a tier. Unlike other controlled vocabularies, only the values that are actually used in a file are stored in the EAF annotation document. Also, each item in an ECV is identified by an XML ID which is stored as an attribute of annotations referring to an ECV item. The value of an item is stored in the EAF to have it immediately available for visualisation and for searching purposes,

[1] http://www.ru.nl/corpusngtuk
[2] http://www.lat-mpi.eu/tools/elan
[3] http://www.lat-mpi.eu/tools/lexus
[4] http://www.sign-lang.uni-hamburg.de/ilex

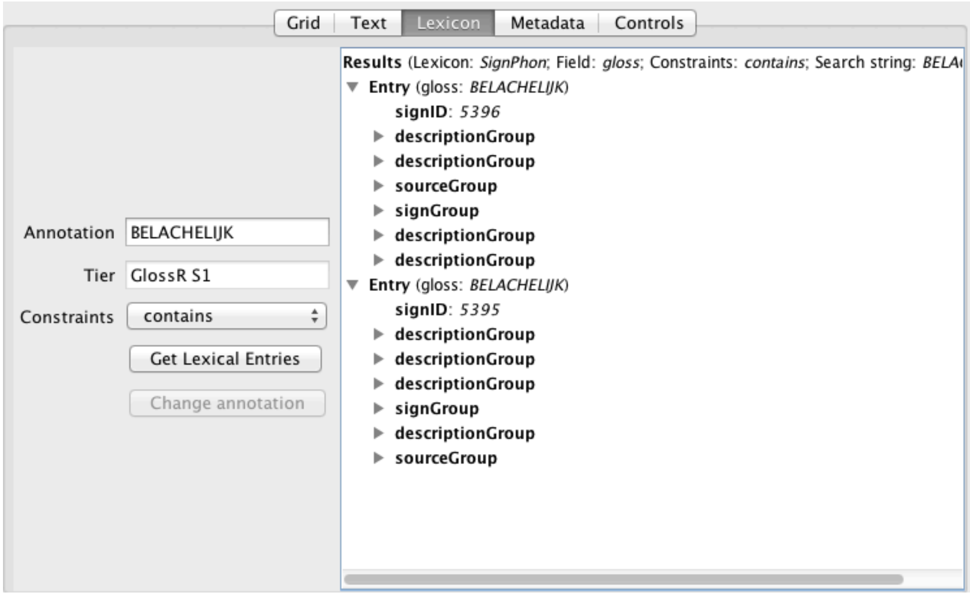



Figure 1: A search for the gloss 'BELACHELIJK' yields two hits in the LEXUS database that is specified for the linguistic type of the gloss tier

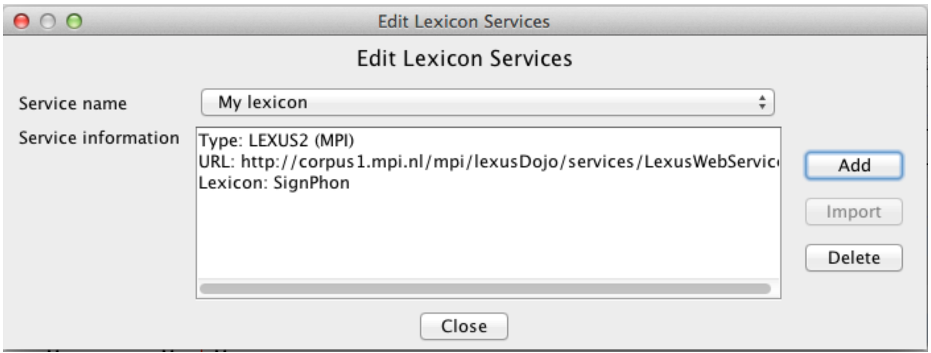



Figure 2: Adding a new lexicon service

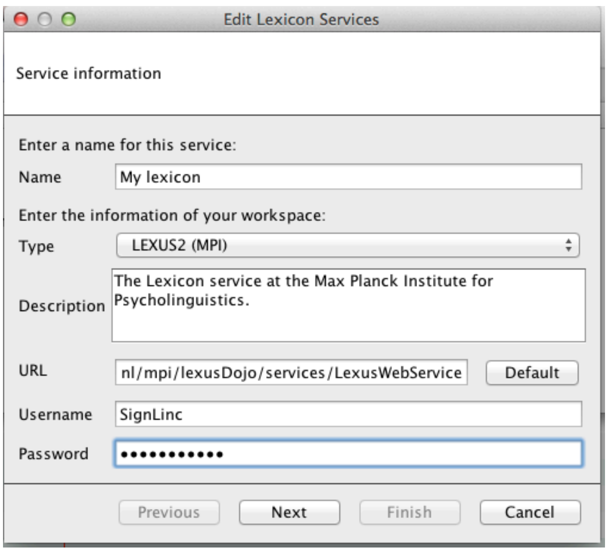



Figure 3: Logging in to LEXUS to create a new lexicon service in ELAN

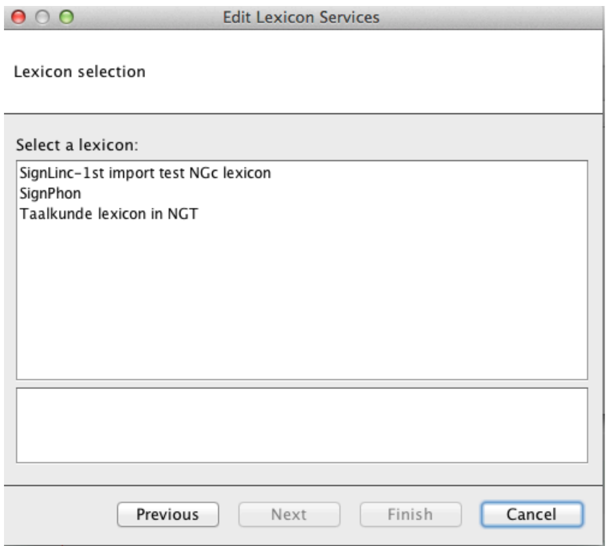



Figure 4: Selecting a lexicon

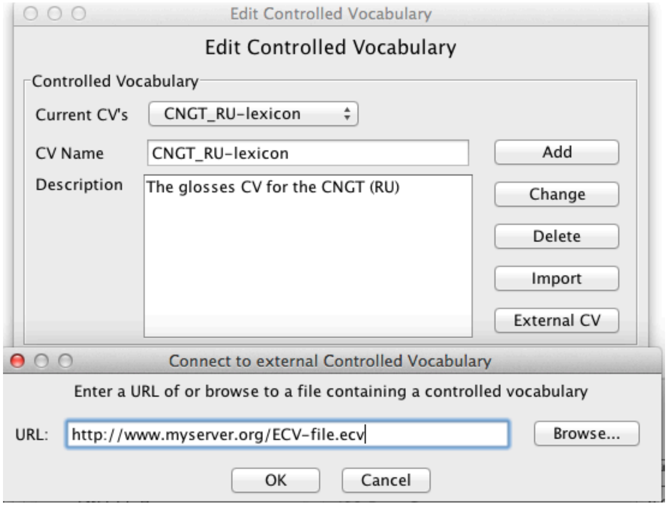



Figure 5: Specifying the URL for an ECV

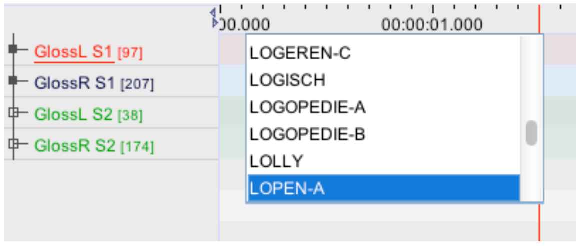



Figure 6: The suggest panel for an ECV

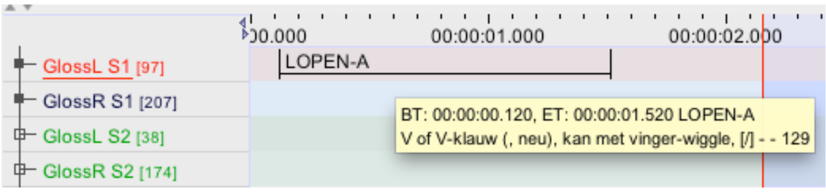



Figure 7: Inspecting the description properties through the tooltip

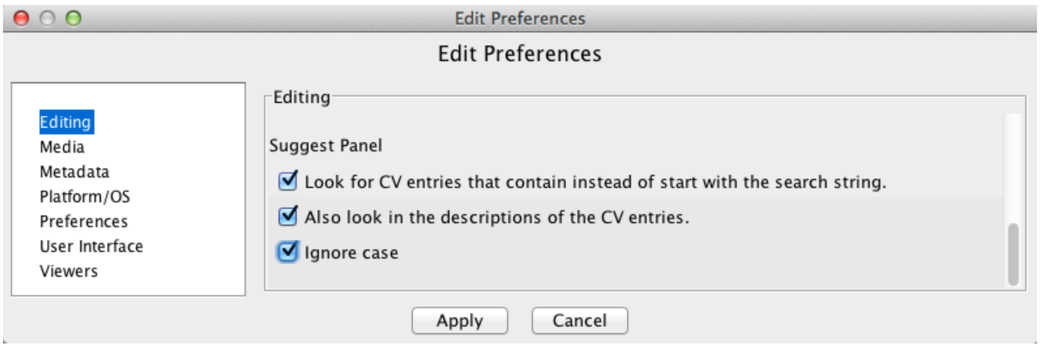



Figure 8: Specifying preferences for the behaviour of the suggest panel

but the XML ID is used to validate and update the values in the EAF in order to keep them in sync with the vocabulary (and eventually with the lexicon).

When a user creates an annotation on a tier with a CV or ECV link, a 'suggest panel' appears, offering the items from the vocabulary as suggestions (Figure 6). These suggestions can be overridden by the user if necessary.

The key advantage of an external CV is that its contents can be centrally maintained for a large set of annotation documents or even multiple corpora, so that there is no risk that the CV list starts to diverge in different documents.

## 4. Expanding the functionality of the ECV

Subsequent development of the ECV functionality in ELAN has led to two important improvements for the user.

### 4.1 Displaying additional information for annotations with an ECV link

The description field in the ECV is now visible in the tooltip that is displayed in ELAN when the mouse hovers over an annotation that has a link to an ECV entry (Figure 7). For the Corpus NGT, this description is filled with information from the lexicon, so that users have access to the phonological form of the lexical item as well as lexico-semantic properties. It has thus become easier for the user to inspect whether the selected ID-gloss actually refers to the sign in question. Current development is targeted at presenting this description property also in the drop-down list that users get to see upon creating a new annotation (the 'suggest panel'). Only in that way mistakes in the selection of the right gloss can be prevented, presupposing that the annotator can read the text string that represents phonological description of a sign.

### 4.2 Facilitating the selection of items from a large ECV

Secondly, users can specify in the preferences what is displayed in the suggest panel: the characters that are typed in to search elements in the ECV to be displayed in the list can be specified to match the start of the item (default), any text in the item, and/or also information in the description field (Figure 8).

It is especially the latter function that represents an important step forward in ensuring that users create correct ID-gloss annotations. As signs are recognised more easily on the basis of meaning rather than form, there is a natural tendency to want to translate the sign in order to create a gloss, rather than to select the ID-gloss from a list. As there are typically different Dutch translations of an NGT sign, this can lead to different glosses for the same sign. By storing potential translation variants of signs in the description field of ECV items, typing in a string like 'area' will also return a gloss like SPACE at the top of the suggest panel, alerting the annotator to the fact that AREA is not the ID-gloss that is listed in the lexicon.

### 4.3 Impact for the workflow of the annotation of the Corpus NGT

These initial lexicon-like facilities in ELAN have led to a workflow in the annotation of the Corpus NGT where both the glossed part of the corpus and the related lexicon grow at the same time. As soon as a significant number of new ID glosses are added to our Excel table and described in terms of phonological categories, translation variants, and homonyms, the ECV list is updated using a Perl script that runs on the text export of that table, and another Perl script double-checks that all instances of glosses that appear in the new ECV are assigned an ECV reference, and will thus display the description field in the timeline viewer in ELAN. From that point onwards, changes in either the gloss string or in the description field can be made in either the Excel table, and with the first following update will be visible in all instances of that gloss in any annotation document.

## 5. Conclusion & future developments

The features described in this paper have created a workflow in which ID-glosses can be created on the various gloss tiers in a more reliable way. The mere fact that the list of currently agreed-upon glosses is available upon the creation of a new gloss annotation reminds annotators of the conventions that apply and of the fact that multiple glosses may apply to the sign form at hand. At the same time, the suggest panel still remains a list of words (the essence of an ID-gloss), and does not yet provide phonological or semantic information that can help the annotator in selecting the right gloss. Presenting the information from the description field as part of the suggest panel would form a next major step in improving gloss consistency and reliability.

As far as the Corpus NGT data are concerned, the next step in this development will be that the elementary lexicon on which the ECV is based is converted to a LEXUS database. Only after such a conversion will users be able to access all the information from the lexicon in the lexicon tab. Here, data are presented in a more structured view than in the gloss tooltip in the timeline viewer.

More substantial development of LEXUS and ELAN will be necessary to facilitate the updating of the ECV based on information in LEXUS, or alternatively to merge the functionality of the ECV and LEXUS by letting ELAN generate the items in the suggest panel for a new annotation directly from a LEXUS database. Adding of lexical items to a database and modifying existing items should in the end be an integrated part of the corpus annotation process, creating a coherent set of resources for a language.

## 6. Acknowledgements

The developments described in this paper were made possible by CLARIN-NL grant 09-003 'SignLinC' and ERC starting grant 210373 'On the Other Hand' awarded to Onno Crasborn.

## 7. References

Crasborn, O., Hulsbosch, M., Sloetjes, H., Schermer, T. & Harmsen, H. (2010) Linking ELAN and LEXUS. Poster presented at the *Fourth Workshop of the Sign Linguistics Corpora Network*, Berlin. http://www.ru.nl/slcn

Crasborn & de Meijer (this volume). From corpus to lexicon: the creation of ID-glosses for the Corpus NGT.

Johnston, T. (2008). Corpus linguistics and signed languages: no lemmata, no corpus. In O. Crasborn, E. Efhimiou, T. Hanke, E. Thoutenhoofd, & I. Zwitserlood (Eds.), *5th Workshop on the Representation and Processing of Signed Languages: Construction and Exploitation of Sign Language Corpora* (pp. 82-87). Paris: ELRA.

# A GSL Continuous Phrase Corpus: Design and Acquisition

**Athanasia–Lida Dimou*, Vassilis Pitsikalis**, Theodor Goulas*, Stavros Theodorakis**, Panagiotis Karioris*, Michalis Pissaris*, Stavroula–Evita Fotinea*, Eleni Efthimiou*, Petros Maragos****

*Institute for Language and Speech Processing – R.C. "Athena",
Artemidos 6 & Epidavrou, GR 151 25 Maroussi, Greece
**NTUA, School of E.C.E., National Technical University of Athens, Greece
Heroon Polytechneiou 9, GR-15780 Zografou Campus
E-mail: ndimou@ilsp.gr, pitsik@cs.ntua.gr, tgoulas@ilsp.gr, sth@atrion.gr, pkarior@ilsp.athena-innovation.gr, mpissaris@gmail.com, evita@ilsp.gr, eleni_e@ilsp.gr, petros.maragos@cs.ntua.gr

**Abstract**

In this study, we present a newly built Greek Sign Language (GSL) corpus. The procedures followed during its implementation, consists of the linguistic design and validation, the studio and hardware acquisition configuration, the implementation and supervision of the acquisition itself, and the post-processing of the annotations for the release of accompanying linguistic/annotation resources. The reported GSL phrase corpus forms the basis for machine learning and training to serve continuous sign language processing and recognition.

**Keywords:** GSL, Kinect Camera, SL resources, sign recognition, continuous sign recognition

## 1. Introduction

The corpus presented in this article is composed of a limited number of Greek Sign Language (GSL) sentences and was created in order to provide additional data to the already obtained corpus during the first year of the Dicta-Sign project (Matthes et al., 2010). More specifically this corpus intended to serve as the ground upon which a significant part of the recognition process would be tested and evaluated, more precisely, the continuous sign language recognition algorithms developed in the project.

Given the targeted nature of this corpus we next present step by step the constraints as well as the procedure followed in order to obtain it.

## 2. Methodology

Methodologically, the creation of the GSL phrase corpus made re-use of already existing data that were acquired within the Dicta-Sign project as part of the project's "parallel" corpus and lexicon resources. However, the initially acquired data were characterised by a number of restrictions relating to their specific discourse content and, most significantly, not incorporating the whole range of parameters required for running the continuous sign language recognition experiments.

Limitations noticed in respect to the content of the semi-spontaneous Dicta-Sign corpus, were noticed in relation to the variety of sign formation parameters, significant for recognition processing, such as location and handshape. Such parameters had limited a lot the volume of useful segment of the initially acquired corpus. The GSL phrase corpus presented in the rest of this paper, contains GSL phrases of simple to modest complexity.

To meet sign language recognition experimentation requirements, the phrases that formed the corpus were selected to be significant for Sign Language linguistic analysis and also for their employment in automatic recognition tasks, given that in sign language recognition terms, simple phrases constitute a task of intermediate complexity when compared to a more open and unconstrained continuous corpus[1].

### 2.1 The GSL Dicta-Sign Corpus

The GSL phrase corpus is directly related to parts of the previously acquired Dicta-Sign parallel corpora (Matthes et al., 2010).

The existing Dicta-Sign GSL corpus has been exploited in order to extract from it small in length and simple in structure phrases that would become the stimuli for the GSL phrase corpus.

The Dicta-Sign corpus was built upon using real life discourse situations between native GSL signers e.g. being at the airport, travelling etc. Those real life situations were divided into tasks that the signers had to reproduce in front of the camera after being instructed thoroughly on how to do so. The complete GSL Dicta Sign corpus was created with the use of nine different elicitation tasks that were performed by eight pairs of native adult Deaf signers.

#### 2.1.1 Task 4 of the GSL Dicta-Sign Corpus

The GSL phrase corpus covers lexically one of the nine elicitation tasks of the Dicta-Sign parallel corpus, namely the fourth one. This task (Dicta-Sign Task 4) treats a specific topic; the situation in the Airport entailing communications on issues of check-in, luggage depository, boarding, safety instructions, meals, take off and landing. There are many reasons for choosing this

[1] A first exploitation of the newly acquired Continuous Recognition Training Phrase Corpus will be seen via the Dicta-Sign project Demonstrator (currently available at: http://signwiki.cmp.uea.ac.uk/dictasign).

specific task, the most important of which is that in all four project sign languages, this was the task that presented the strongest similarities in terms of common topics addressed and resulted in the most comparable parts of the Dicta-Sign corpus among the involved four languages. For this reason this task was also given priority in the manual annotation procedure, the latter being a time-consuming and labour-intensive process.

#### 2.1.2 Annotation of the GSL Dicta-Sign Corpus

Availability of annotations for the complete Task 4 segment was another reason to select this specific task. The annotation in the GSL segment of the Dicta-Sign corpus included the following four tiers:

- GSL Lemmas (creation of the lexicon entries in Greek that would most suitably fit the GSL lemmas),
- Clause boundaries (the continuous signing was cut into clauses),
- English translation of the clauses,
- HamNoSys transcription of the GSL lemmas.

### 2.2 The "list of 1000 Common Concepts"

In order to obtain the lexical entries that the phrases of the GSL phrase corpus were going to be composed of, we compared the lexical entries found in Task 4 of the Dicta-Sign corpus to a list of concepts assigned lexical entries in all four project languages, that served during the first two years of the Dicta-Sign project as the basic common lexicon among the project sign languages.

Throughout the course of the project these lexical entries, consisted a means of exchanging data upon a common ground, which henceforth will be referred to as "The list of 1000 Common concepts".

These "1000 common concepts" apart from being a reference point and a visible outcome of the project[2], became the object of another recording that took place in year 2 of the project for GSL and German Sign Language (DGS). This recording aimed at obtaining data on the handshape, and the movement (trajectory, orientation etc) that is effectuated during the performance of each one of these signs. For this reason extensive footage sessions took place that tracked with a 3D camera the trajectories the signers performed for each one of the 1000 corresponding signs in each language. In notation terms, the whole set of these lexical entries are represented via HamNoSys notations.

The representation of signing of the concepts was transformed computationally, in a way to relate the signing with Postures, Detentions, Transitions and Steady Shifts (PDTS). PDTS is a sequential model proposed by Johnson & Liddell (2011) to capture and label the sequential structure in sign language at the level of linguistic phonetic units.

Every sign is further represented with HamNoSys notation. This is a phonetic transcription for signing language lemmas (Hanke, 2004). The first step in the adopted procedure was to transform the HamNoSys representation into structured sequences of labels in the form of Gestural SiGML (Signing Gesture Markup Language). This is a form of representing gestures in a structured sequence, better understandable by human readers than the HamNoSys form. The next step was to convert the results into segmented SiGML and finally summarise the segmentation in a set of PDTS labels. These sets of labels were used as a basis for the training of the algorithm treating sign location and movement. Combining this information with skeleton tracking information it was possible to align sequences of structured labels with visual data segments (Pitsikalis et al., 2011)

The above mentioned procedure as well as the measurements that took place were essential and allowed to support the recognition processing the linguistic treatment, as well as the synthesis procedures maintaining the ability to represent lemmas by means of HamNoSys notations.

### 2.3 Comparison of the two lists

The comparison of the two lists of sign lemmas, deriving from the Dicta-Sign resources, resulted to an overlapping between the two lists that consisted of 113 GSL lemmas. For each of these lemmas the following information is available: a. Gloss (written in Greek), b. HamNoSys transcription, c. English translation, d. kinect based skeleton tracking information. These 113 GSL lemmas from the Dicta-Sign corpus became the repository upon which the GSL phrase corpus was based.

## 3. The GSL phrase corpus

The 113 GSL lemmas that came out of the comparison of the two different lists, were initially examined in terms of the linguistic resources to be considered, namely: the vocabulary units in relation to the various lexical types and the frequency of occurrence of the considered lemmas. The methodology adopted in order to obtain the GSL phase corpus can be divided into two parts, which are presented in subsections 3.1 and 3.2.

### 3.1 Original Phrases from the Corpus

The 113 GSL lemmas were located within the transcripts of the recordings of Task 4, as they were originally uttered by the GSL signers of the Dicta-Sign Corpus.

Given that Task 4 was one of the Tasks that were fully annotated very early in the timeline of the project, ,we were able to locate the 113 GSL lemmas within the sentence-level annotations.

The number of phrases that contained one or more of the 113 lemmas within the eight[3] transcripts of Task 4 was

[2] http://www.sign-lang.uni-hamburg.de/dicta-sign/consign/demo/cs/cs_51.html

[3] Task 4 was a task that was performed only from one of the two signers who participated in each session of recordings for the Dicta-Sign corpus. So in this case the 113 lemmas were crosschecked across the eight transcripts, one signer per pair.

more than 300; nonetheless, our goal was to obtain those phrases that contained the maximum number of items from the 113 lemmas set in each phrase. With this criterion a selection was made leaving out those phrases that contained only 1 or only 2 lemmas from the 113.
After checking vocabulary coverage, a qualitative criterion was applied: many of the retrieved phrases were significantly long and could not be performed by the signer.. By excluding the complex and long phrases we reached the basic set of 137 simple ones that constituted the pilot part of our corpus.

#### 3.1.1 Creating the videos for the elicitation material

The original videos of the phrases have been cut into new separate video files according to the video time codes found in the related transcripts.
The software that has been used for the annotation of the GSL sign language corpora is iLex (Hanke & Storz,.2008).
This procedure, even though it may seem trivial, entailed some major difficulties as, for example, the correct time boundaries extraction out of the iLEX transcript files.
Using information from transcript files, the phrases of interest were located and a list of them was created. According to this list and the transcription file, where information on video time codes was available, the original video files were chopped in smaller ones so that every phrase of interest was entailed into a separate video file.

#### 3.1.2 Elicitation procedure – Recording sessions

The video files of the selected phrases were employed to construct the elicitation material to be presented to the GSL native signer. The signer was asked to repeat the phrase that he saw on the monitor, as close to the original production as possible. The signer was allowed as many repetitions as he wished.
Unfortunately, in naturally uttered signed speech, signers very rarely perform pieces of language that can be reproduced with the minimum set of instructions by other signers. This is mainly the reason why the original phrases in which the 113 lemmas were found served only as a pilot study for the data acquisition process of the set of formally defined simple phrases presented in 3.2.

### 3.2 Formally Defined Simple Phrases

The above mentioned procedure as well as the repository of the 137 original phrases of Task 4 functioned as the cast upon which 56 formally defined phrases were produced. These are 56 phrases that were put together by means of gloss ordering, which combined the 113 lemmas into phrases that obey the grammar rules of GSL.
The 56 phrases were evaluated by a native GSL signer and they are partitioned in the following sets:

a) *Phrases Set I (PS1):* Contains a set of 20 simple continuous phrases;
b) *Phrases Set II (PS2):* Contains a set of 28 of slightly more complex continuous phrases;
c) *Supplementary Phrase Set III (PS3):* Contains a set of 8 phrases to lexically cover for missing signs.

#### 3.2.1 Elicitation procedure – Recording sessions

Since the formally defined phrases are phrases that were put together by ordering the involved glosses, there were no available video to show to the signer.
During the recordings a GSL interpreter performed each phrase and, if needed, explained each phrase to the Deaf adult native signer. There were no limitations in the number of repetitions to be performed other than the fatigue of the signer.

### 3.3 The signers

Four signers in total participated in these recording sessions. Only two out of the four signers performed all the above phrase sets (I, II & III) as well as the list of the 113 lemmas (Lexicon (L)) in isolated mode, three times each. Their recordings served as a pilot test bed upon which the final recordings were based.
The final recordings took place with two signers, the Official Signer A" and the "Official Signer B" who performed multiple times the 56 phrases as well as the list of the 113 lemmas of the Lexicon (L). Their acquired data served as the database used in experimentation.

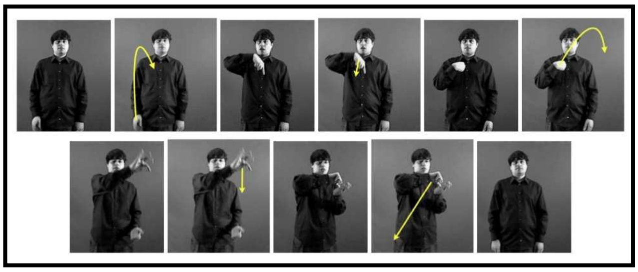


Figure 1: Sample of continuous signing utterance: "TOMORROW I ARRIVE ATHENS" from the GSL Continuous Phrases Dataset (arrows indicate each transition)

## 4. Data acquisition

The data acquired consisted of:

- High Definition (HD) appearance data employing a High Definition camera with frame rate 25 fps
- Depth and appearance data employing a Kinect sensor and
- Skeleton tracking as obtained utilizing the depth data from the Kinect sensor.
- To collect as much data as possible a second Kinect sensor was used to record also the interpreter.

Signer and interpreter were place opposite one another and the Kinect sensors were placed in the middle, one facing the interpreter and the other facing the signer.
Each camera/sensor was controlled by a different person. The acquisition was supported by a moderator who supervised the whole procedure, annotated mistakes,

stated the need of extra repetitions, and marked transcriptions updates.

Figure 2: Setup studio setting for the GSL Phrase Corpus

## 5. Discussion

Herewith presented newly acquired corpus features:

1. topic specific linguistic content
2. structure that simulates simple phrases
3. sharing of linguistic content/vocabulary with a larger and more complex continuous corpus, which can be employed in parallel
4. acquisition of High Definition video data
5. parallel acquisition with the recent high-tech Kinect sensor accounting for both Depth and Skeleton Tracking.

This data set served the purpose of experimentation towards development of continuous sign language recognition algorithms.

Although the corpus is of limited scale, the above features render these data a highly appealing test-bed for interdisciplinary research in the domain of Sign Language and Gesture technology.


## 6. Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/ 2007-2013) under grant agreement No 231135. We would also like to thank Helen Cooper in University of Surrey for providing the Kinect acquisition software.


## 7. References


Hanke, T. (2002). iLex - A tool for sign language lexicography and corpus analysis. In Proceedings of the 3rd International Conference on Language Resources and Evaluation. Las Palmas de Gran Canaria, Spain. ELRA Paris pp. 923—926.

Hanke, T. (2004). HamNoSys – representing sign language data in language resources and language processing contexts. In LREC 2004, Workshop proceedings: Representation and processing of sign languages. ELRA Paris, 2004, pp. 1-6.

Hanke, T. iLex: http://www.sign-lang.uni-hamburg.de/ilex/

Hanke, T., Storz, J. (LREC 2008) iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In LREC 2008, Work shop proceeding of the 3rd Workshop on the Representation and Processing of Sign Languages, Maroko 2008.

Johnson, R. E. and Liddell, S. K. (2011). A segmental framework for representing signs phonetically. sign language studies. 11(3).

Matthes, S., Hanke, T., Storz, J., Efthimiou, E., Dimou, A-L. Karioris, P., Braffort, A., Choisier, A., Pelhate, J. and Safar, E. (2010). Elicitation Tasks and Materials designed for Dicta-Sign's Multi-lingual Corpus. In Dreuw, P. et al., (eds.), *LREC-2010, Proceedings of 4th Workshop on Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pp. 158-163.

Pitsikalis, V., Theodorakis, S., Vogler, C., Maragos P. (2011) Advances in Phonetics-based Sub-Unit Modeling for Transcription Alignment and Sign Language Recognition, Workshop on Gesture Recognition and launching of a Benchmark program (CVPR2011), 2011 Colorado, Springs, USA.

# Improvements of the Distributed Architecture for Assisted Annotation of video corpora

**Rémi Dubot, Christophe Collet**

IRIT
Université de Toulouse
remi.dubot@irit.fr, christophe.collet@irit.fr

**Abstract**

Progress on automatic annotation looks attractive for the research on sign languages. Unfortunately, such tools are not easy to deploy or share. We propose a solution to uncouple the annotation software from the automatic processing module. Such a solution requires many developements: design of a network stack supporting the architecture, production of a video server handling trust policies, standardization of annotation encoding. In this article, we detail the choices made to implement this architecture.

**Keywords:** Annotation, Sign Languages, Automatic Annotation, Distributed Architecture

Linguists need annotations of sign language video corpora. Video corpus annotations are mainly done manually on Annotation Tools (ATs). This work is really time-consuming and frequently repetitive. To support this claim, we can consider the literature regarding annotation tools of the last couple of years. There are two remarkable trends: collaborative annotation and automatic annotation. Works on collaborative annotation focus on the workflow (Hofmann et al., 2009; Brugman et al., 2004). About automatic annotation, we had a look on ELAN and ANVIL. ELAN (Auer et al., 2010) provides a plug-in interface, called "Recognizer API", for automatic processing. A trick is given to run the module on a different machine through the network but it is restricted to local network. ANVIL (Kipp, 2010) provides different kinds of processing based on co-occurrence measurements on annotations and motion characteristics extraction from motion capture.

As far as we know, most of automatic tools are developed as standalone prototyping software, finally remaining as in-house systems that continue to be used only by the team that have developed them. Barriers to their use in production are numerous: "temporary" privacy of sources and executables, deployment difficulty (run only in a specific environment, need a powerful machine), integration constraints (incompatible programming languages, running on different OS, etc.). We think that the architecture we have developed greatly simplifies this integration.

We have designed an open and distributed system for the integration of automatic processing in annotation tools. We consider that it will encourage collaboration initiatives. Our idea is to let every part of the system do what it does best. Each automatic processing tool will run on the module fitting the best. Videos are hosted by a dedicated server. And finally, manual annotation will be done on the annotator's computer.

What we provide is the solution to make all this working together. We see three parts in this problem:

- Service discovery,
- Security,
- Compatibility.

This article first shows the global architecture and the agents involved. Then, it details stage by stage the stack of protocols. Finally, it presents, as an evaluation, the automatic annotation modules already done and a preview of the potential offered.

## 1. Architecture

An early version of this architecture was presented in (Collet et al., 2010). At that time, the architecture was only partially implemented. The finalization leads us to make several adjustments and the new developed architecture is presented now. The system is based on four types of agents, from which three have been already presented in the introduction:

- The first type of agent is Annotation Tools (AT). Any annotation tool can be extended to allow its integration in the architecture. We already provide an annotation tool: AnColin.
- The second type of agent is Automatic Annotation Assistants ($A^3$).
- The third type of agent is the server hosting the videos (VFS: Video-File Server)
- The last type agent is a service directory, it is in charge to reference $A^3$s. When an annotator wants to use an automatic process, his AT will retrieve the list of all currently available functions from the $A^3$S.

These four agents are represented in figure 1.

We call an instance of this architecture a Distributed Annotation System (DAS).

## 2. Stack

### 2.1. Network

As agents are not necessarily located in the same local network, they communicate over the Internet. One of our objectives is to facilitate the deployment of automatic processing, this implies to let the implementation of the processing part as free as possible from constraints coming from the architecture. The consequence is a high heterogeneity

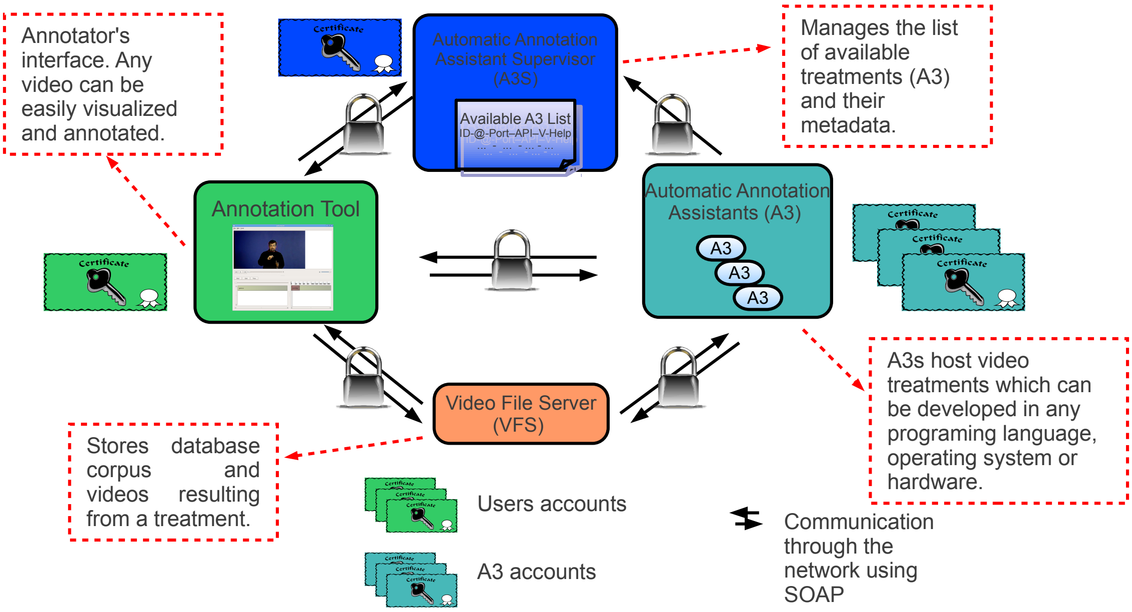



Figure 1: Overview of the system

in terms of programming language, operating system, etc. To insure the low level compatibility between elements we use SOAP (W3C, 2007). SOAP is an open source, multi-platform middle-ware. It uses the HTTP protocol for communications which is important as it allows to pass through proxies. Because some information traveling on the network might be confidential, a TLS (former SSL) layer is added under SOAP. This layer encrypts all the communications. We will see below that TLS provides us several more services.

The Internet, SOAP and TLS are enough to have flexible and secure end-to-end communication, now we have to look at how agents find one another on the network. An instance of the architecture must have one VFS and one $A^3$S. It is the solid part of the architecture. Instances of the two other agents, ATs and $A^3$s, have to be configured to use a VFS and an $A^3$S. Therefore the VFS and the $A^3$S must have static addresses. At launch, $A^3$s provide their status and their network addresses to the $A^3$S. ATs get the $A^3$s' addresses with their specifications when it retrieves the list of available functions. This sequence is summarized in the figure 2.

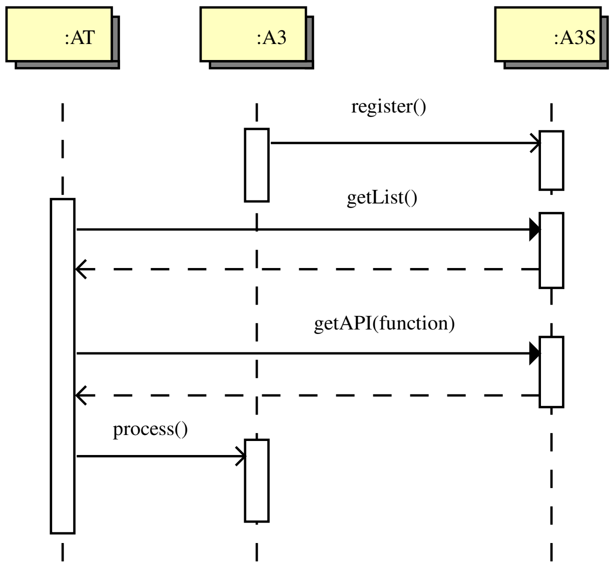



Figure 2: $A^3$S working sequence

## 2.2. Authentication

The authentication, in a communication, refers to the mean of identifying the end-users.

Most of our authorization/trust policy relies on identities consequently we must have a solution to identify agents.

On computers, there is a common solution for authentication which makes use of asymmetric cryptography. Concretely, each agent which has to identify itself must have a certificate, signed by a Certification Authority, which links its identity (name, email, etc.) to its public key. Such certificates are called X.509 certificates. The TLS protocol provides an authentication service using X.509 certificates.

Each agent must have its own certificate. Individual certificates are shown as keys in figure 1. The complete task of authentication is done by TLS.

Signed certificates are not necessarily free, in most cases Certification Authorities sell this service at high costs.

However, there are two low cost solutions:

- When many sites are involved, we recommend the use of certificates focusing only on the email address which are generally free.
- When one or few sites are involved, it is possible to setup a local Certification Authority.

### 2.3. Authorization

The authorization refers to rights policy, agents having the possibility of knowingly choose whether to serve another agent or not.
Videos are a sensitive point, mainly because they are under restrictive image rights. Consequently, the VFS has to manage trust in AT users and $A^3$s regarding the videos. Our implementation of the VFS allows a fine rights management for user access to videos and uses a mechanism of jobs to limit to its minimum the data given to $A^3$s. This is symbolized by certificate collections under the VFS in figure 1.
$A^3$S and $A^3$s might manage authorization too, but currently we do not see any reasons not to let them publicly available.
Finally, as ATs do not provide any service to other agents, there is no authorization to manage on this side.

### 2.4. Encoding

We use Annotation Graphs (Bird and Liberman, 2001) as our standard annotation structure for exchanges. However, the AG model was designed to be as general as possible and consequently has gaps. To overcome this, we have made an extension of AGs. This extension is backward compatible with AGs. This means all ATs handling AGs (ELAN, ANVIL, etc.) are already able to deal with the files generated by an $A^3$. We will not detail all the features brought by this extension, only the two directly linked to the distributed architecture. First, we add the concepts of a track (identical to the concept of a tier) and a group of tracks forming a hierarchy (a tree or a lattice). Second, we add the concept of a type. Here, we talk about types for the content of annotations (the values inside segments). The need of track hierarchy is self-evident. We are going to address in detail the question of why and how to type the values.
Untill now, nothing has been done in the field of annotation encoding to encourage data homogeneity. Consequently, there is a high heterogeneity in data encoding. It leads to problems like “In these vectors, which component is the horizontal one?” or “How was this bounding box encoded? two corners? a corner and a size?”, etc. And while it is annoying for humans, it is really problematic for automatic tools. The compatibility between $A^3$s depends on the homogeneity of data. An $A^3$ producing bounding boxes and an other processing bounding boxes must share the same encoding, even if there were developed fully independently. This is a constraint to allow users to process the results of the first $A^3$ with the second one. The homogeneity is also necessary to have a smart display and edition of annotations. Our solution is to affect types (as in computer languages) to annotation data. The type system we present below has been able to handle all annotations we use in our team currently.
We split the task in two parts:

- Making a type system able to deal with all kind of data appearing in annotations.
- Making types for all the common data and collect them all in a library.

#### 2.4.1. Types

To describe types, the model defines 3 atomic types and 4 construction rules. The 3 atomic types are:

- `Integer`: An integer.
- `Float`: An approximation of a real number.
- `String`: A string.
- `Empty`: This is intended to be a base to build a kind of boolean types. It is used when the information is contained in segments' positions and segments' values are meaningless.

The 4 construction rules are:

- `Copy`: Makes a copy of a type. Allows to makes synonyms.
  Example of definition:
  `Distance:= Copy(Float).`
  Example of valid instance of `Distance`:
  `12.3.`
- `List`: Describes a list of values sharing the same type.
  Example of definition:
  `Glose:= List(String).`
  Example of valid instance of `Glose`:
  `["to drive","car"].`
- `Struct`: Describes a set of named and typed fields.
  Example of definition:
  ```
  HeadPose:= Struct(roll:Float, yaw:Float,
                    pitch:Float).
  ```
  Example of valid instance of `HeadPose`:
  `{roll:3.14, yaw:42.0, pitch:2.72}.`
- `Union`: Allows to take values between multiple types.
  Example of definition:
  `Head:= Union(HeadPose,String).`
  Example of valid instance of `Head`:
  `"profile".`
- `Constant`: Allows to make a constant of a given type.
  Mainly used coupled with `Union` for enumerations.
  Example of definition:
  `hello:= Constant(String,"hello").`
  Example of enumeration `Head`:
  `voc:= Union(Constant(String,"hello"),Constant`

The abstract type system may be extended if needed. But we want the type encoding to stay stable. The format we chose is XML Schema (Fallside and Walmsley, 2004). XML Schema is far more generic than our type system and allows many extensions of the abstract system.
We give translation schemes between abstract type system and XML Schema.

#### 2.4.2. Common Type Library

To standardize the encoding of annotations, we maintain a standard type library which takes the form of an XML Schema Definition (XSD). It is intended to be online.
The detail of the current content of the CTL is the following:

- `Vector2D`.
- `Point2D`.
- `BoundingBox`: Encodes bounding box with the upper left corner as a `Point2D` and the size as a `Vector2D`.
- `List2DPoint`: Encodes a list of Point2D.
- `Hamnosys`: Encodes Hamnosys description.
- `Empty`.

This list will grow quickly as soon as $A^3$s will be used in production.

## 3. Use

We maintain our own AT, AnColin, which fully integrates this architecture. Up to now, four $A^3$s has been developed: Signing detection, sign segmentation, body part tracker (Gonzalez and Collet, 2011) and facial feature tracker.
The first $A^3$ concerns the detection of signing activity. In a corpus of dialogue, the signers take turns to talk. This means that there are times where one of the participant is just listening. This $A^3$ detects where in the sequence the informant is actually signing. The core of this $A^3$ has been provided by Helen Cooper from University of Surrey.
The second $A^3$ regards the segmentation of continuous sign language. It uses hand movement analysis to detect limits between signs.
The third $A^3$ tracks hands and head using a particle filter based approach. It robustly handles hand-over-head occlusion using a template before occlusion (Gonzalez and Collet, 2011).
The last $A^3$ uses a small quantity of hand labelled images to learn a set of facial feature trackers, which can be applied across segments of video. The tracker is based on the linear predictor flock method as described in (Ong and Bowden, 2011).
$A^3$s produce a great amount of technical data. Rendered as text, data is difficult to visualize and to edit for humans. The use of automatic processing induces a new need for visualization and editing tools. The type feature is the base for such features. We have introduced two kind of tools to AnColin:

- Head-Up Display (HUD) modules which are in charge to display and edit annotations directly on the video. We currently have two modules: one for bounding boxes, the other for cloud of points.
- Segment display and edition modules which are in charge to display and edit annotations on the tracks. We currently have one generic module able to generate forms from types.

As an example of possible applications, we are able to chain $A^3$s and finally display results with HUDs, making workflows. Examples of workflows already possible:

- $SigningDetection \rightarrow FaceTracking \rightarrow HUD(BoundingBox)$
- $SigningDetection \rightarrow FaceFeatureTracking \rightarrow HUD(PointCloud)$

## 4. Conclusion

We have introduced an architecture for distributed annotations and the protocol stack of communications in this architecture. Everything presented here has been completely implemented. On the other hand, many parts are not stable enough to be used in production. Some elements which were written as prototypes are being rewritten and others are being stabilized. However we have experimented with using the architecture. The list of $A^3$s is currently short but presenting state-of-the-art tools and expected to grow.

## 5. References


E. Auer, A. Russel, H. Sloetjes, P. Wittenburg, O. Schreer, S. Masnieri, D. Schneider, and S. Tschöpel. 2010. ELAN as flexible annotation framework for sound and image processing detectors. In *International workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (LREC)*, Valletta, Malte.

S. Bird and M. Liberman. 2001. A formal framework for linguistic annotation. *Speech Commun.*, 33(1-2):23–60.

H. Brugman, O. Crasborn, and A. Russel. 2004. Collaborative annotation of sign language data with peer-to-peer technology. In *4th International Conference on Language Resources and Evaluation*, pages 213–216.

C. Collet, M. Gonzalez, and F. Milachon. 2010. Distributed System Architecture for Assisted Annotation of Video Corpora. In *International workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (LREC)*, pages 49–52, Valletta, Malte, May. European Language Resources Association (ELRA).

D. C. Fallside and P. Walmsley. 2004. XML Schema: W3C Recommendation. http://www.w3.org/TR/schema.

M. Gonzalez and C. Collet. 2011. Robust body parts tracking using particle filter and dynamic template. In *IEEE Int. Conf. on Image Processing*, pages 537–540, Brussels, Belgium.

C. Hofmann, N. Hollender, and D. Fellner. 2009. Workflow-based architecture for collaborative video annotation. *Online Communities and Social Computing*, pages 33–42.

M. Kipp. 2010. Multimedia annotation, querying and analysis in ANVIL. *Multimedia Information Extraction*, 19.

E.-J. Ong and R. Bowden. 2011. Robust Facial Feature Tracking Using Shape-Constrained Multiresolution-Selected Linear Predictors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:1844–1859.

W3C. 2007. Recommendation of the W3C. http://www.w3.org/TR/soap/.

# Semi-Automatic Annotation of Semantic Relations in a Swiss German Sign Language Lexicon

**Sarah Ebling*, Katja Tissi**, Martin Volk***

* Institute of Computational Linguistics, University of Zurich, Binzmühlestrasse 14, 8050 Zurich, Switzerland
{ebling,volk}@cl.uzh.ch
** University of Applied Sciences of Special Needs Education, Schaffhauserstrasse 239, 8050 Zurich, Switzerland
katja.tissi@hfh.ch

**Abstract**

We propose an approach to semi-automatically obtaining semantic relations in Swiss German Sign Language (Deutschschweizerische Gebärdensprache, DSGS). We use a set of keywords including the gloss to represent each sign. We apply GermaNet, a lexicographic reference database for German annotated with semantic relations. The results show that approximately 60% of the semantic relations found for the German keywords associated with 9000 entries of a DSGS lexicon also apply for DSGS. We use the semantic relations to extract sub-types of the same type within the concept of double glossing (Konrad 2011). We were able to extract 53 sub-type pairs.

**Keywords:** Sign language, Swiss German Sign Language, semantic relations, double glossing

## 1. Introduction

Sign language lexica have been annotated with various types of linguistic information in the past, including semantic relations. For example, Konrad (2011) reports work on the German Sign Language Corpus Project[1] in which the semantic relations synonymy and antonymy were annotated manually.[2] We present an approach to the semi-automatic annotation of semantic relations. We are compiling a German–Swiss German Sign Language corpus of train announcements as part of our efforts in building a machine translation system for this language pair. Departing from the semantic relations obtained, we also experiment with automatically obtaining sub-types that belong to the same type within the concept of double glossing (Konrad 2011, 145–155).

In Section 2 we briefly describe two common sign language notation systems, of which one (Section 2.1) is a coding system, and the other (Section 2.2) is a transcription system according to the typology of van der Hulst and Channon (2010). We then introduce Swiss German Sign Language and describe an existing lexicon for this language (Section 3). This is the lexicon that we will extend with the semantic relations extracted as a result of our approach. We describe our approach in Section 4. In Section 5 we introduce the typology of signs by Johnston and Schembri (1999), which is the basis of the concept of double glossing. Double glossing is described in Section 6, where we also present our approach to automatically extracting sub-types of the same type.

## 2. Sign Language Notation Systems

### 2.1. Gloss Notation

Sign language glosses are semantic representations of signs that usually take the default form of the corresponding spoken language word.[3] For example, in Swiss German Sign Language (cf. Section 3), the gloss GESCHWISTER, a German word, is used to represent the sign for 'siblings'. Glosses can also consist of multiple words, e.g., SICH-SETZEN ('to have a seat').[4]

Glosses allow for alphabetic sorting in a lexicon. However, from a conceptual point of view it is problematic to express the vocabulary of one language (i.e., a sign language) by means of another (i.e., a spoken language). A further problem with glosses is that they are not standardized; the same sign may be denoted with multiple glosses. Moreover, glosses typically convey only limited facial expression and information about body movement. This means that they cannot, e.g., differentiate between different movement paths of the hands through which a signer associates different objects with individual locations in the signing space (Huenerfauth, 2006).

These shortcomings imply that glosses are merely sufficient to refer to entries in a lexicon. For all other purposes, e.g., for investigating the sublexical components of a sign, a more powerful notation system is needed. The Hamburg Notation System for Sign Languages (HamNoSys) (Prillwitz et al., 1989) has been developed for this.

### 2.2. HamNoSys

HamNoSys consists of approximately 200 symbols. It takes explicit account of sublexical components: each of the components handform, hand position (with extended finger direction and palm orientation as subcomponents), location, and movement is transcribed. Figure 1 shows the example of the sign NATION, VOLK ('nation', 'people') in Swiss German Sign Language that contains one instance of each component.[5]

---

[1] `http://www.sign-lang.uni-hamburg.de/dgs-korpus/`

[2] This was done based on relations between the underlying images of signs, i.e., based on iconicity.

[3] By 'spoken language' we mean a language that is not a sign language.

[4] We follow the convention of writing glosses in all caps.

[5] A sign generally consists of at most two syllables, with the maximum syllable represented as Hold–Movement–Hold, as in the sign BASEL ('Bâle') in Swiss German Sign Language.

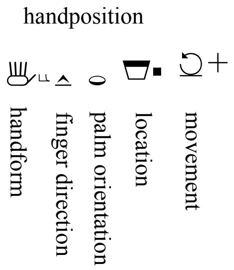



Figure 1: HamNoSys sublexical components for the sign NATION, VOLK (‘nation’, ‘people’) in Swiss German Sign Language

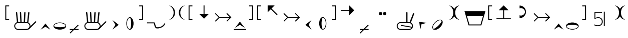


Figure 2: HamNoSys transcription of GEBÄRDENSPRACHKURS (‘sign language course’) in Swiss German Sign Language

HamNoSys is machine-readable: it offers an XML representation, the Signing Gesture Markup Language (SiGML) (Elliott et al., 2000), which can be used to drive an avatar.
The most significant disadvantage of HamNoSys is that its inventory does not include symbols to encode non-manual features. A set of two-letter tags was introduced in the ViSiCAST project to capture aspects like shoulder, body, and head movement as well as eye gaze and facial expression (Hanke, 2001). However, these tags are not officially part of the HamNoSys symbol set. Moreover, HamNoSys transcriptions are rather complex, as can be seen from Example 3, the transcription of the sign GEBÄRDENSPRACHKURS (‘sign language course’) in Swiss German Sign Language. Nevertheless, HamNoSys is considered by many to be the state-of-the-art sign language notation system. It is also used in a lexicon of Swiss German Sign Language. In the following section, we describe Swiss German Sign Language in more detail and introduce the lexicon.

## 3. Swiss German Sign Language

Swiss German Sign Language (Deutschschweizerische Gebärdensprache, DSGS) is the sign language of the German-speaking area in Switzerland. It has approximately 6000 users (Lewis, 2009) distributed across five different dialects. We focus on the Zurich dialect. The mouthings used in DSGS are derived from Standard German rather than from one of the Swiss German dialects. In 1996, work started on a DSGS lexicon (Boyes Braem, 2001). Currently the lexicon contains about 9000 signs. Each sign is represented with a German gloss as well as with a set of German keywords. Keywords are included since glosses often reflect only one meaning of the sign or do not sufficiently distinguish the sign’s meaning from the meanings of other signs. For example, there are two glosses GLAUBEN: one () carries the keywords *glauben*, *annehmen* (‘believe’, ‘assume’), the other () the keywords *glauben*, *Glaube*, *gläubig*, *religiös* (‘believe’, ‘faith’, ‘religious’).
In the lexicon, a video clip is available for the citation form of the sign. The possible modifications (e.g., pluralisation, verbal aspect) as well as additional linguistic information are given in textual form. Efforts are made to transcribe all of the signs in HamNoSys.
Our aim is to enrich the DSGS lexicon with information on semantic relations. We identify semantic relations in a two-step process: first, a set of pairs of semantically related entries is generated automatically; second, the set is filtered through manual screening by a native signer. In what follows, we describe the first step of this process in more detail.

## 4. Identifying Semantic Relations

### 4.1. Step 1: Automatic Identification

| Relation | Word Class N | A | V | Reverse Rel. | Type |
|---|---|---|---|---|---|
| hyperonymy | ✓ | ✓ | ✓ | hyponymy | conceptual |
| hyponymy | ✓ | ✓ | ✓ | hyperonymy | conceptual |
| meronymy | ✓ | ✗ | ✗ | holonymy | conceptual |
| holonymy | ✓ | ✗ | ✗ | meronymy | conceptual |
| entailment | ✗ | ✗ | ✓ | | conceptual |
| causation | ✗ | ✗ | ✓ | | conceptual |
| association | ✓ | ✓ | ✓ | | conceptual |
| synonymy | ✓ | ✓ | ✓ | synonymy | lexical |
| antonymy | ✓ | ✓ | ✓ | antonymy | lexical |
| pertonymy | ✓ | ✓ | ✓ | | lexical |
| participle | ✗ | ✓ | ✗ | | lexical |

Table 1: Sub-types of semantic relations in GermaNet

We use the sign language glosses and keywords available for each sign in the DSGS lexicon (cf. Section 3) as indicators of the underlying semantic concepts. We search for these words in GermaNet. GermaNet is a lexicographic reference database for German word senses containing annotations of semantic relations (Hamp and Feldweg, 1997). It is based on the Princeton WordNet for English (Miller, 1993).[6] For the semantic relations in GermaNet to be fully applicable to the entries in the DSGS lexicon, the semantic networks of the two languages DSGS and German would have to be exactly the same. This is not the case; yet we expect the networks to be similar enough for us to use the assumption of concept equality as a heuristic. We believe it is promising to further investigate the applicability of spoken language concepts to sign languages.
The basic units in GermaNet are semantic concepts, which are called *synsets*. Each synset combines a set of lexical units by which the corresponding concept is denoted. Lexical units may take multiple orthographic forms, e.g., *fantastisch* (‘fantastic’) may appear as *fantastisch* or *phantastisch* (new and old spelling, respectively). Lemmas can belong to one of the three word classes noun, verb, and adjective.

(1)
```
<synset id="s2376" category="adj">
```

[6] Wordnets also exist for other languages, including Dutch, Italian, Spanish, German, and French (Vossen, 2004).

```
<lexUnit id="l3637" sense="1" source="core"
namedEntity="no" artificial="no"
styleMarking="no">
<orthForm>erforscht</orthForm>
</lexUnit>
<lexUnit id="l3638" sense="1" source="core"
namedEntity="no" artificial="no"
styleMarking="no">
<orthForm>erkundet</orthForm>
</lexUnit>
</synset>
```

There are two main types of semantic relations in GermaNet: *conceptual relations* hold between synsets, and *lexical relations* hold between lexical units. Table 1 lists the sub-types of semantic relations available in GermaNet: for each relation, the word class(es) with which it occurs (N–noun, A–adjective, V–verb), the reverse relation (where available), and its type (conceptual or lexical) are specified. Example 1 shows the synset `s2376` containing the two lexical units `l3637` and `l3638` that are each represented by a single orthographic form: *erforscht* ('investigated'), and *erkundet* ('explored'). Synonym relations exist between the lexical units.

We use version 6.0 of GermaNet, which contains 93,407 lexical units distributed across 69,594 synsets. 81,852 conceptual relations hold between the synsets, and 3562 lexical relations exist between the lexical units. For each entry pair in the DSGS lexicon we check whether the glosses themselves or one of the keywords on each side are present in the GermaNet database. If this is the case, we extract the set of relations that exist between the lexical units under consideration or between the synsets to which the lexical units belong. For example, a synonym relation exists between the two entries ANGEBOT and ANTRAG (both 'offer'). A hyponym/hyperonym relation exists between the two synsets to which the entries ALGERIEN, ALGERISCH ('Algeria, Algerian') and LAND ('country') belong.

| Sub-type | Absolute count | Percentage |
|---|---|---|
| Hyponym/hyperonym | 5435 | 77.87 |
| Synonym | 817 | 11.70 |
| Meronym | 279 | 4.00 |
| Antonym | 277 | 3.97 |
| Pertainym | 59 | 0.85 |
| Participle | 40 | 0.57 |
| Related-to | 39 | 0.56 |
| Causation | 34 | 0.48 |

Table 2: Relation sub-types extracted along with their frequencies

We identified 6980 relations. The distribution according to relation sub-types is as shown in Table 2. Below is a sample output of our approach (Examples 2 to 7): six antonyms of the sign ALT ('old').

(2) JUNGTIER, JUNGES, JUNG ('young animal', 'young')

(3) FRISCH, NEU ('fresh', 'new')

(4) JUNG, JUGENDLICHER, JUGEND ('young', 'youth')

(5) KLEIN, JUNG, KLEINES, JUNGES ('small', 'young')

(6) NEU, BRANDNEU ('new', 'brand new')

(7) NEU ('new')

### 4.2. Step 2: Manual Screening

Step 2 of our process consists of manually filtering the semantic relations that were retrieved automatically during Step 1. This task was carried out by a native signer who is also a member of our project. We presented her with 500 randomly selected statements of the kind displayed in Examples 8 and 9 and asked her to rate them with True or False. She rated 302 out of 500 statements with True (60.4%) and 198 statements with False (39.6%).

(8) LAND ('country') is hypernym of ALGERIEN, ALGERISCH ('Algeria, Algerian') ✓

(9) GRUND, UMSTAND, MOTIV ('cause, reason') und LANDSCHAFT, UMGEBUNG, GEGEND ('landscape, neighbourhood') have the same meaning ✗

(10) GOTT, ALLMÄCHTIGER, HERR, VATER ('God', 'Almighty', 'Lord', 'Holy' 'Father') is hyponym of GROSSVATER, OPA, GROSSPAPA ('grandfather', 'granddad').

The statements rated as False are relations that do not apply for DSGS. There are two possible reasons for this:

1. Our system found the relation at hand based on a sense of a German keyword that was incorrect in the given context. Hence, the relation is also not valid for German. For example, the (false) statement shown in Example 10 is due to an ambiguity of the word *Vater*, which can mean both 'Holy Father' as well as 'father'. In this case, the former ('Holy Father') is the intended meaning, whereas the hyponym GROSSVATER, OPA, GROSSPAPA ('grandfather', 'granddad') proposed by our system is based on the latter meaning ('father'), which in this context is incorrect.

2. The relation is valid for German but not for DSGS. Hence, it accounts for a difference in the semantic concepts of German and DSGS. For example, in German, *Künstler, Künstlerin, Kunstschaffender, Kunstschaffende* ('artist') is a hypernym of *Musiker, Musikerin, Musikant, Musikantin* ('musician'). In DSGS, however, KÜNSTLER, KÜNSTLERIN, KUNSTSCHAFFENDER, KUNSTSCHAFFENDE is restricted to the meaning of a visual artist, i.e., a painter. Hence, the relation does not hold for DSGS. Similarly, TRAINING, TRAINIEREN ('practice') is confined to the domain of physical activity in DSGS, whereas it may involve any sort of training in German. Hence, *Training, trainieren* is a valid hyponym of *Lehren, unterrichten, schulen, belehren, erklären* ('instruct', 'teach') in German but not in DSGS. As a third example, *Haushalt,*

*haushalten* ('household') is a hypernym of *Buchhaltung, buchhalten* ('book-keeping') in German but not in DSGS, where BUCHHALTUNG, BUCHHALTEN cannot have a financial aspect to its meaning (as in 'financial household').

We will incorporate the semantic relations identified as valid for DSGS into the DSGS lexicon (cf. Section 3).
Semantic relations between signs can be used to obtain sub-types that belong to the same type within the concept of *double glossing* (Konrad, 2011, 145–155).[7] Double glossing builds on the typology of signs by Johnston and Schembri (1999). In what follows, we introduce this typology. In Section 6.1 we then explain the concept of double glossing. In Section 6.2 we present our approach to automatically generating sub-types of the same type.

## 5. Typology of Signs

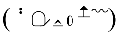
Johnston and Schembri (1999) assume two main types of signs in sign languages: productive signs and conventional signs.[8] For productive signs, the sum of meanings of the sublexical components (cf. Section 2.2) yields the overall meaning of the sign. Productive signs are iconic: their form can be traced back to an underlying image. Moreover, they are context-dependent. They are derived spontaneously, which is why they are also referred to as "on-the-spot" signs. It is for this reason also that productive signs do not have a stable citation form; hence, they do not appear in the lexicon.
Conventional signs are idiomatic in that their overall meaning is not composed merely of the meanings of the sub-lexical components. They are "off-the-shelf" signs, which makes them similar to spoken-language words. Most conventional signs are originally iconic, yet have developed into self-contained form-meaning units that can be used without drawing on iconic value. Nevertheless, the iconic value of many conventional signs can be reactivated by modifying the sign, e.g., by pluralising it. This process is called delexicalisation, or reiconisation. By contrast, the process during which productive signs are turned into conventional signs is referred to as lexicalisation.
Konrad (2011) extends the scheme of Johnston and Schembri (1999) by further differentiating conventional signs according to two usages: a conventional usage, and a productive usage. The usages are characterized by the way in which signs are combined with mouthings: a conventional usage of a conventional sign implies a habitual combination of sign and mouthing, while a productive usage is constituted by an unusual or occasional sign-mouthing combination.
Figure 3 summarizes the typology of signs proposed by Johnston and Schembri (1999) and extended by Konrad (2011). It is important to note that the distinction between productive and conventional signs is one of degree rather than kind. The concept of double glossing applies to conventional signs only. This is the part enclosed in the dotted frame in Figure 3.

## 6. Double Glossing

### 6.1. Concept Overview

The concept of double glossing includes a two-level hierarchy: the upper level consists of (lexical) types, the lower level consists of sub-types that are identical in form, underlying image, and image producing technique.[9] Identical images and image producing techniques imply that signs have the same iconic value. Hence, the governing principle behind double glossing is iconicity. As an example, consider the type FLACH ('flat'), which unites, among others, the sub-types BASIS ('base'), BODEN ('ground'), FELD ('field'), and TISCH ('table') (Konrad, 2011). The common underlying image here is that of a flat surface. Other examples of type names are VIERECK ('square'), BEREICH ('area'), BEHÄLTER ('container'), RUND ('round shape'), and MATERIAL ('material') (Konrad, 2011).

### 6.2. Identifying Sub-Types from Semantic Relations

Departing from the semantic relations obtained from our two-step process (cf. Section 4), our goal was to automatically extract sub-types of the same type within the concept of double glossing. Hence, we were concerned with semantically related sub-types. This is the area outlined in grey in Figure 4. Outlined in black is the area in which double glossing as a whole takes place.
We looked at pairs of hyponyms of the same hypernym that are form-equivalent, i.e., have the same HamNoSys notation. For example, BACH ('brook') and FLUSS ('river') have the same form (  ) and the same GermaNet hypernym, *Gewässer* ('stretch of water'). The heuristic here is that two form-equivalent signs that are semantically related have the same underlying image (in this case, a tracing movement sketching the flow of a river or a brook, cf. Figure 5) and the same image producing technique. Note that this is not always true. Konrad (2011, 175) names the example of two different signs KRIPPE that are equivalent in their underlying image but not equivalent in form and image producing technique.
We identified 53 sub-type pairs of this kind. Note that it would be conceptually wrong to name the corresponding types after the mutual hypernyms, e.g., to choose GEWÄSSER as the name of the type comprising the sub-types BACH and FLUSS; this is because not all sub-types of a type are necessarily semantically related.

## 7. Conclusion

In this paper we showed that using a spoken-language resource to obtain information about a sign language leads to 60% correct semantic relations. We used GermaNet to automatically obtain candidates for semantic relations in Swiss

[7] Efforts are underway to include double glossing in the DSGS lexicon. Double glossing is implemented in iLex (Hanke and Storz, 2008).

[8] A third type exists that subsumes various smaller categories, e.g., finger alphabet, initialised signs, index signs, number signs, gestures, etc.

[9] The terms *type* and *sub-type* are from König et al. (2010). Note that they are different from the terms of the same name introduced in the context of semantic relations in Section 4. Image producing techniques are described in Langer (2005).

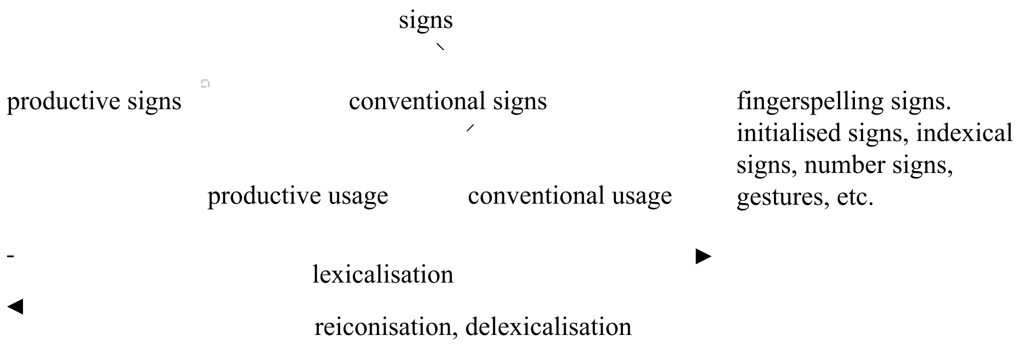



Figure 3: Typology of signs based on Johnston and Schembri (1999) and Konrad (2011)

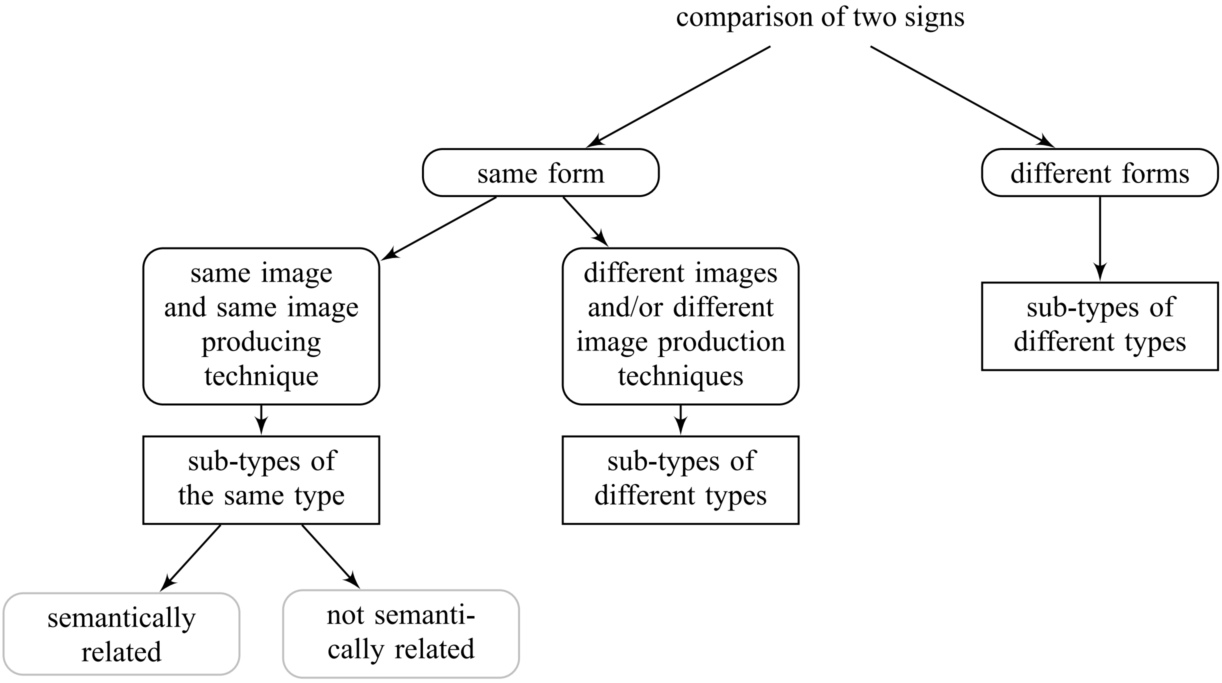



Figure 4: Double glossing: Identifying sub-types of the same type (based on Konrad (2011))

Figure 5: Form of the signs BACH ('brook'), FLUSS ('river'), and WEG ('path') (source: DSGS lexicon)

German Sign Language. We also looked at semantically related sub-types within the concept of double glossing. We extracted 53 sub-type pairs. Our approach contributes to a comparison of semantic and iconic networks: it yields further insight into the question raised by Konrad (2011, 238), "ob und inwieweit ikonische und semantische Netzwerke zur Deckung gebracht werden können" ('whether and, if so, how iconic and semantic networks can be brought to overlap').

In the future, we would like to look into ways of operationalizing the criteria of 'identical image' and 'identical image producing technique' so as to be able to extract more sub-types automatically. Given the high cost of double glossing (Konrad, 2011, 151) it seems reasonable to automate as much of this task as possible.

We will also investigate the possibility of extracting additional semantic relations for German from Swiss German Sign Language form equivalences. In doing so, we will pursue the opposite direction to that commonly investigated, i.e., we will attempt to arrive at additional knowledge of spoken languages using information from sign languages.

## 8. Acknowledgements

We thank Penny Boyes Braem for providing her DSGS lexicon as well as giving advice at various stages of this work, and Reiner Konrad for his helpful comments.

## 9. References

Penny Boyes Braem. 2001. A multimedia bilingual database for the lexicon of Swiss German Sign Language. *Sign Language & Linguistics*, 4:133–143.

R. Elliott, J. R. W. Glauert, J. R. Kennaway, and I. Marshall. 2000. The development of language processing support for the ViSiCAST project. In *Proceedings of the fourth international ACM conference on Assistive technologies*, Assets '00, pages 101–108, Arlington, Virginia, United States. ACM.

Birgit Hamp and Helmut Feldweg. 1997. GermaNet – a Lexical-Semantic Net for German. In *Proceedings of ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*, Madrid.

Thomas Hanke and Jakob Storz. 2008. iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In *LREC 2008 Workshop Proceedings*, pages 64–67.

Thomas Hanke. 2001. ViSiCAST Deliverable D5-1: Interface Definitions. Technical report, ViSiCAST project.

Matt Huenerfauth. 2006. *Generating American Sign Language Classifier Predicates for English-to-ASL Machine Translation*. Ph.D. thesis, University of Pennsylvania.

Trevor Johnston and Adam Schembri. 1999. On Defining Lexeme in a Signed Language. *Sign Language & Linguistics*, 2:115–185.

Susanne König, Reiner Konrad, Gabriele Langer, and Rie Nishio. How much top-down and bottom-up do we need to build a lemmatized corpus? Poster presented at the *Theoretical Issues in Sign Language Research Conference (TISLR 10)*, Sept 30–Oct 2, 2010, Purdue University, Indiana, USA. Online: `http://www.purdue.edu/tislr10/pdfs/Konig%20Konrad%20Langer%20Nishio.pdf`.

Reiner Konrad. 2011. *Die lexikalische Struktur der Deutschen Gebärdensprache im Spiegel empirischer Fachgebärdenlexikographie. Zur Integration der Ikonizität in ein korpusbasiertes Lexikonmodell*. Gunter Narr Verlag.

Gabriele Langer. 2005. Bilderzeugungstechniken in der Deutschen Gebärdensprache. *Das Zeichen*, 70:254–270.

M. Paul Lewis, editor. 2009. *Ethnologue: Languages of the World*. SIL International, Dallas, Tex., 16th edition. Online version: `http://www.ethnologue.com/`.

George A. Miller. 1993. WordNet: a Lexical Database for English. In *Human Language Technology, Proceedings of a workshop held at Plainsboro, New Jersey*, pages 409–409, Plainsboro, New Jersey.

Siegmund Prillwitz, Regina Leven, Heiko Zienert, Thomas Hanke, and Jan Henning. 1989. *HamNoSys: Version 2.0: Hamburg Notation System for Sign Languages. An introductory guide.* Signum, Hamburg.

Harry van der Hulst and Rachel Channon. 2010. Notation systems. In Diane Brentari, editor, *Sign languages*, Cambridge language surveys, pages 151–173. Cambridge University Press, Cambridge.

Piek Vossen. 2004. EuroWordNet: A Multilingual Database of Autonomous and Language-Specific Wordnets Connected via an Inter-Lingual Index. *International Journal of Lexicography*, 17:161–173.

# Sign Language technologies and resources of the Dicta-Sign project

**Eleni Efthimiou[1], Stavroula-Evita Fotinea[1], Thomas Hanke[2], John Glauert[3], Richard Bowden[4], Annelies Braffort[5], Christophe Collet[6], Petros Maragos[7], and François Lefebvre-Albaret[8]**

[1]Institute for Language and Speech Processing, [2]Universität Hamburg, [3]University of East Anglia, [4]University of Surrey, [5]LIMSI/CNRS, [6]Université Paul Sabatier, [7]National Technical University of Athens, [8]WebSourd

[1]{eleni_e,evita}@ilsp.gr, [2]thomas.hanke@sign-lang.uni-hamburg.de, [3]J.Glauert@uea.ac.uk, [4]R.Bowden@surrey.ac.uk, [5]annelies.braffort@limsi.fr, [6]collet@irit.fr, [7]maragos@cs.ntua.gr, [8]francois.lefebvre-albaret@websourd.org

**Abstract**

Here we present the outcomes of Dicta-Sign FP7-ICT project. Dicta-Sign researched ways to enable communication between Deaf individuals through the development of human-computer interfaces (HCI) for Deaf users, by means of Sign Language. It has researched and developed recognition and synthesis engines for sign languages (SLs) that have brought sign recognition and generation technologies significantly closer to authentic signing. In this context, Dicta-Sign has developed several technologies demonstrated via a sign language aware Web 2.0, combining work from the fields of sign language recognition, sign language animation via avatars and sign language resources and language models development, with the goal of allowing Deaf users to make, edit, and review avatar-based sign language contributions online, similar to the way people nowadays make text-based contributions on the Web.

**Keywords:** Sign language technologies, multilingual sign language resources, sign language Web 2.0 applications, Deaf communication, human-computer interfaces (HCI) for Deaf users

## 1. The Dicta-Sign framework

When introducing the components and objectives of the Dicta-Sign project (FP7-ICT) at the 4th Workshop on the Representation and Processing of Sign Languages in the framework of LREC-2010 (Efthimiou et al., 2010), we underlined its major aim, which was to improve the state of web-based communication for Deaf people by allowing use of sign language in a way similar to what holds for oral languages. In the three years of its life-cycle, Dicta-Sign significantly progressed knowledge in the domains of recognising and generating authentic signing. It also created significant sign language resources to be exploited in future research both in sign language linguistics and sign language technologies. Dicta-Sign developed several technologies demonstrated via a sign language-aware Web 2.0 application, a wiki, combining work from the fields of sign language recognition, sign language animation via avatars and sign language linguistics. The goal is to allow Deaf users to make, edit, and review online sign language content presented via a signing avatar, in a similar way to how people make text-based contributions on the Web. In this context, several successful proof-of-concept demonstrators have been created. Furthermore, to serve its technological goals, Dicta-Sign created a multilingual lexicon and annotated comparable corpora resources for four European sign languages: Greek, British, German, and French Sign Language, which were exploited in the definition of models for sign language lexicon and grammar, they informed the project's technologies and are now available to the international research community.

Dicta-Sign (http://www.dictasign.eu/) actually developed three proof-of-concept prototypes: a Search-by-Example Tool that works as a sign language dictionary (Cooper et al,. 2011), a Sign Look-up Tool that allows for multilingual translation of signs in the four project sign languages (Elliot et al., 2011), and a Sign-Wiki, a sign language-based Wiki which enables end-users to view, edit, create and upload sign language content, while preserving their anonymity through the use of a signing avatar for the display of so created content.

## 2. Technological advancements and multilingual SL resources

Dicta-Sign contributed new knowledge in regard to sign language resources creation and the whole range of sign language technologies. The scientific domains addressed are:

- Image processing
- Advanced computer vision
- Statistical methods for continuous sign recognition with multimodal fusion and adaptation
- Virtual human technology
- Sign language modelling
- Grammar and lexicon design and development
- Corpus construction

The project has investigated sign language processing in different input and output modalities Within these input and output modalities, different channels (manual signing, eye gaze, facial expression and body posture) present complementary communicative information. Dicta-Sign has analysed and synthesised information within these channels to encompass the full range of expressiveness within sign languages.

In so doing, Dicta-Sign has achieved the following goals:

- Establishment of the world's only extensive parallel multilingual corpus of annotated sign language video data to inform further sign language research and provide a resource to inform sign language processing technology (http://www.sign-lang.uni-hamburg.de/dicta-sign/portal/).
- Development of advanced sign language annotation tools to provide access to this corpus integrating the

recognition, generation and animation technologies developed within the project (Dicta-Sign technical report: D5.4; Gonzalez & Collet, 2012)

- Use of image processing and computer vision technology and the development of a statistical framework to enhance and progress beyond the state-of-the-art in continuous sign language recognition, also experimenting with sign language linguistic models and exploiting multimodal fusion and adaptation (Dicta-Sign technical reports: D1.3. and D2.3)
- Development and dissemination of a morphophonetically based orthography based upon HamNoSys incorporating the manual and non-manual channels of sign language communication used both for annotation purposes and as an intermediate representation for both sign language generation and recognition within system software,
- Provision of large cross-lingual sign lexicons (http://www.sign-lang.uni-hamburg.de/dicta-sign/consign/demo/cs_list_eng.html),
- Extension of sign language generation and virtual human animation technology to advanced realistic signing (Dicta-Sign technical report: D3.3),
- Development of a prototype sign-to-sign translation service (Dicta-Sign technical report: D7.2).

These objectives are structured through the development and delivery of the following outputs:

- A parallel multilingual corpus for four national sign languages – German, British, French and Greek (DGS, BSL, LSF and GSL respectively).
- A multilingual lexicon of 1000+ signs from each of the four project sign languages,
- A continuous sign language recognition system that has achieved significant improvements and also has researched the novel directions of multimodal sign fusion and signer adaptation,
- A language generation and synthesis component, covering in detail the role of manual, non-manual and placement within signing space,
- Annotation tools which incorporate these technologies providing access to the corpus and whose long term utility can be judged by the up-take by other sign language researchers,
- Three bidirectional integrated prototype systems which showcase the utility of the system components beyond the annotation tools application.

All project prototypes have been exhaustively evaluated by Deaf end users. Their details, emphasising on their relation to project resources, are presented in section 3 below.

## 2.1 Sign language recognition

In order to support work in sign recognition, from its early stages, Dicta-Sign has created 3D data of sign language by shooting the whole corpus with Bumblebee depth capturing cameras, which were incorporated in the studio set up used for the project corpus acquisition. With the arrival of the Kinect, this new device was adopted as a 3D data acquisition instrument and Kinect data have since fed experimentation in both visual tracking and feature extraction as well as in continuous sign recognition research.

### 2.1.1. Visual tracking and feature extraction

In order to recognise isolated signs, a low level feature description of the signers' actions need to be extracted. Visual tracking and feature extraction work in Dicta-Sign has covered several tasks including hand detection and tracking in 2D and 3D (Hadfield & Bowden 2011 and 2012), visual hand detection emphasising on tracking and motion feature extraction (Roussos et al., 2010a, Roussos et al., 2010b), body part detection with estimation from depth (Holt et al., 2011), facial feature localisation and tracking (Ong & Bowden, 2011), face modelling for tracking and related feature extraction (Rodomagoulakis et al. 2011; Antonakos et al. 2012), handshape modelling and related feature extraction (Rousos et al. 2010a; Cooper et al., 2012, Pugeault 2011), as well as mapping of body motion to HamNoSys features.

### 2.1.2. Continuous sign language recognition

For recognition of dictation style sign the project developed a new spatio-temporal feature selection approach called Sequential Pattern Trees (Ong et al,. 2012) which generalises well to unseen signers and forms the heart of the Sign Wiki recognition module. For continuous sign language recognition, the project developed an efficient visual front-end for spatio-temporal processing of the corpus video data and statistical (HMM-based) models for both data-driven and phonetics-based Sub-units (SUs) which have allowed Dicta-Sign to make significant advances in continuous SL modelling and recognition. Specifically, the project's contribution to new knowledge in continuous sign recognition includes: 1) dynamic vs. static data-driven SUs (Pitsikalis et al. 2010, Theodorakis et al. 2011a), 2) phonetics-based SUs (Pitsikalis et al. 2011; Vogler 2011), 3) raw and canonical SU models (Theodorakis et al. 2012, Maragos et al. 2012), which provide novel solutions in case of multiple signers and system adaptation to them, 4) facial event detection and recognition of related linguistic cues (Antonakos et al. 2012), and 5) fusion of multiple cues/modalities for improving continuous sign recognition (Roussos et al. 2010, Theodorakis et al. 2011a, Theodorakis et al. 2011b, Theodorakis et al. 2012). In addition, exploitation of SL grammar rules has also led to interesting results on continuous video data.

## 2.2 Sign language generation and animation

Dicta-Sign has improved the state-of-the-art in synthesis and animation of sign language through 3D virtual characters (Glauert & Elliott, 2011). Synthesis is based on the use of SiGML, Signing Gesture Markup Language, which is an XML language based heavily on HamNoSys, the Hamburg Notation System, since HamNoSys transcriptions can be mapped directly to SiGML. SiGML is primarily implemented through the JASigning software which is supported on both Windows and Macintosh platforms. During the project, enhancements have been made to SiGML that allow precise control of the timing of animations (Jennings et al., 2010). This was exploited to show that an animation synchronised with a sign language video can be produced by annotating the signs used, along with their timings, and playing back HamNoSys transcriptions of the signs (Hanke et al., 2011), where fine control of the location of sign postures and the details of transitions between postures has been provided.

## 2.3 SL linguistic modelling

Work on linguistic modelling was concerned with the high-level linguistic framework relevant to sign languages. The aim was to develop lexicon and grammar models to assist in SL generation and recognition.

On lexical level, the proposed models provide formal descriptions that may be used by a sign recognition system, whose input is a video of signed utterances, or as input to an animation system, whose output is signed by an animated virtual signer. To do so, these models need to include both phonetic representations based on HamNoSys and on the Johnson & Liddell phonetic notation, as well as abstract level representations based on Zebedee[1], a sign description model based on a geometric tool kit and time structuring that allows signs and their variability in the signing space to be specified.

Regarding grammar modelling, the main issues include flexible sign order in sign phrase, signing space representation, non-manual gestures, synchronisation of multiple body articulators, role shift, and prosody (Filhol 2012). Grammar modelling work in Dicta-Sign concentrated on a set of phenomena which provide cues to all above:

1. Enumeration: sets of unordered elements signed in sequence;
2. Alternatives, where options of a choice are listed sequentially;
3. Qualification/naming, where one sign in a sequence denotes an entity specified (named, finger-spelt, adj-qualified) by the others (Filhol & Braffort 2012);
4. Neutral questions, i.e. the case where the speaker is directly asking for an answer which he is not able to predict;
5. Quantifiers (small/big);
6. Announcing titles: announcing the topic of a discourse section.

The first three in the above list cover utterance components structure, the fourth one deals with the whole clause structure, the fifth affects lexical units, while the last one touches upon discourse structure.

## 2.4 Sign language annotation tools

Concerning annotation tools, Dicta-Sign proposed a complete solution that answers the need to integrate various tools into the annotation process.

The proposed solution is specified as a distributed system architecture called the Automatic Annotator Assistants System (Collet, Gonzalez, Milachon 2010). It aims to use software developed in different programming languages, operating systems and software platforms, outsourcing annotation tasks to other machines such as a computing cluster, easily adding automatic processing to support the annotation task within existing annotation tools.

This architecture is multi-platform and multi-language (including RealBasic, C/C++ and Java) and the model used for data exchange is adaptable to a number of annotation formats. It also includes a security mechanism to ensure annotation data, video files and program code integrity (Dubot & Collet 2012). As a prove of concept, we have implemented the following automatic annotation assistants (A3): Body Part Tracking (Gonzalez & Collet, 2011); Facial Feature Tracking (Ong & Bowden, 2011); Signing Detection; Sign Segmentation (Gonzalez & Collet, 2012). All specifications of API for A3 and the source code of A3 template and A3S are publicly available.

## 2.5 The Dicta-Sign sign language resources

Within the Dicta-Sign project sign language resources were compiled for four European languages: British, French, German, and Greek Sign Language. These resources were used to inform progress in other research areas within the project, especially sign recognition, linguistic modelling, and sign generation.

A multilingual lexical database providing a core lexicon of approximately 1000 entries in the four project sign languages was the first resource to be built. The shared list of concepts chosen for the lexicon is of everyday use or specifically related to the field of Dicta-Sign's main topic, European travel. Signs were recorded for each language and annotated by assigning gloss labels, form description (HamNoSys) and a rough meaning. However, the biggest achievement in the area of language resources within the project is Dicta-Sign's multilingual corpus on the domain "Travel across Europe". Prior to the project, parallel corpus collection for sign languages had only been undertaken in minimal sizes or for spoken language simultaneously interpreted into several sign languages, but not for semi-spontaneous signing by native signers, based on a well designed methodology for corpus elicitation, which eliminates dominant oral language interference (Matthes et al. 2010).

Data collection took place in all four countries involved in the project, using seven different cameras to film the informants from different perspectives (front, side, and top view) as well as with additional stereo cameras that provide video footage for automatic processing. In each country, 14 to 16 informants were filmed in sessions lasting about two hours each. This resulted in applicable language data of at least 8 hours per language.

After various steps of video post-processing the data were annotated using the iLex annotation environment (Hanke

[1] http://perso.limsi.fr/filhol/zebedee/index.html

& Storz 2008).
The detailed annotation, conducted for parts of the corpus data, includes segmentation of the continuous signing into individual signs, lemmatisation (i.e. assigning glosses), a form description of the signs using HamNoSys as well as English translation. Additionally, content tags were assigned to most of the corpus data that reflect the topics the informants signed about. These tags allow finding video sections with comparable content across individual informants and languages.
The Dicta-Sign language resources are made available via a dedicated web portal that provides data on different access levels and with different approaches to access the data. The access levels range from publicly available data to restricted access for researchers or future project partners.
Different approaches are offered to access the data, e.g. by selecting a certain language and informant, by choosing a certain task, or by selecting a specific content tag to see data of various languages and informants signing about the same topic.

## 3. Showcase prototypes

The main objectives of Dicta-Sign concentrated towards the development of an integrated framework that allows contributions in the project's sign languages. The realisation of this goal is crucially based on the assumption that Deaf users may directly insert their own content via Kinect devices. This content is recognised by the sign language recognition component and converted into a linguistically informed internal representation, which is used to animate the user's contribution with an avatar, and/or to translate the individual signs composing a specific contribution into the other respective three sign languages, if this is requested by the user.

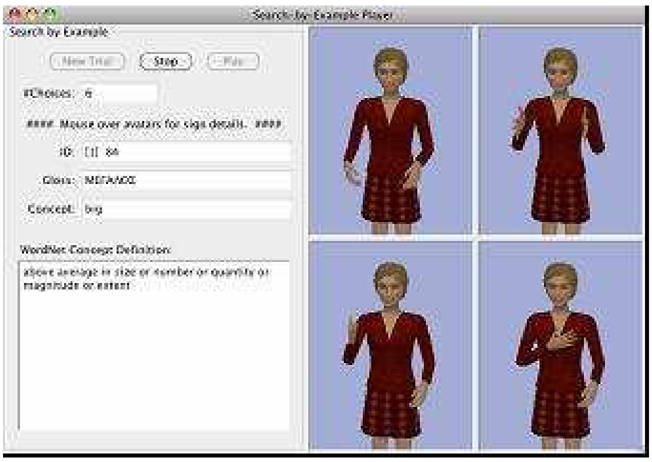


Figure 1: Search-by-Example Tool - Client Display

The project's three proof-of-concept prototypes have successfully demonstrated the feasibility of this initially formulated objective. In all cases, the user may insert his/her own signs or sign utterances either by choosing lemmas from the system's lexicon or by signing in dictation tempo in front of a Kinect camera. All prototypes are Java applications, which use a client-server architecture.
The aim of the *Search-by-Example* prototype (Figure 1), is to allow a Deaf user to perform a sign or a component of a sign, and to search the Dicta-Sign lexical database on the basis of features extracted from the performance. Once candidate lexical signs have been identified, the database delivers information about the form of the sign, in HamNoSys notation, and other linguistic information, that becomes available to the user. An animation of the sign is provided using an avatar.
While there are a few initiatives towards exploitation of sign recognition technologies for lexicon search purposes (Athitsos et al., 2010; Wang et al., 2010), most lexicon applications, known to us, use video to present sign lemmas, while lemma search is based on either groupings according to basic handshape or typing of the gloss or the rough equivalent lemma in the case of bilingual dictionaries. Typical examples are dictionary applications like *The Online Dictionary of New Zealand Sign Language* (http://nzsl.vuw.ac.nz/) and the *English to ASL Dictionary* (http://www.lifeprint.com/dictionary.htm), and also.
Rather than using video, some systems adopt avatar technology. The terminology dictionaries of ASL such as *The Signing Math Dictionary of ASL* (http://www.vcom3d.com/index.php?id=aslani), which are developed by Vcom3D in order to be used in Deaf education, use advanced avatar technology to display lemmas. Other examples of avatar use for lemma representation apply primitive signing avatar versions, such as the one used in *The National Business Aviation Association K8AIT Sign Llanguage Dictionary* (http://wings.avkids.com/Book/Signing/abc.html), which provides mainly aviation related lemmas and has no search facilities other than by groups of letters of the Latin alphabet, while it displays lemma related concept definitions in written English.
The *Sign Look-up Tool* (Figure 2) also has two components – sign input and recognition, and lexical sign data presentation – in a client-server relationship. The most important extension introduced in this system is the ability to handle multiple languages.

Figure 2: Sign Look-up Tool - Sign Matches Display Mode

*Sign Matches Display* mode allows the user to view all entries in the ranked list of signs potentially matching the latest user input, based on the list supplied by the server. *Translations Display* mode allows the user to view translations of any of the current set of matching signs in the four sign languages covered by the Dicta-Sign Lexicon. Moreover, the Dicta-Sign Lexicon contains WordNet descriptions for each concept. The user can see the WordNet entry for a concept, in the panel at the bottom right of the window, by hovering over the avatar currently displaying that concept.

Finally, the *Sign-Wiki* prototype (Figure 3) integrates all technologies and resources developed during the project.

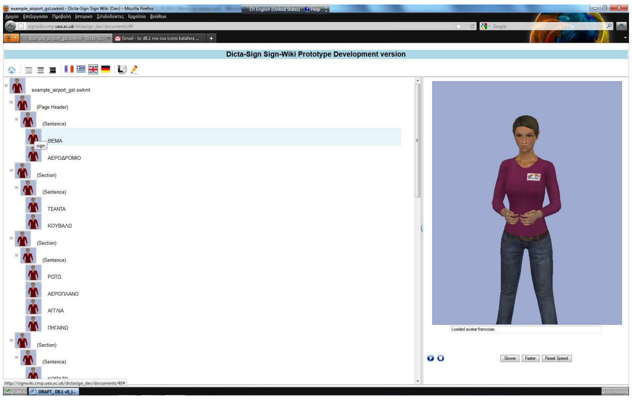
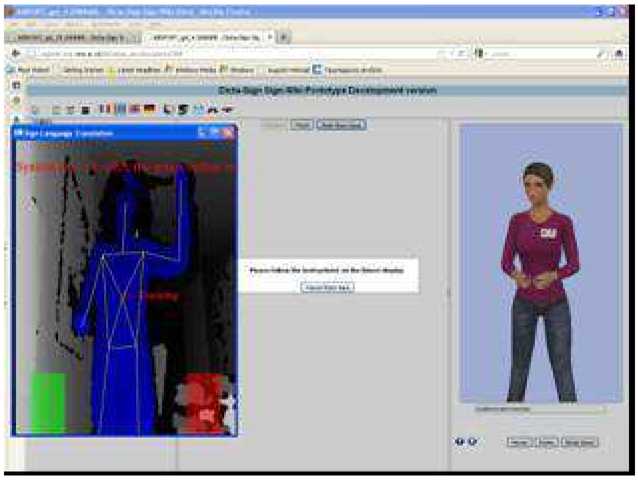



Figure 3: Sign-Wiki typical page

The Dicta-Sign sign-Wiki demonstrates the potential of sign languages to participate in contemporary Web 2.0 applications where user contributions are editable (Figures 4, 5) by an entire community and sign language users can benefit from collaborative editing facilities.
The server developed within the project provides the same service as a traditional Wiki, but using sign language. Sign language content may be directly inserted in dictation style by means of a Kinect camera (Ong 2012). Newly created or previously existing content may be edited, saved and uploaded for presentation. Instead of using text as the output medium, a signing avatar presents information. The system matches the user's signs against a stored dictionary, and the matched signs are used to generate the movements of the signing avatar. The use of an avatar preserves the anonymity of the user and facilitates modification and reuse of information present on the site. If a Kinect device is not available, sequences can be created by selecting signs from the system's lexicon, using spoken language synonyms to expand the range of choices.



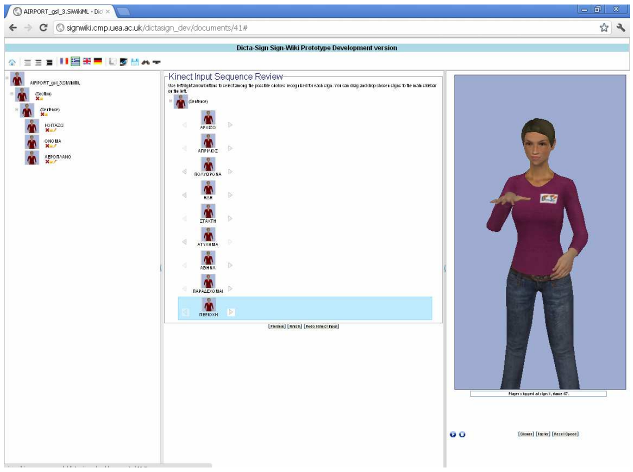




Figure 4: Sign-Wiki - Input reviewing

Innovation in respect to communication via sign language is summarised in the following:

I) Sign-Wiki users may view SL information uploaded by other individuals. This may involve information in one's own sign language or may require translation support in order to be comprehended.

II) In the latter case, the user may find support by a Sign-Look-up translation module, which allows search of signs in four sign languages. Multilingual correspondences of the same concept increase the possibility of its understanding.

III) The user may edit previously uploaded signs or sign phrases by applying i.e. simple copy-paste procedures on pieces of SL utterances or by changing basic components of a sign, using a visual sign editor or the set of HamNoSys notations.

IV) The user may create new SL content by either entering his/her own productions to the system by means of a Kinect devise by exploiting the project's sign language recognition technologies and/or using the sign creation tools and linguistic models also used for editing purposes. In the case of real time input from the part of the user, single signs or sign phrases are performed in dictation style.

V) The user may save, upload and present his/her content preserving his/her anonymity, since performance of sign language content happens by means of a signing avatar exploiting sign animation technologies.

Detailed reporting on end-user evaluation of the DICTA-SIGN Sign-Wiki, is the subject of project deliverable D8.2: Evaluation report of Sign-Wiki demonstrator.


Figure 5: Sign-Wiki – Sign Builder Tool

Finally, Dicta-Sign prototypes have been exposed to end-user evaluation procedures that have provided comments relating to all levels of implementation, crucially emphasising on the Deaf user's preferences in respect to interaction with the systems, thus, gaining advanced human-computer interface design for Web 2.0 sign language applications, that can be best viewed in implementation of the sign-Wiki prototype, also serving as the project demonstrator.
Especially in respect to the sign-Wiki, since the prototype is usable online, all functions were tested via internet by end-users using one of the four project sign-languages (LSF, GSL, DGS, GSL) thanks to the translation option. Gained results revealed that the wiki is actually used equally in order to create new utterances and to modify existing utterances. While it would also be possible to use the Wiki interface key concepts in pedagogical applications or for information providing purpose in combination with other existing solutions like 3DSigner (www.3DSigner.fr), besides possible applications, the testers pointed out provided anonymity as the major strength of such an application.

## 4. Conclusion

Dicta-Sign has undertaken fundamental research and development in the combined use of image processing and advanced computer vision techniques, statistical methods for continuous sign recognition with multimodal fusion and adaptation, virtual human technology, sign language modelling, grammar & lexicon design and development as well as corpus construction. The Dicta-Sign demonstrator focused on the end user's requirements as regards human-computer interaction via sign language. Under this light, the main aim here has been to underline the range of actions and interaction possibilities that are finally offered to signing users of Web 2.0, resulting from research work that exploits properly annotated language resources.

## 5. Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 231135.

## 6. References

Antonakos, E., V. Pitsikalis, I. Rodomagoulakis and P. Maragos (2012). "Unsupervised Classification of Extreme Facial Events using Active Appearance Models Tracking for Sign Language Videos". submitted for publication in: *Proc. Int'l Conf. on Image Processing (ICIP-2012)*, Orlando, Florida, USA .

Athitsos, V., Neidle, C., Sclaroff, S., Nash, J., Stefan, A., Thangali, A., Wang, H., and Yuan, Q. (2010). "Large Lexicon Project: American Sign Language Video Corpus and Sign Language Indexing/Retrieval Algorithms". *Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT)*, pages 11-14.

Collet, C., M. Gonzalez and F. Milachon (2010). "Distributed system architecture for assisted annotation of video corpora". *Proc. of the LREC Workshop on the Representation and Processing of Sign Languages : Corpora and Sign Languages Technologies*, May 17th–23rd 2010, Valetta, Malta., pp. 49-52.

Cooper H, Pugeault N, Bowden R (2011). "Reading the Signs: A Video Based Sign Dictionary". In *Proc IEEE Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Streams (ARTEMIS)*, Proc ICCV11.

Cooper, H., N. Pugeault, E.-J. Ong and R. Bowden (2012). "Sign Language Recognition using Sub-Units." *Journal of Machine Learning and Recognition: Gesture Recognition*. (to appear).

Dubot, R., Collet, C. (2012). "Improvements of the Distributed architecture for assisted annotation of video corpora". *Proc. of the LREC Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon*. May 21-27 2012, Istanbul, Turkey.

Efthimiou, E., S.-E. Fotinea, T. Hanke, J. Glauert, R. Bowden, A. Braffort, C. Collet, J. Maragos and F. Goudenove (2010). "Dicta-Sign: Sign Language Recognition, Generation, and Modelling: A Research

Effort with Applications in Deaf Communication". *Proceedings of the Language Resources and Evaluation Conference Workshop on the Representation and Processing of Sign Languages : Corpora and Sign Languages Technologies*, May 17th–23rd 2010, Valetta, Malta.

Elliott R, Cooper H M, Ong E J, Glauert J, Bowden R, Lefebvre-Albaret F (2011), Search-By-Example in Multilingual Sign Language Databases, In *Proc. Sign Language Translation and Avatar Technologies Workshops*, 2011.

Filhol, M. (2012, tbp). "Combining two synchronisation methods in a linguistic model to describe Sign Language" In E. Efthimiou, G. Kouroupetroglou, S.-E. Fotinea (Eds.), *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*, Springer LNCS/LNAI 7206.

Filhol, M. & Braffort, A. (2012, in these proc.). "A Study on Qualification/Naming Structures in Sign Languages". *Proceedings of the Language Resources and Evaluation Conference Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon*, May 27th 2012, Istanbul, Turkey.

Glauert J & Elliott R (2011). "Extending the SiGML Notation - a Progress Report" Proc. *Sign Language Translation and Avatar Technologies*, Dundee, October 2011.

Gonzalez, M. and C. Collet (2011). "Robust body parts tracking using particle filter and dynamic template" *18th IEEE International Conference on Image Processing (ICIP 2011),* 11-14 Septembre 2011, Brussels, Belgium. pp 537-540.

Gonzalez, M. and C. Collet (2012). "Signs segmentation using dynamics and hand configuration for semi-automatic annotation of sign language corpora". *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication, GW 2011*. E. Efthimiou, G. Kouroupetroglou and S.-E. Fotinea. Lecture Notes in Computer Science/Lecture Notes in Artificial Intelligence Vol. 7206: Springer-Verlag (to appear).

Hadfield, S. and R. Bowden (2012). "Go With The Flow: Hand Trajectories in 3D via Clustered Scene Flow". Proc. *International Conference on Image Analysis and Recognition (ICIAR)*, June 25-27, 2012, Aveiro, Portugal (Accepted to Appear).

Hadfield S, Bowden R (2011). "Kinecting the Dots: Particle Based Scene Flow From Depth Sensors". In Proc. *Int. Conference Computer Vision, ICCV2011*, Barcelona, Nov 2011.

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*. ELRA, Paris, pp. 64-67.

Hanke T, Matthes S, Regen A, Storz J, Worseck S, Elliott R, Glauert J, Kennaway R. (2011). "Using Timing Information to Improve the Performance of Avatars" in Proc *Sign Language Translation and Avatar Technologies*, Dundee, October 2011

Holt, B., E.-J. Ong, H. Cooper and R. Bowden (2011). "Putting the pieces together: Connected Poselets for Human Pose Estimation". In Proceedings of the *13th Int. Conference on Computer Vision,: Workshop on Consumer Depth Cameras for Computer Vision*, November 6-11, 2011, Barcelona, Spain.

Jennings, V., Elliott, R., Kennaway, R.and Glauert. J. (2010). "Requirements for a Signing Avatar". Proceedings *Workshop on Corpora and Sign Language Technologies (CSLT)*, LREC 2010, Malta, pp133-136, May 2010.

Maragos, P., Pitsikalis, V., Theodorakis, S., Rodomagoulakis, I., and Antonakos, E. (2012). Report on Integrated Continuous Sign Recognition. DICTA-SIGN Tech. Rep. D2.3, NTUA, Greece, Feb.2012.

Matthes, S., Hanke, T., Storz, J., Efthimiou, E., Dimiou, N., Karioris, P., Braffort, A., Choisier, A., Pelhate, J. and Safar, E. (2010). "Elicitation Tasks and Materials designed for Dicta-Sign's Multi-lingual Corpus". In Dreuw, P., Efthimiou, E., Hanke, T., Johnston, T., Martínez Ruiz, G. and Schembri, A., (eds.), LREC-2010, *Proceedings of 4th Workshop on Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pp. 158-163.

Ong, E.-J. and R. Bowden (2011). "Robust Facial Feature Tracking using Shape-Constrained Multi-Resolution Selected Linear Predictors". *IEEE Trans Pattern Analysis and Machine Learning* 33(9): pp1844-1859.

Ong. E.-J. and Cooper, H. M. and Pugeault, N., and Bowden, R., (2012). "Sign Language Recognition using Sequential Pattern Trees". In Proc. *Computer Vision and Pattern Recognition 2012*. To appear.

Pitsikalis, V., Theodorakis, S., Maragos, P. (2010). "Data-driven sub-units and modeling structure for continuous sign language recognition with multiple cues". In: Proc.*LREC Workshop: Corpora and Sign Language Technologies*.

Pitsikalis, V., Theodorakis, S., Vogler, C. and Maragos, P. (2011). "Advances in Phonetics-based Sub-Unit Modeling for Transcription Alignment and Sign Language Recognition". *IEEE CVPR Workshop on Gesture Recognition*, Colorado Springs, USA, 2011.

Pugeault N, Bowden R, (2011). "Spelling It Out: Real-Time ASL Fingerspelling Recognition", In *IEEE Workshop on Consumer Depth Cameras for Computer Vision, Proc ICCV 2011*.

Rodomagoulakis, I., S. Theodorakis, V. Pitsikalis and P. Maragos (2011). "Experiments on Global and Local Active Appearance Models for Analysis of Sign Language Facial Expressions". Proc. *9th International Gesture Workshop (GW 2011): Gesture in Embodied Communication and Human-Computer Interaction*, May 2011, Athens, Greece.

Roussos, A., S. Theodorakis, V. Pitsikalis and P. Maragos (2010). “Affine-Invariant Modeling of Shape-Appearance Images applied on Sign Language Handshape Classification”. Proc. *International Conference on Image Processing (ICIP-2010)*, Sep. 2010, Hong Kong.

Roussos, A., S. Theodorakis, V. Pitsikalis and P. Maragos (2010). “Hand Tracking and Affine Shape-Appearance Handshape Sub-Units in Continuous Sign Language Recognition”. Proc. *11th European Conference on Computer Vision (ECCV): Workshop on Sign, Gesture and Activity (SGA)*, Sep. 2010, Heraklion, Crete, Greece.

Theodorakis, S., Pitsikalis, V. and Maragos, P. (2011a). “Dynamic vs. Static Sub- Unit Modelling for Continuous Sign Language Recognition with Movement- Position Cues”. *IEEE Trans. On Systems Man and Cybernetics,* 2011 (under review).

Theodorakis, S., Pitsikalis, V. and Maragos, P. (2011b). “Advances in Dynamic-Static Integration of Manual Cues for Sign Language Recognition”. *Proc. 9th International Gesture Workshop (GW 2011): Gesture in Embodied Communication and Human-Computer Interaction*, May 2011, Athens, Greece.

S. Theodorakis, V. Pitsikalis, I. Rodomagoulakis, and P. Maragos (2012). “Recognition with Raw Canonical Phonetic Movement and Handshape Subunits on Videos of Continuous Sign Language”, submitted for publication in: Proc. *Int’l Conf. on Image Processing (ICIP-2012)*, Orlando, Florida, USA, Sep. 2012.

Vogler C, (2010) Extraction of segmental phonetic structures from HamNoSys annotations. DICTA-SIGN Tech. Rep. D4.2, ILSP, Greece.

Wang, H., Stefan, A., Moradi, S., Athitsos, V., Neidle, C. and Kamangar, K. (2010). ”A System for Large Vocabulary Sign Search”. Proc. *ECCV-2010 Workshop on Sign, Gesture and Activity (SGA)*, Sep. 2010, Heraklion, Crete.

Sign language Dictionaries:

The Online Dictionary of New Zealand Sign Language: http://nzsl.vuw.ac.nz/

English to ASL Dictionary http://www.lifeprint.com/dictionary.htm

The National Business Aviation Association K8AIT sign language dictionary http://wings.avkids.com/Book/Signing/abc.html

# SignWiki – an Experiment in Creating a User-based Corpus

**Sonja Erlenkamp[1], Olle Eriksen[2]**

[1]University-College of Sør-Trøndelag
Department of Teacher and Interpreter Education
7004 Trondheim, Norway

[2]Møller Resource Center
PO box 175 Heimdal
7473 Trondheim, Norway
E-mail: sonja.erlenkamp@hist.no, olle.eriksen@statped.no

**Abstract**

In comparison to other signed languages, Norwegian Sign Language (NTS) is not well researched and documented while at the same time the need for documentation of NTS in a corpus based dictionary has been apparent to the field for quite a while. Despite some high quality applications to raise funding for corpus work, the field in Norway has not succeeded to gain enough understanding in governmental research funding institutions for the need of a corpus based dictionary, mainly because of the rather small population of NTS users. As a result, a new approach is used by involving the NTS community to create a database of signs, including their use, distribution and as far as possible other metadata. Tegnwiki (=Signwiki) is a first attempt at creating a user-based database of NTS by allowing users to contribute videos and information on isolated signs on a Wiki platform. Like Wikipedia, the Signwiki will be open accessible, but administered by a group of experts. Obviously a Signwiki cannot replace a scientific corpus. But if this experiment is successful it might be a good starting point for countries with no or little funding for corpus projects where involvement of users is the key factor.

**Keywords:** Norwegian Sign Language, Wiki, Sign Database

## 1. Introduction

Norwegian Sign Language (henceforth NTS for Norsk TegnSpråk) is one of the known yet little described signed languages in Europe. It is one of the genuine Norwegian languages used by a language minority of several thousand deaf and hearing users. Since 1825 it has been school language in Norway at schools for the deaf. As early as 1875 Norwegian Sign Language was labelled as a language (Skavlan, 1875), but as in many other western countries this attitude towards a signed language did not survive the period of oralism first in the late 1970s and early 1980s, the idea of NTS as a natural full-fledged grammatical language evolved again. Through the past 3 decades, several official documents and articles (e.g. Bergh, 2004; Erlenkamp, 2007; Erlenkamp et al. 2007) have operated with a number of about 4000 to 5000 deaf Norwegians and an unknown number of hearing Norwegians using NTS as one of their first languages. It is estimated that about 15.000 of the 5 million Norwegians use this language as a first or second language.

Thus, the language community is rather small. By now, NTS has gained a relatively wide acceptance in the Norwegian Society; on April 28th 2009, a proposition was passed by the Norwegian Parliament that NTS should become one of several official languages (GP35 2008).

Sign language studies and interpreting studies have been offered at several Universities and University Colleges since the mid-1990s. Moreover, in the 1990s the government established a 40 weeks free course in NTS for hearing parents of deaf children to help closing the gap between the hearing parent's signed language knowledge and skills and the practical skills of their deaf children in NTS.

NTS is however, in comparison to other signed languages, not well researched and documented. Basic aspects about NTS grammar, acquisition and sign variation (e.g. dialects, sociolects) have not been documented yet, and the documentation of Norwegian Sign Language has so far only been conducted by a handful of researchers (e.g. Greftegreff, 1991; Vogt-Svendsen, 1990, 2001; Selvik, 2006; Erlenkamp, 2009, 2011 a and b). Furthermore, reliability of many research projects depend on the availability of large amounts of annotated language data. Thus, the need for documentation of NTS in a corpus based dictionary has been apparent to the field for quite a while and for that reason scientific groups in a number of European countries are currently developing signed language data corpus. In Norway the establishment of any collection of mentionable size has been prevented by the lack of a standard written representation which results in an extremely time consuming process where the visual data must be annotated manually. Despite some high quality applications to raise funding for corpus work, the field has not succeeded to gain enough understanding in governmental research funding institutions for the need of a language corpus and a corpus based dictionary, mainly because of the rather small population of NTS users. The existing sign language dictionary project "Norsk tegnordbok" developed by the Møller Resource Centre has due to economic limitations and limited availability of expertise reached a state where single signs are searchable online, but crucial information on linguistic categories of the signs, their usage and affiliation to regional or social varieties is not available. Thus the glossary consists of videos of simple isolated signs (in total about 6500) and corresponding Norwegian translations (http:tegnordbok.no). The translation lemmata are the basis for a search in this glossary.

As a result, the field is trying out a new approach by involving the NTS community to create a larger database of signs, including their use, distribution and as far as possible other metadata. Tegnwiki (=Signwiki) is a first attempt at creating a user-based database of NTS. If successful, this project could contribute in a large scale to increased accessibility of data on isolated signs in NTS. In that case, Tegnwiki will be functioning as an aid for interpreters in need of vocabulary, as well as part of signed language learning. Furthermore, users can suggest and discuss new signs on the Signwiki and thus the Signwiki will open up for a more democratic process on the development of new signs in underdeveloped domains of NTS.

## 2. Project Goals

The project has the following goals:

1) developing a user interface based on a common wiki user interface, modified to allow easy integration of videos even for users with little experience in using video tools
2) developing a standard for each article (each article will be linked to one sign), including slots for metadata about the usage of the sign, as well as opportunities for user discussions about the sign
3) informing and encouraging the NTS community to participate in this project

Above all, this project is an attempt to involve signers in a project about their own language and gather some information of signs based on the user's knowledge. As a consequence, the expectations on what can be collected and the level of quality of each article have to be kept at a reasonable level. It is, however, planned to make the signs from the already existing glossary available on the wiki as well, in an attempt to obtain more information about these signs and thus hopefully to create a synergy effect between the Signwiki and the dictionary project.

## 3. Technological, Economical and Scientific Requirements

Research on any sign language during the past decades has been shaped by two main limitations due to the visual modality the language data are based on. Both the lack of a standard written representation of signed languages which also could serve as a base for an annotation system, and the technological limitations regarding storage and access of large amounts of signed language data, have until recently made it impossible to work on large amounts of sign language data. Technological developments have already improved the latter, through streaming and other methods of accessing films. The former limitation is currently undergoing a major change due to the development of software tools for annotations of visual data like syncWRITER (Hanke & Prillwitz 1995), ELAN and iLex (Hanke & Storz, 2008).

These developments do, however, not change the fact that the process of developing, annotating and storing a large signed language data corpus depends on several external factors:

1. availability of staff trained in signed language and annotation of visual data (or the opportunity to train the staff)
2. access to research based knowledge about signed linguistic categories for the respective signed language used as background for tagging
3. access to (and funding of) the necessary technical equipment, e.g. internet access, servers, cameras
4. funding of the time consuming process of creating a data base

Any signed language corpus project is highly dependent on all of these four factors. Thus, so far only countries with enough expertise, financial background and technical infrastructure have implemented such a project. Countries that for whatever reasons do not manage to meet these factors will have to develop other strategies for building accessible information about signs/signed language. As Norway definitely has the sufficient means for every factor, willingness of prioritizing this kind of project in research politics is a crucial part of the funding process.

In distinction to large data corpus project, a Signwiki has to deal with only two of the four factors:

1. access to research based knowledge about signed linguistic categories
2. access to (and funding of) the necessary technical equipment, e.g. internet access, servers, cameras

The first requirement will be met with a small group of experts to develop a standard for linguistic data and eventually work with securing the quality of information on the Wiki.

The second factor is not an issue in Norway, since Norway has one of the highest internet access per capita rates in the world. Furthermore, other needed equipment like web cameras are affordable for most people in Norway and many deaf and hearing Norwegians already possess one.

In order to achieve the highest possible accessrate for users in the Wiki, the project aims nevertheless towards mainstream technical solutions, i.e. standard hardware, in spite of possible decrease in video-quality. The user interface depends heavily

on the use case and must be easily accessible with a low learning demand. Thus, the SignWiki aims at using a common Wiki user interface, with integrated functions for use of videos in order to record and play videos.
The user does not have to store data on her/his own computer. All data are manipulated on the central server. The user needs an internet connection with "Fast broadband connection".
Ideal for video recording would be at least 25 frames per second and a standard PAL resolution (720x576), which however cannot be expected from most home-used webcameras. Thus, we can distinguish between contributions from the project group (with data from the sign dictionary) and contributions from the non-expert users. The former group will deliver data meeting the requirements for a data corpus, while the latter might not. This may result in a number of videos less clear, sharp and smooth than sought for in other signed language data corpus projects and will in consequence lead to less adoptability of the Signwiki's data to other projects.

## 4. Availability of Data and Privacy

Like Wikipedia, the Signwiki will be open accessible, but administered by a group of experts. Access to the infrastructure will be through an internet portal. Most data will be publicly accessible, while there might be some data, like experts discussions, that are only accessible to the project group, in particular throughout the establishment of the project.
Norwegian rules about privacy in connection with research projects are very strict and would limit access to a research data base: Access to the raw data would be limited to researchers and application designers; before granted access, users would need to sign agreements about confidential use of the data. By developing an open Signwiki where every user can control her/his own data input, privacy will not be an issue.

## 5. Linguistic Data and Metadata

Each article in the Wiki will cover one single sign with the opportunity to comment on and contribute to meta-information about the sign. This form of data collection leads to some advantages in terms of being user based, but also comprises a number of risks regarding the reliability and quality of the data.
Research on NTS has been performed by only a few researchers since the early 1980s and has thus focused mainly on certain areas of grammatical description, like sign classes (Erlenkamp, 2000), different parameters of signs like mouthing (Vogt-Svendsen, 2001) and hand shapes (Greftegreff, 1991), time expressions (Selvik, 2006) and sentence types (Vogt-Svendsen, 1990). At present, a language model for NTS grammar and iconicity is under development, including grammatical descriptions like word order and grammatical relations (Erlenkamp, 2009; 2011 a and b). Since the number of researchers on NTS is small, most projects on NTS follow international developments on theory and methods and contribute to these developments. The results of earlier research will serve as background for the development of a list of desired metadata about a sign. This list will be matched up against a realistic expectation of what native signers without expertise in signed language linguistics can contribute with, e.g.:

| **List of desirable linguistic data:** | **Realistic expectations towards user provided data:** |
|---|---|
| Phonological data: | Most of these data can be obtained at any time by analyzing the video data by experts, independent from the Wiki users) |
| Morphological/syntactic categories: | Unlikely to be provided by users |
| Data about the distribution of the sign: | Likely to be able to get some indication from users |
| Examples of usage: | Very likely to be able to get reliable data from the users |

One of the major questions is how the data on the Signwiki will be searchable. One obvious solution is a search function based on the Norwegian translation word(s) for each sign. In addition we are going to look into possible solutions for a search based on sign configurations. For that purpose, a Ph.D.-candidate will try to develop and apply her model for searching on the Wiki as part of her project.
Since the Signwiki-project is not meant as a scientific database, metadata common in scientific signed language corpora cannot be expected to be obtained through Signwiki. As part of a Wiki, some metadata are collected automatically, like when the sign was put online and by whom. This kind of metadata, however, are not the most interesting for signed language data comparison.
Obviously a Signwiki cannot replace a scientific corpus. But if this experiment is successful it might be a good starting point for countries with no or little funding for corpus projects were the involvement of users is the key factor.
Technical solutions for video presentation of signed language data might also be of interest for other publishers of websites on NTS. In a best case scenario, Signwiki will serve as a contribution for a sign language dictionary platform and as a supplier of examples as well as a democratic platform for the development of new signs in NTS.

## 6. Acknowledgements

This project has been financially supported by the Norwegian ExtraFoundation for Health and Rehabilitation through EXTRA funds.

## 7. References


Bergh, G. (2004) (ed.). Norsk tegnspråk som offisielt språk. [Norwegian Sign Language as an Official Language]. ABM. ABM-skrift nr. 10. Report.

ELAN. EUDICO Linguistic Annotator. http://www.lat-mpi.eu/tools/elan/

Erlenkamp, S. (2007). Et tospråklig liv med norsk tegnspråk. Språknytt 4:2007, pp. 16-19.

Erlenkamp, S., Gjøen, S., Haualand, H., Kvitvær, H.B.; Petterson P.R.; Schrøder, O.-I. & Vonen, A.M. (2007). Begrunnelser for å gjøre norsk tegnspråk til offisielt språk. Notat frå utval oppretta av styret for Norges Døveforbund. [Arguments for why Norwegian Sign Language Should Become Official Language. Report of the Committee Established by the Norwegian Deaf Association]. Oslo: Norges Døveforbund.

Erlenkamp, S. (2009). "Gesture Verbs" - Cognitive-visual Mechanisms of "Classifier Verbs" in Norwegian Sign Language. CogniTextes 3. http://cognitextes.revues.org/250

Erlenkamp, S. (2011). Grunntegnstilling i norsk tegnspråk. Erlenkamp, S.; Amundsen, G.; Halvorsen, R. P. & Raanes E. (Eds.), Norsk Lingvistisk Tidskrift 29(1), Oslo: Novus, pp. 87-116.

Erlenkamp, S. (2011). Norsk tegnspråk: helt norsk og veldig annerledes. Skisse av en ny beskrivelsesmodell for norsk tegnspråk. Erlenkamp, S.; Amundsen, G.; Halvorsen, R. P. & Raanes E. (Eds.), Norsk Lingvistisk Tidskrift 29(1), Oslo: Novus, pp. 26-37.

GP35 (2008). Governmentproposition no. 35 (2007-2008) "Mål og meining. Ein heilskapleg norsk språkpolitikk". http://www.regjeringen.no/pages/2090873/PDFS/STM200720080035000DDDPDFS.pdf

Greftegreff, I. (1991). Håndformer og håndformendringer i norsk tegnspråk: En innledende undersøkelse av foneminventaret. M.A.-thesis in linguistics, Trondheim: University of Trondheim.

Hanke, T. & Prillwitz, S. (1995). syncWRITER: Integrating Video into the Transcription and Analysis of Sign Language. In Bos, H. & Schermer, G. (Eds.), Sign Language Research 1994: Proceedings of the Fourth European Congress on Sign Language Research, Munich, September 1-3, 1994. Hamburg: Signum, pp. 303-311. http://www.signum-verlag.de/BTitel/pdf/3-927731-57-9b.pdf.

Hanke, T. & Storz. J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguisticsand Sign Language Lexicography. Proceedings of 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora at LREC 2008.

Selvik, K.-A. (2006). Spatial Paths Representing Time. A Cognitive Analysis of Temporal Expressions in Norwegian Sign Language; Oslo: Department of Linguistic and Nordic Studies, University of Oslo. Ph.D.-thesis.

Skavlan, S. (1875). Throndhjems Døvstumme-Institut Program, udgivet i Anledning af Institutets 50-aarige Bestaaen. Trondheim: Statped (original: Lie & Sundts Bogtrykkeri).

Vogt-Svendsen, M. (1990). Interrogative strukturer i norsk tegnspråk. En analyse av nonmanuelle komponenter i 86 spørsmål. Trondheim: Faculty of History and Philosophy, University of Trondheim. Ph.D. thesis.

Vogt-Svendsen, M. (2001). A Comparison of Mouth Gestures and Mouthings in Norwegian Sign Language (NTS). In Boyes Braem, P. & Sutton-Spence, R. (Eds.), The Hands are the Head of the Mouth: The Mouth as Articulator in Sign Language. Hamburg: Signum, pp. 9-39.

# Transcribing and evaluating language skills of deaf children in a multimodal and bilingual way: the sensitive issue of the gesture/signs dynamics

**Isabelle Estève**

Laboratoire Lidilem, Université de Grenoble
Université Stendhal UFR des Sciences du Langage BP 25 38040 Grenoble cedex 9
isabelle.esteve@u-grenoble3.fr

**Abstract**

Transcribing and evaluating the narrative productions of 6 to 12 year-olds deaf children in their multimodal and bilingual dimensions confront us to the central question of gestures/signs distinction. This paper aims to discuss how the narrative skills of 30 deaf children schooled in different education settings – oralist, bilingual and "mixed" – led us to create transcription/annotation tools in ELAN allowing to take into account the dynamics between verbal and non-verbal material involving especially within the gestural modality. We will focus on two central points of our reflections. How to delimit productions in units into taking into account the semiotic and the structural dynamics aspects of production? How to describe and categorize the gestural processes non systematized in a linguistic form to report the developmental dynamics?

**Keywords:** multimodality, deafness, gestures/signs dynamics, transitional skills

## 1. Introduction

Our contribution proposes to envisage the thematic of the interplay between sign language corpora and lexicons in a particular way. We emphasize the issues raised by the evaluation of the skills of deaf children schooled at primary school. The aim of this paper is to approach the sensitive issue of the gestures/ signs distinction in deaf children's productions whose language skills are still in linguistic development. The reflections that we have propose to expose here concerning the transcription/annotation tools are based on corpora of language productions, lexical and narratives, of 30 deaf children aged 6-12years and schooled in different education setting – oralist, bilingual and "mixed". Given the diversity of the deaf children profiles in terms of familial environment, educational background, degree of deafness etc., our corpora constitute a representative sample of the reality of the schooling spaces in the context of deafness. Our approach of the deaf child orality[1] development is thus fundamentally empiric, anchored in actual data.

We will first anchor the theoretical context of our research in the light of the hypotheses on the multimodal aspect of language in the hearing context. Then, we will focus on two central points of transcription/annotation schemas which we have built in ELAN. Firstly, we have will argue how specific gestural dynamics, and in a more broader perspective bimodal dynamics, in deaf children's production incite to shift the focus away from the linguistic production to conceive an integrative approach to the interplay of the verbal and non-verbal material. More specifically, we have narrowed down the idea of global units of segmentation that we have propose for a transcription/annotation grid. We – Agnès Millet and I – have worked in ELAN to transcribe deaf children productions in the multimodal and bilingual aspects. Secondly, we will focus more specifically on the sensitive issues of the criteria used to distinguish gestures and signs. These issues led us to develop tools which allow to catch, in a more dynamic way, the state of the development of the deaf children's gestural symbolizations skills as well as their evolution, regardless whether these skills are systematized in a linguistic form or not. We will conclude by exposing the limits of our transcription / annotation schemas, in their current state of elaboration. And finally we will open the prospects which are still to be explored in order to provide answers to description challenges emerging from the specific shapes of the language dynamics involved in the orality development in the context of deafness.

## 2. Theoretical context of the research

### 2.1. Crucial issues of a multimodal approach to the orality in the deafness context

The multimodal aspect of the language has been less explored in the context of deafness and, in particular, in later language development. However, given the diversity of the sociolinguistics contexts in which the deaf children make their first steps in language, the linguistic skills of deaf children in primary school are still under development and especially in sign language (SL) – 90% of deaf children are born into hearing families and thus, for the most of them, are not exposed to a SL model before beginning school. Although this observation is largely shared by many researchers, locally and internationally, current research focus on the linguistic level of the competence, and even, for the most part, on a single linguistic component of the skills developed by the deaf child : exclusively in SL (Schick et al., 2005) or exclusively in Vocal Language (VL) (Spencer

[1]Given the ambiguity how the term "oral" is used in the context of deafness which is always used in a interchangeable way or in the place of the term "vocal" to refer to speech and to avoid therefore a misunderstandings of the concept of *orality*, we have to precise at the outset that this concept is used here to an opposition with *scripturality*. As my approach takes a multimodal perspective, *orality* has to be conceived here in its broader sense as to include all vocal and gestural resources, in their verbal and non-verbal dimensions.

and Marschark, 2006). Only a few recent studies are take into account the sociolinguistics realities in which the development of deaf child orality is anchored in approaching the question of the interplay of the two oral components of linguistic skills – i.e. in SL and VL – (Plaza-Pust and Morales-Lopez, 2008). Indeed, the deaf child is led, through the diversity of his daily interactional contexts which he has to face, towards building in bilinguality [2]. Although these studies integrate bilingual dynamics of the skills of deaf children, most of these studies focus on a linguistic perspective of description/evaluation too. However, the point is, as for us, to capture the set of language dynamics which are implied in symbolization development skills of deaf children, as broadly as possible, and then no matter whether and how far these skills are systematized in a linguistic form or not. Therefore, in this reasoning, multimodality offers an unique perspective to understanding the transitional states of the development of deaf children's symbolic abilities, and especially in the gestural modality.

### 2.2. Cross-boundaries of bilingualism and multimodality perspectives: An integrative approach

Then, on the margins of the majority of researchs in the context of deafness, our approach of the orality is anchored in an effective application of the concept of *communicative competence* proposed by Hymes (1984), integrating a broad conception of language – *i.e.* including these verbal and non-verbal dimensions – such as elaborated by McNeill (1992). The orality in the context of deafness and its development offers the opportunities to dialogue perspectives of the research fields of bilingualism and multimodality : deaf bimodality is indeed implied specifically in a both verbal or non-verbal potential. The deaf children's orality development thus offers an unique window on the multimodal aspect of language : gestuality and vocality are both in tense to a linguistic potential, respectively up to a SL (LSF, in our context of research) and up to a VL (French, in our context). So, cross multimodality and bilingualism constitutes an integrated perspective to fully explore the inter- and intra-modality dynamics, in their whole, at work in the development of the *integrated* (McNeill, 1992) bimodal language system, bilingual in progress. The transposition of the Kita's information packaging hypothesis (Kita, 2000) allows us to highlight the perspectives opened by the language development in the context of deafness.

### 2.3. Proposal to modelling deaf orality development

The key idea of the cognitive hypothesis proposed by Kita is that the hearing speaker has at his disposal in bimodality two alternative or concurrent manners of organizing the representation of the events supported by two modes of thinking. One is linked to the manner how to perceive events in the concrete world, which is underlain by *spatio-motric thinking*, and, the other, conforms to the way of organizing the information in a decontextualized and hierarchically structuring form in a particular language, which is underlain by the *analytic thinking*. The application of this hypothesis to evaluate the symbolization skills under linguistic development opens perspectives, particularly relevant for the application to deaf children. Indeed, Kita arguments that the concurrence/complementarity between these two modes of thinking is revealed on-line in the production of hearing children and in particular in the gestures-speech discordant combinations. These types of bimodal combinations can be interpreted as a symbolization acquired in the spatio-motric thinking but not yet systematized in a linguistic form. So, bimodal combinations give a direct insight on the transitional phase of two modes of thinking and allows to investigate, in a more sensitive manner, the question of phases of language acquisition. Note that if in hearing children, the transition takes place between gestures and speech, concerning deaf children, a double transition is implied concurrently in both modalities : from gestures to signs in one part, and from onomatopoeias [3] to words in the other. So, on the basis of Kita's hypothesis, our proposition of modelling language development in deaf children can be represented as follows :

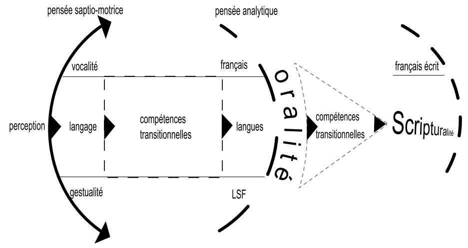



Figure 1: Proposal to modelling the deaf orality development

Indeed, in deaf children development, gestuality and vocality are both anchored in the movements of a symbolization i) firstly anchored in the sensorimotor perception : visual for the one, auditory for the other – most of deaf children have actually hearing aids or have a cochlear implant – ii) and then gradually moves away from this perceptive anchors to integrate a decontextualized and hierarchically structuring way to encoding the information in a particular language, in a SL for one and in a VL for the other. We know that, given the deafness, primary symbolic abilities of deaf children are developed in the gestural modality no matter of the presence of a SL in their familial environment – see Da Cunha Pereira and De Lemos (1994), Goldin-Meadow (2005), Mohay (1994), Van den Bogaerde (2000). Given the importance of the contact with the VL, this transition is initiated concurrently, nonetheless

[2] No matter of the education setting – i.e. oralist or bilingual for example – in which they are involved in their childhood and no matter otherwise of the type of their family environment (deaf or hearing) in which they have grown up, deaf adults are used to use and combine the two languages – SL and VL – and the two modalities – vocal and gestural – in order to manage their daily interactions. Then deaf adults become, to varying degrees, bilingual bimodal (SL/VL) speaker – our observations (Millet et al., 2008) are consistent in this point with other works (Lucas and Valli, 1992; Van den Bogaerde, 2000).

[3] "Onomatopoeias" are conceived here in a broader sense than usually subsumed under this term, and must be interpreted as designing all "symbolic vocalisations" which are not necessary specified in a conventional meaning as it is the case for the cock crow for example.

too, although later, in the vocal modality[4]. Nonetheless, the fact that skills of deaf children integrated progressively the manner of encoding events linguistically in a particular language will depend on the presence and on the accessibility of the linguistic models in their environment, familial initially and at school in subsequently. So, we can assume that the symbolization abilities developed in each modality which are not totally still in range in perception but which are not yet systematized in a linguistic form will be represented a more longer transitional phase, which can be pursued beyond the pre-linguistic period up to schooling and even maybe beyond this period. So therefore we propose to conceive these skills as an intermediate state between language – in its broad sense – and languageS – in its restricted sense of linguistic component – under the concept of *transitional skills*. These *transitional skills* include all the forms of symbolization developed by the deaf child which index the exploration of the potential range of organizing the events in the two channels that are at their disposal – a linear and temporal organization underlies the VL system, and a spatial organization which takes more place for simultaneity underlies the SL system – and then index a progression up to the analytic modes of thinking. The development of bimodality in deaf children thus provides an unique insight on the cognitive processes implied in the language development, in their diverse phases of transition between spatio-motric and analytic thinking which represent a constant transition in different aspects of language acquisition during childhood.
Given the theme of the workshop, this paper focusses particularly on the questions emerging from the dynamics involved in the gestural modality. Nonetheless note that bimodal dynamics in their whole, including dynamics involved on vocal modality too, open a lot of perspectives on the comprehension of the language development in the context of deafness in particular, and in a broader context, of childhood in general. These perspectives must be briefly introduced in order to replace the gestures/signs reflections in their global context.

### 2.4. Issues emerged from the evolution of gestural symbolization abilities : to welcome the plurality of deaf children's development trajectories

The heterogeneous gestural development trajectories of our corpora shake up, in different levels and by different manners, the boundaries between gestures and signs. Indeed, deaf children who have not been formally exposed to LSF develop transitional gestural symbolic skills that we cannot ignore. Note that our observations are consistent with those by other researchers : their gestural skills differ nonetheless remarkably from the gestures used by their hearing peers (see, Estève (2011),Estève and Batista (2010)) and are close in many ways to the linguistic processing existing in SL (see, Fusellier-Souza (2004),Goldin-Meadow (2005)). These trajectories of gestural development are not less legitimate than those which have developed, before their entry at school, linguistic skills in SL. So, tensions between gestures and signs observed in deaf children of our corpora have led us to re-consider the transcription/ annotation tools for gestures and signs and to re-think the *a priori* fixed categorization between verbal and non-verbal gestures – whose the apply is not without problems for the description of deaf adults discourse too. Our perspective is, beyond providing information on the heterogeneity in deaf children's orality skills, in a more long-term perspective, the elaboration of tools assessment which can be used to situate each deaf child in a specific progression between spatio-motric and analytic thinking. These tools assessment will allow us, in the longer run, to provide the reflections on adapted didactic practices which will be able to support this progression up to the development of the linguistic skills in SL – and more broadly, in our research interests, in the two languages.
In this paper, we are focussing on some central points of the challenges emerging from the deaf children's language realities of our corpora to which we are trying to answer in our grid of transcription/annotation elaborated in ELAN in a multimodal and bilingual perspective of description. We will emphasize specifically on two of the central points emerged from the description of deaf children's skills which are still in development :

- How to delimit units into taking into account the semiotic and the structural dynamics of the production ?
- How to describe and categorize the gestural processes non systematized in a linguistic form in order to report the developmental dynamics ?

These points raise a lot of other sub-questions which will be tackled in more in details in the progress of the argumentation.

## 3. Methodological aspects

The corpora on which our reflections are conducted is composed of narratives productions of 30 deaf children schooled in divers settings : oralist, bilinguals – LSF/written French *vs* LSF/oral and written French – and "mixted" – designing a class in which children with different educative projects, integrating or excluding the LSF, are taught together. The profiles of children are heterogeneous not only on the education setting, but also concerning their family environment (hearing *vs* deaf), their degree of deafness (severe to profound), their age (5 up to 12 years old), their level of schooling (GS up to CM2), etc.The narrative tasks consist on a retelling of a *Tom & Jerry*'s cartoon.

## 4. Delimiting units : towards a global approach on verbal/non-verbal phenomenona

The multidimensional aspects of the issues raised by the gestures/signs dynamics in deaf productions fundamentally challenge the existing tools used for transcribing and annotating non-verbal gestural resources. Tools usually adopted to describe the gestural productions of hearing children cannot be applied as such to report deaf children's gestural production. Rather we have to transpose these by rethinking them, in a more dynamic way, in relation with those used to the linguistics gestures. Integrated in an international

[4] The delay of VL skills development in deaf children has been highlighted by a lot of works : see, amongst others, the works of Lederberg (for a synthesis, see Lederberg and Spencer (2001)) and Pizzuto et al. (2001) on the vocabulary development.

project conducted by J.M. Colletta on the later multimodality development in children in different languages, our reflections on tools have precisely been anchored in this challenge of the necessity to adapt an existing grid, developed to transcribe and annotate the narrative conducts of hearing children – for a global presentation of this grid, see Colletta et al. (2009). The first issue raised by the specific shapes of the dynamics involved in the gestural modality in the context of deafness is to rethink the separately approaching of verbal/non-verbal resources. Rather than conceived these two kinds of resources as independently, it will be necessary to capture, in a single perspective, the co-constructed production as a whole. Therefore, the first step of our reflections was to investigate the question of segmentation units.

### 4.1. State of the art in the fields of SL and multimodality

Challenges emerging from the description of SL and multimodality data are joined together in numerous transcription/annotation points. Transcribing SL and transcribing multimodality data have in common the fact that they confront researchers with the problem of taking into account of the productions of distinct articulators which can interact simultaneously to contribute to the elaboration of utterances. However, in most of the studies in these two fields of research, even if the transcription adopted is multi-linear, these lines does not seem to be hierarchically organized and take place more as an exhaustive list integrating a transcription's line for each resource or each articulator involved in the expression. In hearing multimodality research, the delimitation units of gestures and speech are usually considered independently and their transcription/annotation is generally carried out in two separate blocks of lines. This common practice can be explained by the fact that the perspective of description is still linguistic-centered. Indeed, in most multimodal transcription systems, the central line for transcribing/segmenting productions is usually speech and the semiotic contribution of gestures is generally reported to the content of the speech syntactic units with which they are temporally linked. This is the perspective adopted by Colletta et al. (2009) to transcribe bimodal narrative productions of hearing children. Otherwise, in the research on SL, the most common system adopted is centered, for the manual components, on lexical glosses – correspond to the lexical unit in VL which is the closest to the semantic content of the lexical signs produced. The non-manual components – head, eye gaze, facial expressions, etc. – are generally transcribed in separate lines. And, in the same way as with what is observed in systems for multimodal data, the interplay of the multi-articulators contributing to the elaboration of discourse is reported to a line conceived as central: that of manual components, in link with lexical elements. Let us underline that, in the case of the SL, the contribution of the articulators is situated more at a morpho-syntactic or morpho-lexical level than at a semiotic one. Given these shared practices of transcription in the SL and VL research fields, it is not surprising that researchers who have looked at the bimodal bilingual practices of SL/VL bilingual speakers use parallel transcriptions, splitting global production into two separate lines: VL production on one first line and the SL production on a second one [5].

To our knowledge, in all existing systems of transcription, the vocal and gestural modalities are usually approached as two parallel productions and the question of a global unit of segmentation that would integrate all interplaying resources is still left open. The specifics language shapes of deaf children's productions have led us to rethink a unit of segmentation at the global production level in order to grasp the intra- and inter-modality dynamics that contribute, trough complex semiotic interplay, to the elaboration of utterances.

### 4.2. Towards a global unit of segmentation: an effective application of *growth point* (McNeill, 1992)

Indeed, the diversity of language configurations in the utterances produced by deaf children provides arguments supporting the language conception of the utterance formation proposed by McNeill (1992, 35),which he has conceptualized under the term of *growth point*[6]. As the following example (see the figure 2) illustrates, in the utterances of deaf children each resource can play both a specific and a complementary role in the elaboration of the structural and/or semiotic aspect of the utterances. While each unit is meaningless on its own, together they contribute to compound a global production that takes on complex meaning, through this interplay of plural resources. This example illustrates a typical structure, amongst others, of specific dynamics observed in deaf children's productions: one resource introduces the theme – tin this case, the French in this case 'the baby bird', another introduces the rheme – in this case, gestures in this case which depicting the trajectories of the baby bird (still an egg at the point in the cartoon) – and a third one is adjoined to specify a characteristic of the action for example – in this case, ononomatopeias underlining the brief and repetitive aspect of the egg's jump. Each resource represents an essential element of the structural and/or semantic aspect of the global production, without which the latter would be incomplete. The meaning of the following example acquired, through the complex interplay of verbal/non-verbal resources, can be translated as : there is a baby bird, an egg, and suddenly he has jumped everywhere a lot.

Therefore, in our grid, we opt for a unit of segmentation at the level of bimodality which corresponds to an effective application of the concept of *growth point* integrating all resources, vocal and/or gestural, verbal or non-verbal involved in the elaboration of utterances. Then, the first encoding action of the transcription process consists, in fixing

[5] See for examples of transcriptions: the grid recently proposed in ELAN by Pichler et al. (2010) to transcribe hearing children productions growing up in deaf families or more anecdotally (Bishop et al., 2006; Emmorey et al., 2005; Van den Bogaerde, 2000).

[6] "[...] when gesture and speech combine, they bring into one meaning system two distinct semiotic architectures [...] The GP [growth point] is the name we give to an analytic unit combining imagery and linguistic categorial content."(McNeill and Duncan, 2000, 144).

**Noa**
frs: le petit oiseau et là
*French: the baby bird and then*
onom.: poum poum poum poum poum poum
gestes: 2M jointes (conf.boule) rebondissent en différents points de l'espace
*gestures : 2Hands joined (conf.bawl) are jumping in various points of space*

Figure 2: Examples of the interplay of resources in theme/rheme elaboration

exclusively temporal boundaries, in order to delimit coherent semiotic sets by taking account of all resources that participate in elaborating each utterance. This takes place on a central line, which conditions the others 4 major lines of transcription – French, Onomatopoeias, LSF and Gestures –, In addition to semiotic criteria, syntactic criteria are used to delimit units where applicable. Given the state of development of linguistic skills in deaf children, the criteria of syntactic segmentation cannot always be operated. Therefore, we define these units under the concept of *semantic-syntactic units*. A *semantic-syntactic unit* can be built on one resource exclusively, vocal or gestural, verbal or non-verbal, or plural resources used in diverse combinations, either simultaneity or in sequentiality, as the following figure illustrates. This extract from the narrative of Driss, 5 years 1 months, schooled in an oralist setting, can be delimited into 4 *semantic-syntactic units* compounded by two units supported by bimodal non-verbal combination (unit 1 and 4), one unit supported by verbal/non-verbal combination (unit 2) and one unit supported by a non-verbal resource (unit 3).

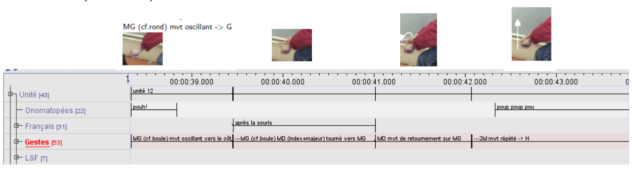


Figure 3: Examples of *semantic-syntactic* units segmentation

Thus integrating verbal and non-verbal resources in a single segmentation of language units allows us, at this primary step of description, to make the question of gestures/signs distinction less crucial and, in a broader sense, to do the same for the question of verbal and non-verbal status of the productions. Indeed, delimiting narratives on *semantic-syntactic units* allows deaf children's productions to be captured in their entirety without considering the question of their status and no matter the language form in which they are expressed.

### 4.3. Verbal and non-verbal resources dynamics : an unique insight into skills under development

Adopting an integrative perspective towards description, enables the effective symbolic abilities of deaf children to be accounted for, as a whole, no matter their state of development and no matter how far they have systematized them in a linguistic form. As illustrated by the following example (see figure 1), an extract from the narrative productions of Abdel, a young deaf child (6 years 4 months) schooled in an oralist setting, this perspective allows us to take into account the complex gestural processes that can be elaborated by deaf children despite not having been formally exposed to LSF.

Figure 4: Example of complex gestural processes elaborated by a deaf child schooled in an oralist setting

Abdel make use of the principles of spatiality and simultaneity that underpin encoding of events in the gestural modality so as to represent complex narratives events and parallel action, *i.e.* the egg arriving on the back of the mouse which is sleeping. As this extract shows, gestural productions of deaf children, even if they are not formally exposed to LSF, are not, in most cases, isolated gestural units contrary to what can be observed in the gestural productions of hearing children. The differences are more striking when it comes to description of cartoon scenes involving simultaneous actions, as we has shown elsewhere in a comparative analysis between the narrative productions of young hearing and deaf children schooled in an oralist setting in the first year of primary school (Estève and Batista, 2010; Estève, 2011). While deaf children represent the scene in space by making use of the possibilities for simultaneous representation afforded by gestures, the gestures of hearing children follow the linear and temporal description of the scene in keeping with linguistic encoding in French. The two following chosen examples correspond to the retelling of a sequence of the cartoon during which the baby bird is picking the chair on which the mouse is sitting.

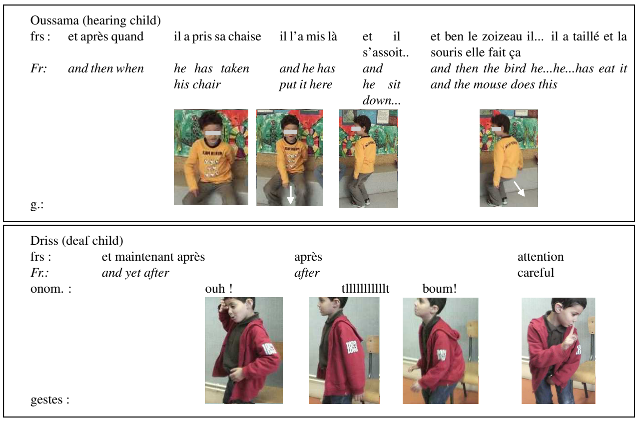



Figure 5: Comparative gestures processes used by hearing and deaf child for a scene implying simultaneous actions

So, as these examples suggest, for deaf children, the gestural modality is not simply an alternative way in which to represent events closest to the manner to perceive their organization in the real world as it's the case for hearing children. Rather it represents an alternative way to encode the information in a hierarchically organized and decontextualized symbolization, underlying the specific manner to encode events linguistically in SL. This way of representing the events is, given the deafness, closer to the way deaf children perceive and conceive of the events and is therefore their preferred manner of organizing and encoding the information. This contrast between hearing and deaf children's gestural production give full weight to the concept of *transitional competence* that we have developed : these crucial differences in gestures productions have to be linked with the fact that gestural skills in deaf children, even if they are not formally exposed to LSF, are in tension towards a linguistic potential.

### 4.4. Towards a linguistic potential: Some crucial cues for linguistic acquisition in progress

In the previous examples, Abdel (see figure 4) and Driss (see figure 5), both deaf children non formally exposed to LSF, re-invent, indeed, the processes used in narratives conducted in LSF. The *corporal proform* (Millet, 2002) – that other researchers consider as a *body classifier* (Morgan and Woll, 2003) – to refer to the mouse in the two cases – is maintained in action across successive units – the sleeping mouse for Abdel, the sitting mouse for Driss. This allows the simultaneous actions of the second character to be represented in parallel. Driss uses vocal resources to describe on onomatopoeic way the simultaneous action of the baby bird which is pecking at the chair on which the mouse is sitting. Whereas Abdel uses his other hand, which is not mobilized in the representation of the mouse's actions, to describe in gestures the trajectory of the baby bird – still in the egg at this point in the story – which is arriving behind the sleeping mouse. These two productions, and especially the Abdel's one, are consistent with the processes observed in adults' narratives behaviour within our copora. It should be note that in contrast to observe in this child production, for the same event of the cartoon, adults integrate in the movement, depicting the trajectory of the egg, the conventional manual configuration used to refer to a small and round object in LSF (*stf-objet-rond*).


Figure 6: Examples of narratives processes used by adults in LSF

This formal difference in the manual configuration led us to consider these units as a *manual proform* (Millet, 2002) in adults production. In contrast, in the child production we have considered this description as gestures and not signs. Thus, this observation can provide argument to evaluate more precisely the state of gestural skills development. Indeed, this deaf child seems has to be integrated some lexical or more precisely morpho-lexical elements of narrative process specific to the SL, and more specifically those implies in the anaphoric references : the *corporal proform*. While others are still not being systematized in a linguistic form, as it is the case for the manual anaphoric manner to represent a referent : *manual proform*.

On the other hand we have to note that, in the production of this child, the recourse to French serves the need to introduce a new referent in an isolated NP ("la souris" [the mouse], "le petit oiseau"[the baby bird]). And then the gestural representation of the two referents constitutes to a certain degree an anaphoric reference to the referent introduced in French. This specific cross-modal construction of the narratives processes is very frequent in our corpora.

The example from Abdel's narrative allows illustrates the fundamental complementarity of the symbolic skills that this deaf child, at his stage of language development, had developed in both vocal and gestural modality. Gestural and vocal skills of this child seem to be implied in contrasting steps of progression towards analytic modes of thinking, in contrasting steps of linguistic systematization, and in contrasting levels of the linguistic competence : lexical for vocal modality and morpho-lexical/morpho-syntactical for gestural modality. Indeed, on the basis of this examples, Abdel's competence can be evaluated as follow regarding the contrasting skills developed in the vocal and gestual modalities : he appears to have systematized some lexical skills in French and some morpho-lexical and morpho-syntactical skills in SL and more broadly in the gestural modality, which have not yet been integrated in a linguistic form in keeping with a particular SL, *i.e.* LSF.

So, Abdel's language behaviour provides strong arguments to underscore the crucial importance of taking into account how deaf children mobilize, combine and make progress in the two alternatives forms of encoding events which are available in the two channels of communication. On the one hand, in the vocal modality, a temporal and linear mode of speaking underpins the manner to represent events and, on the other hand, in the gestural modality, they call upon a spatial mode of speaking giving more place for representing simultaneity. Our integrative perspective allows to situate more precisely the deaf children's skills by integrating in the evaluation processes all the effective symbolic skills developed in each modalities. We can be able therefore situate each child in a particular state of progression towards analytic modes of thinking in two different modalities languages.

While the segmentation into *semantic-syntactic units* allows us to apprehend dynamics intra- and inter-modality as a whole, evaluating deaf child's symbolic skills more precisely implies also calling into question the tools used to transcribe and annotate each resource in order to be able to describe, in a more dynamic way, the evolution and continuity between non-verbal and verbal resources in the development trajectories of each child.

## 5. Evaluating the value of gestural units: when productions entail rethinking boundaries

### 5.1. Criteria for distinction between gestures/signs

The criterion used to distinguish between gestures and signs is, as usually, the reference of the deaf adults productions. However, not only their application is in fact very sensitive but the reference of adults productions does not solve all the issues surrounding the central question of gestures/signs distinction in the narrative productions of deaf children whose gestural skills are stills in development. On the one hand, this criterion tends indeed to set a separation between gestures and signs and it does not allow to take into account the proximity or the tension with SL. This tension incite in fact to consider the emerging of an intermediary language value between the both, which we have proposed to design under the concept of *quasi-linguistic*(Cosnier, 1982). Moreover, given the fact that narrative processes in LSF corresponds to the same basic processes that are used in the gestures of hearing speaker, but that have been systematized in a linguistic form in LSF (Millet and Colletta, 1997), fixing boundaries between gestures and signs is not without difficulty. And last but not least, applying this criteria implies necessary interpretation mechanisms (over-interpretation ?) of the gestural production and thus inevitably normalization mechanisms which can lead to aligning the children's productions with an "adultomorphic representation" (*représentations adultomorphes*) (Morgenstern, 2009). These transformation/transposition processes remove the materiality of the effective formal realization of the gestural production. This not only tends to freeze deaf children's productions in an adult form, but can also lead to substantial bias due to the fact the transcriber interprets children's production through a normalizing mechanism aligning them with an adult standard model.

Applying the criteria of adult references therefore implies a certain numbers of sensitive mechanisms. In order to reduce the part and place of the individual instinct of the transcriber, and given the state of description of narrative structures in LSF, we collected a corpus of 3 deaf adult narrative discourses retelling the same cartoon. The complexity of the examples which we have to confront with in our child corpora have lead us, to conceive annotation tools of gestures which are able to report the gestural shapes in their childhood reality.

### 5.2. Categorizing the non-verbal processes : proposition for a typology

The existing tools destined to annotate gestures of hearing children cannot be applying directly to the annotation of gesture's value for hearing gestures in the existing grid (Colletta et al., 2009) can't be apply to deaf children productions. Indeed, the category which retains our attention is the one of *referential gestures* which is conceived as "*gestures which have for function to design a referent if their can be perceived or representing it in space*" (Millet and Colletta, 1997). The perspective of the suggested typology is to detail all the gestural processes supporting common matrices shared between the hearing and the deaf, and which are particularly used in the encoding of narrative events. The following table summarizes some of the most frequent kinds of gestures used by deaf children of our corpora and which approach, in various ways, the narrative processes used in LSF.

| Types of gestures | Description |
|---|---|
| **Mimetic-action** | mimics the action or the behaviour of a referent by a global corporal interplay |
| **illustrative** | describes the characteristics of a referent (size, form, etc.) by the manual configuration or depicts the referent or a characteristic of the action in the space |
| **Spatiographics** | depicts in space the arrangement of the elements of the referential universe and/or gives a topographic representation of the arrangement of the elements in the actual space |
| **Endophorics pointing** | manuals or cephalics pointing gestures which refers to a locus, before (anaphoric) or later (cataphoric), assigned to a referent |
| **Trajectory-mimetic** | manual gestures depicting the trajectory of the referent |

Table 1: Typology proposed for the annotation of more frequent gestures used by deaf children

This typology is a first response to re-thinking the dynamics between gestures and signs in order to qualify how the processes that have not been systematized in a particular SL approaches verbal processes. Nonetheless, in their actual state of development, these tools do not enable us to assess precisely to the degree of proximity of these nonverbal processes structuring with the linguistic processes in LSF and the degree of the systematization of the units compounded the structure of these gestural processes. The perspectives opened by this final remark are still in exploration or waiting to be explored.

## 6. Less a conclusion, than a beginning

Rather than concluding, we will outline the perspectives and directions for further reason that our propositions have opened up. Of the numerous research questions raised by our propositions we shall start by pinpointing the key issue of delimiting gestural units, and especially which are not systematized in SL. Questions emerged both in the delimiting processes implied within the sets at the level of *semantic-syntactic* units and inside the blocs at the level of the units compounding a gesture. Indeed to situate precisely the evolution of gestural skills we have to be able to describe, with a fine granularity perspective, the degree of appropriation of the elements integrate in the gestural narratives process elaborated by deaf children. We therefore had to describe each piece of information encoded in the gestures categorized as non-verbal. Rather, this set of information is usually considered and transcribed as one single gesture. The example of Driss' narrative (see example 3) provides essential material for this discussion. Gestures used by this child – that we have transcribed as single gestures in the current transcription – can be describe in a more fine-grained perspective. For for bi-manual gestures, for example, it will be necessary to detail independently each gesture product and their components as different units – as manual configuration and movement amongst others. This description will allow us to contrast gestural processes elaborated by children in our corpora which encode different pieces of information in their gestures and which are nonetheless categorized as the same kinds of gestures in our current typology. These different encodings however provide cues on contrasted states of appropriation of the spatial encoding structure as the comparison of the extract from Driss and Abdel narratives shows. For example, it can be emphasized on the appropriation of the gestural processes which can support a double perspective of description, manual and corporal or manual/manual, for simultaneous actions. A fine-grained description of units will allow the precise identification of what kinds of elements of the specific manner of the linguistic encoding information in a particular SL are integrated. At this point, it will be necessary to consider in which way tools for SL description could be applied to gestures.

Furthermore, this article has concentrated on the gestural dynamic, especially given the focus of this workshop, but the reality of the bimodal dynamics encourage to think over, in a more general sense, of the way in which non-verbal/verbal dynamics are integrated in the transcription/annotation tools. A crucial challenge is supported by taking into account the *transitional skills*, inherent to the development processes, and emerging alongside the continuum between the verbal and non-verbal extremes which are usually conceived as two static states of skills. The ”non-verbal” components of school-age deaf children’s productions are particularly neglected in the schooled context as in research. While our corpora based on a representative sample of heterogeneity of deaf children primary age-schooled show the importance of the deaf children skills which are not systematized in a particular language(s) – no matter it is vocal or/and gestural. These language realities incites to concentrate our efforts of comprehension of the later development in the context of deafness on the crucial issues raised by *transitional skills*. In our view, these challenges involve two central points: providing the elements about individual development trajectories in the context of deafness and modelling the various steps of language progression up to linguistic skills on them multiple forms. Our current work engages precisely with these two perspectives. A longitudinal corpora is currently being compiled, with 17 nursery school children and 25 primary school children. For the first time, this will provide cues on the heterogeneity of the transitional steps between language – in a broader sense – and linguistic skills in SL and VL.

# 7. References


M. Bishop, S. Hicks, A. Bertone, and R. Sala. 2006. Capitalizing on Simultaneity: Features of Bimodal Bilingualism in Hearing Italian Native Signers. In Ceil Lucas, editor, *Multilingualism and sign languages: From the great plans to Australia*, pages 79–118. Gallaudet University Press, Washington D.C.

J.-M. Colletta, R. Kunene, A. Venouil, V. Kaufmann, and J.P. Simon. 2009. Multi-track annotation of child language and gestures. *Multimodal corpora*, 5509:54–72.

J. Cosnier. 1982. Communications et langages gestuels. In J. Cosnier, J. Coulon, A. Berrendonner, and C Orecchioni, editors, *Les voies du langage, Communications verbales gestuelles et animales*, pages 255–292. Bordas, Paris.

M.C Da Cunha Pereira and C. De Lemos. 1994. Gesture in hearing mother-deaf child interaction. In Virginia Volterra and Carol J. Erting, editors, *From gesture to language in hearing and deaf children*, pages 178–186. Gallaudet University Press, Washington D.C.

K. Emmorey, H. Borinstein, and R. Thompson. 2005. Bimodal bilingualism: code-blending between spoken English and American Sign Language. In *Proceedings of the 4th International Symposium on Bilingualism*, pages 663–673. Somerville : 30 avril - 3 mai 2003.

I. Estève and A. Batista. 2010. Utilisation et construction de l’espace dans les conduites narratives d’enfants sourds oralisants et d’enfants sourds entendants. In B. Aiosa, F. Naït-Bouda, and M. Thévenon, editors, *Repères et Espace(s). De la pluralité à la polysémie*, Grenoble. Presses Universitaires de Grenoble.

I. Estève. 2011. *Approche bilingue et multimodale de l’oralité chez l’enfant sourd : outils d’analyse, socialisation, développement.* Ph.D. thesis, Université de Grenoble, Grenoble.

I. Fusellier-Souza. 2004. *Sémiogenèse des langues des signes. Étude de langues des signes primaires(émergentes) pratiquées par des sourds brésiliens.* Ph.D. thesis, Université Paris VIII.

S. Goldin-Meadow. 2005. *The resilience of language: What gesture creation in deaf children can tell us about how all children learn language.* Psychology Press, New-York.

D.H. Hymes. 1984. *Vers la compétence de communication.* Hatier, Paris.

S. Kita. 2000. How representational gestures help speaking. In David McNeill, editor, *Language and gesture*, pages 162–185. Cambridge University Press, Cambridge.

A. R. Lederberg and P. E. Spencer. 2001. Vocabulary development of deaf and hard-of-hearing children. In D. Cark, M. Marschark, and M. Karchmer, editors, *Context, Cognition and Deafness*, pages 88–112. Gallaudet University Press, Washington, D.C.

C. Lucas and C. Valli. 1992. Language contact in the american deaf community. In Ceil Lucas and Clayton Valli, editors, *Linguistics of American Sign Languages*, pages 458–480. Academic Press Inc, San Diego.

D. McNeill and S. Duncan. 2000. Growth point in thinking-for-speaking. In David McNeill, editor, *Language and gesture*, pages 141–161. Cambridge University Press, Cambridge.

D. McNeill. 1992. *Hand and mind: What gestures reveal about thought.* University of Chicago Press, Chicago.

A. Millet and J.-M. Colletta. 1997. Gestualité référentielle des entendants et LSF: des matrices communes. In *Actes du Congrès International des Linguistes*, Toulouse : 20-25 avril 1997. [on line].

A. Millet, I. Estève, and L. Guigas. 2008. Pratiques communicatives de jeunes sourds adultes. Rapport pour la délégation générale à la langue française et aux langues de france, Lidilem, Université de Grenoble. [on line].

A. Millet. 2002. Les dynamiques iconiques et corporelles en langue des signes française (lsf). *Lidil*, 26:27–44.

H. Mohay. 1994. The interaction of gesture and speech in the language development of the two profoundly deaf children. In V. Volterra and C. J. Erting, editors, *From gesture to language in hearing and deaf children*, pages 197–204. Gallaudet University Press, Washington D.C.

GD Morgan and B. Woll. 2003. The development of reference switching encoded through body classifiers in british sign language. In K. Emmorey, editor, *Perspectives on classifiers construction*, pages 271–296. Lawrence Erlbaum Associates, Mahwah, New Jersey.

A. Morgenstern. 2009. *L’enfant dans la langue*. Presses Sorbonne Nouvelle, Paris.

C. Pichler, J.A. Hochgesang, D. Lillo-Martin, and R.M. de Quadros. 2010. Conventions for sign and speech transcription of child bimodal bilingual corpora in ELAN. *Language, Interaction and Acquisition/Langage, Interaction et Acquisition*, 1(1):11–40.

E. Pizzuto, A. Barbara, M.C Caselli, and V. Volterra. 2001. Cognition and language in italian deaf preschoolers of deaf and hearing families. In D. Cark, M. Marschark, and M. Karchmer, editors, *Context, Cognition and Deafness*, pages 49–70. Gallaudet University Press, Washington, D.C.

Carolina Plaza-Pust and Esperanza Morales-Lopez, editors. 2008. *Sign Bilingualism : Language development, interaction, and maintenance in sign language contact situations.*, volume 38 of *Studies in Bilingualism.* John Benjamins Publishing Company, Amsterdam.

B.S. Schick, M. Marschark, and P.E. Spencer, editors. 2005. *Advances in the sign language development of deaf children.* Oxford University Press, New York.

P. E. Spencer and M. Marschark, editors. 2006. *Advances in the spoken language development of deaf and hard-of-hearing children.* Oxford University Press, New York.

B. Van den Bogaerde. 2000. *Input and interaction in deaf families.* LOT Publishers, Utrecht.

# Linking an ID-gloss database of ASL with child language corpora

**Julia Fanghella[1], Leah Geer[2], Jonathan Henner[3], Julie Hochgesang[4], Diane Lillo-Martin[1], Gaurav Mathur[4], Gene Mirus[4], Pedro Pascual-Villanueva[1]**

[1]University of Connecticut, [2]University of Texas at Austin, [3]Boston University, [4] Gallaudet University
{julia.fanghella, diane.lillo-martin}@uconn.edu,{julie.hochgesang, gaurav.mathur, gene.mirus}@galluadet.edu,
leah.geer@utexas.edu, jhenner@bu.edu, pedrodpedro@gmail.com

**Abstract**

We describe an on-going project to develop a lexical database of American Sign Language (ASL) as a tool for annotating ASL corpora collected in the United States. Labs within our team complete locally chosen fields using their notation system of choice, and pick from globally available, agreed-upon fields, which are then merged into the global database. Here, we compare glosses in the database to annotations of spontaneous child data from the BiBiBi project (Chen Pichler et al., 2010). These comparisons validate our need to develop a digital link between the database and corpus. This link will help ensure that annotators use the appropriate ID-glosses and allow needed glosses to be readily detected (Johnston, 2011b; Hanke and Storz, 2008). An ID-gloss database is essential for consistent, systematic annotation of sign language corpora, as (Johnston, 2011b) has pointed out. Next steps in expanding and strengthening our database's connection to ASL corpora include (i) looking more carefully at the source of data (e.g. who is signing, language background, age, region, etc.), (ii) taking into account signing genre (e.g. presentation, informal conversation, child-directed etc), and (iii) confronting the matter of deixis, gesture, depicting verbs and other constructions that depend on signing space.

**Keywords:** ID-gloss, ASL corpora, lemmatization

## 1. Introduction

A lexical database that lists a unique gloss, also known as ID-gloss, for each sign is indispensable for annotating corpora consistently (Johnston, 2010). The human transcriber, when left to rely on their own memory for retrieval of unique glosses, is more likely to produce errors in the transcript. Continuous use of the database during the transcription process allows the human transcriber a more efficient retrieval system that will reduce the amount of errors in the transcript. As corpora grow, they feed the lexical database in turn, providing tokens of signs that need unique glosses. This paper reports on an on-going project, the ID-Gloss Project as reported in Alkoby et al. (2010), to develop a lexical database of American Sign Language (ASL) as a tool for annotating sign language corpora collected in the United States.

### 1.1. Database design

The design of our ID-gloss database (Alkoby et al., 2010) is unique, and reflects the current scholarly approach to sign linguistics in the United States. Several different research labs work on sign languages, but no set of systematic, consistent, nationally accepted glosses exists. For this reason, our database was developed so as to permit different research groups to provide site-specific information corresponding to a common set of lexical signs. It is not that we wish for a set of standard glosses to be used in ASL research but rather we wish to facilitate cross-lab data comparison, which is the aim of this project. Each lab completes locally chosen fields (gloss, phonological information, word class, etc.) using their annotation system of choice, and picking from a global template of available fields that was designed in collaborative meetings. The fields are then merged into the global database, thereby providing complementary information for each sign. Figure 1 provides a schematic diagram of our system. The diagram illustrates that the global site houses all of the media files, which are linked to each of the local databases and the local database is linked with the respective local template (a subset of the global template).

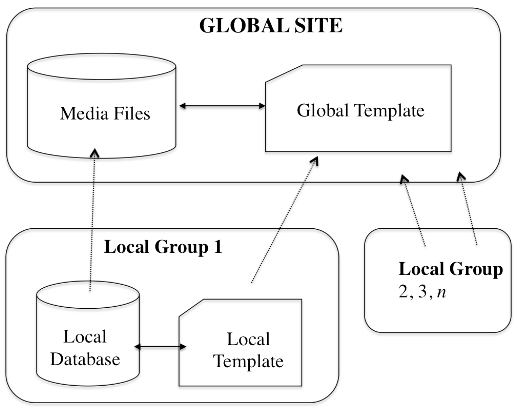



Figure 1: Schematic diagram of the databasae structure.

### 1.2. Approach: methods of database construction

In this section, we will discuss how our database is built. We begin with a description of what types of annotations each group contributes to the database and focus on the phonological and phonetic notes and descriptions we make for each sign. Other annotations include morphosyntactic information like part of speech and sociolingusitic information like regions or age-groups that might use a particular sign, among other things.

#### 1.2.1. Research groups

Our approach currently capitalizes on the annotation perspectives of three different research groups. The first group comprises researchers at Gallaudet University and the University of Connecticut (for convenience, this group is referred to as G/UC). This group follows transcription conventions developed by Chen Pichler et al. (2010). Phonological information about signs is entered using Stokoe Notation (Stokoe et al., 1965) as well as phonetic information about hand configuration using Sign Language Phonetics and Architecture – SLPA – (Johnson and Liddell, 2011). The second group, referred to as BTS henceforth, includes researchers from Boston University and Gallaudet University using the Berkeley Transcription System, or BTS (Slobin and Hoiting, 2002). The third group includes researchers at the University of Texas (UTX). This group also uses SLPA for each of the formational parameters of signs.

#### 1.2.2. Data annotation

The three notation systems mentioned in §1.2. have individual strengths that contribute to the structural integrity of the ID-gloss database. Stokoe's notation system is the oldest and most well known by many sign language linguists. The notation system is based on three major sign parameters, namely the *dez* (handshape), *tab* (location) and *sig* (movement). The handshape in a sign can be described by 19 possible labels and additional diacritics. The location of a sign can be represented by symbols that represent certain areas of the body, from the face to the hips. The possible placements can be more specific (forehead, mid-face, chin, or cheek/temple) or more general (trunk). The movement can be represented by 24 symbols. These include directionality (upward or downward movement), internal movement such as wiggling, and movements to contact and grasping movements.

BTS notation was developed in order to be compatible with the CHAT transcription system (MacWhinney, 2000) used by the CLAN analysis programs in Child Language Data Exchange System (CHILDES). The full BTS system allows for annotation of polycomponential signs, such as classifiers and depicting signs. With respect to handshape notation, BTS allows for more fine-grained distinctions between than Stokoe (e.g., the ASL 'A', 'S', and 'T' handshapes all receive different notations in BTS but receive the same label in Stokoe Notation). The BTS uses 68 distinct subhandshapes, elaborated from 10 more abstract handshape categories.

Johnson and Liddell (2011) propose a segmental approach, the SLPA, to the phonetic notation of signs wherein each segment is notated with information about the handshape, movement, placement (like Stokoe's *location*), contact, and orientation of the hand(s). The use of Johnson and Liddell's notation system is time consuming, but it is useful for gathering phonetic-level detail about the production of each sign in the database, something neither Stokoe nor BTS notations provide. Additionally, while Stokoe Notation provides a single label for each parameter of the sign (e.g., [A] for the handshape pictured in Figure 2), SLPA provides a componential notation for the behavior of separate elements in each part of the sign. For example, the hand configuration in SLPA is represented by a series of symbols that describe the joint behavior of each finger and thumb as well as the arrangement (relationship between fingers, such as crossing) and contact, if any, between the fingers and thumb. Thus, the [A] in Stokoe notation is phonetically annotated as [LEE<1FF=2FF=3FFe=4FFe] in SLPA.

Figure 2: [A] handshape.

The choice of each notation system depends on a research lab's theoretical orientation and research goals. The database allows for each group to use their preferred notation system. Another advantage of the database system is that other research labs can then see what notations are used by other research labs.

### 1.3. Linking the database with an ASL corpus

One advantage to our overall design is flexibility in assigning glosses to signs, since it allows local groups to gloss the same sign in potentially different ways. Furthermore, the different groups may provide different sets of additional information for each sign. While each group gives at least information in the fields *gloss* and *alternative gloss*, the groups provide different phonological information, and choose between optional fields for morphosyntactic, sociolinguistic, and other types of information. Each lab has access to the information entered by other labs, so the work is mutually beneficial. Moreover, the flexibility afforded by the system may lead to eventual convergence on glosses, a desirable outcome. In our project, the annotation of corpora and the building of a lexicon have so far been independent processes, but here we evaluate our progress to date, and discuss concerns for the continued development of these resources.

## 2. Methods

For the present paper, we are comparing the glosses for signs in our ID-gloss database to annotations of spontaneous child ASL data from the BiBiBi project (Chen Pichler et al., 2010). The child data were annotated by coders trained to use consistent glosses, but they worked without access to the developing database. The first set of signs selected for inclusion in the database were expected to be ones that would likely occur in the child language corpus. Thus, this makes a good test case for examining the efficacy of a cyclical approach to simultaneously building a lexicon and corpus. Specifically, as we start comparing the corpus annotations to the signs in the database, we notice that some of our predictions are borne out (e.g., signs we predicted would be used were indeed used), but other signs are missing from the database. For example, consider the table in Table 1.

| Related Signs in Corpus | ID-gloss | Missing (or inconsistently glossed) signs |
|---|---|---|
| EAT/FOOD | EAT | |
| PICK/FIND | PICK | |
| SEARCH/LOOK-FOR | SEARCH | check consistency |
| MY/MINE | POSS(self) | check consistency |
| SAME/SAME-AS | SAME | SAME-TIME |

Table 1: BiBiBi glossing conventions: need verification with ID-gloss database

We can see the potential challenges in glossing: it is tempting to use context to distinguish 'search' from 'look-for', for example, but doing so is incompatible with the goal of maximizing searchability by having a unique gloss per sign type. This process helps to fuel the cyclical process to adding signs to the lexical database then returning to transcription with the expanded list. This report includes comparison of approximately 650 of the 1000 signs currently in our database, which have been assigned ID-glosses by the G/UC group.

Five sessions of spontaneous ASL data from one child in this corpus were selected for analysis. These sessions were collected when the child, Ben, was age 1;07 (years;months), 1;10, 2;01, 2;04, and 2;07. The total number of individual child productions at each session is given in the second column of Table 1. Out of these productions, the following were eliminated: uninterpretable productions (coded as YYY or XXX according to conventions), gestures, mouthing (in the absence of a manual sign), fingerspelling (coded as FS), pointing (coded as IX or POSS), depicting (coded as DV). The remaining items are lexical tokens, the number of which is given in the third column of the table. Finally, repeated tokens of the same type within a session were reduced, providing the number of lexical types, given in the fourth column.

| Age | Total child utts | Lexical Tokens | Lexical Types |
|---|---|---|---|
| 1:07 | 459 | 105 | 42 |
| 1:10 | 854 | 340 | 77 |
| 2;01 | 445 | 175 | 60 |
| 2;04 | 625 | 275 | 95 |
| 2;07 | 454 | 213 | 81 |

Table 2: Data set used for analysis

Each of the lexical types produced by Ben was compared against the set of ID-glosses entered by the G/UC team. We calculated the proportion of Ben's lexical types that were shared with the database and contrasted them with the those that were not yet in our database (unshared types).

## 3. Results

The results of our analysis are presented in Figure 3.

Overall, 63% of Ben's lexical types are included in the database. As the figure illustrates, the proportion of shared

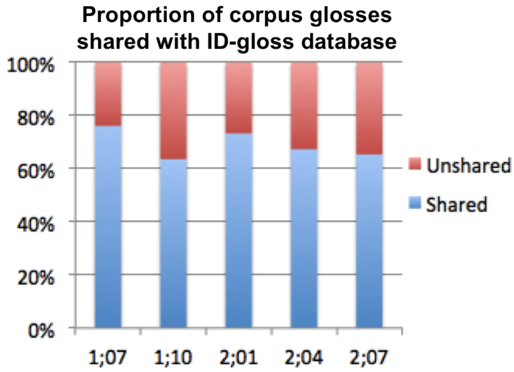



Figure 3: Proportion of glosses shared with database.

types decreases over the year's worth of data, from 76% to 66%.

## 4. Discussion

In this section, we will discuss what we better understand about the process of ID-glossing based on the comparison we have just outlined in §2. and §3. We particularly focus upon the challenges that are presented by the missing glosses (particularly phonological variants) and challenges that are involved with growing the database (e.g., how can we involve the community, a question that we have been concerned with from the onset of our project). But first, a quick discussion of the results presented in §3.

### 4.1. Interpretation of the results

As we saw from Figure 3, the proportion of lexical types produced by Ben that were also in the database decreased. One possible interpretation of this is that Ben's lexicon grew, as one might expect, from 1;07 to 2;07, but the database had not yet been updated with examples of his newer lexical items. This reinforces the circular process we have been describing, wherein the database encourages consistent transcription, but corpus transcription encourages expansion of the database when new items are encountered.

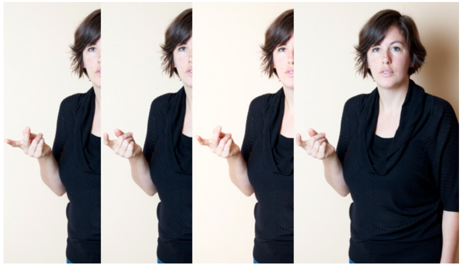
### 4.2. Missing items & glossing challenges

Many of the glosses that do not appear in the database currently are signs that will be added in the near future (e.g., CHICKEN, FALL-DOWN, GRANDMOTHER, WATERMELON). Others are numbers; they are annotated conventionally, but not included in the database. Perhaps a future step regarding numbers might be to include signs with numeral incorporation, as these do vary, and we would want to be able to capture any variation with an appropriate gloss in our database. A few glosses indicate that the annotators did not follow the annotation conventions consistently (e.g., MINE was used although the conventions called for POSS(self)). Despite the overall utility of the database, which reaffirmed the need for ID-glossing, several problem cases were revealed that deserve attention. The issue that will be addressed here is, how many ID-glosses should be assigned for signs that resemble one another and represent the same concept, or the issue of phonological variation.

#### 4.2.1. Treating phonological variation

Consider, for example, the word 'dog': in order to account for different versions of the ASL sign for 'dog', the database includes unique glosses for two distinct phonological forms for the same concept: DOG$_{snap}$ and DOG$_{slaphip}$. The annotators of the child corpus used three different glosses, however: DOG$_{snap}$, DOG$_{slapsnap}$, and DOG. The use of the gloss DOG was the result of human transcriber error. The transcriber should have appended 'snap' to the gloss. This error could have been avoided if the transcriber had access to the database. For the other variant, DOG$_{slapsnap}$, it would be possible to add to the database, and include the phonological information distinguishing the three variants in every annotation in the child database, eliminating the use of the underspecified gloss DOG. An alternative option would be to have one unique gloss DOG and leave the identification of phonological variants to secondary tagging. The question of how many distinct glosses are needed for the different forms of 'dog' – and other items that have multiple phonological forms – is one that needs to be addressed more thoroughly. Johnston (2010) discusses this very issue and suggests using a separate ID-gloss for each phonological variant, and to tag phonological information (such as handshape or movement) onto the gloss. This was the approach taken here with the different versions of 'dog'. An approach similar in principle – provide phonological information to distinguish sign variants – is used by the researchers who provide glosses using BTS. With that system, all phonological variants with different initial handshapes will necessarily be distinctive. DOG$_{snap}$ is glossed [KT]DOG and DOG$_{slaphip}$ is labeled [BU]DOG, where [KT] and [BU] are the names of the initial handshapes used to produce these signs. However, DOG$_{slapsnap}$ would have the same representation as DOG$_{slaphip}$, since they both use the same initial handshape. To the extent that these variants should be differentially glossed, the technique of using handshape information to distinguish varying phonological forms for the same concept is appealing, although it is not sufficient to distinguish all cases.

Another method, which is now the standard for the Auslan corpus (Johnston, 2011a), is to use a single ID-gloss for all *minor* phonological variants, and then specify the phonological information through secondary tagging (e.g. on a separate tier in ELAN). What is unclear, however, is what is meant by "minor" in terms of phonological variants. Does this mean one quantifiable difference, or more? In our view, the changes in the the phonological forms of DOG$_{slaphip}$, DOG$_{snap}$ and DOG$_{slapsnap}$ are not minor. There are several changes form to form, which can be seen in Figure 4. Specifically, DOG$_{snap}$ is produced in neutral space and requires the index finger to flex and make contact with the thumb in a snapping motion which is repeated (Figure 4a). DOG$_{slapsnap}$, pictured in Figure 4b, requires the hand, with all fingers extended to make contact with the thigh then move to neutral space where the fingers change their configuration and snap once. Finally, DOG$_{slaphip}$ does not have the snapping motion of DOG$_{snap}$ or DOG$_{slapsnap}$ but has the patting motion from DOG$_{slapsnap}$ - that is, DOG$_{slaphip}$ is when the hand needs to move to the thigh and make contact twice or so (Figure 4c).



(a) DOG$_{snap}$

(b) DOG$_{slapsnap}$

(c) DOG$_{slaphip}$

Figure 4: Three different tokens of 'dog' in ASL

There are pros and cons to each approach. Under the first approach, the transcriber would be burdened with generating (or requesting) a new ID-gloss for any new phonological variant produced in a text and adding it to the database. However, this would be a burden only for the first time a particular variant is encountered. If the same variant is used again, the transcriber need only select the correct ID-gloss from the database. On the other hand, if we follow Johnston's approach, the transcriber would not be as burdened the first time. Yet, the burden would be delayed until time came for secondary tagging.

The transcriber would have to add the same phonological information on a secondary tier every time the variant appears, leading to redundancy within the transcription.

#### 4.2.2. What is the purpose of the analysis?

While there are clear benefits and challenges with either approach, the following question should be considered carefully before deciding which approach to follow: What is the goal of the analysis? If, for instance, the analysis is more focused on the meaning (syntactic / morphological structure of the child's utterances), and less on phonological variation, then Johnston's approach should suffice. On the other hand, if the focus of the analysis is on the frequency and environment of the phonological variants themselves, then Johnston argues that it will be sufficient to conduct a search on the secondary tagging.

What is clear from this discussion is that comparisons between the annotated data and the glosses in the database validate our need to develop an electronic link between the lexical database and the corpus. Such a link (currently under development) will serve two important purposes: it will help to ensure that annotators use the appropriate ID-glosses, and it will allow new glosses needed to be readily detected (Johnston, 2010; Hanke and Storz, 2008). As the database grows, this strict process will also allow researchers to make accurate estimates of lexical frequency, something that is lacking in most sign language research to date and something which has implications for further work in various subfields of linguistics.

### 4.3. Challenges in growing our database

While modest evidence indicates that the ID-gloss database will be useful for annotating the child language corpus, König et al. (2010) identified several issues that accompany the development of a gloss database. Three of these issues will be discussed here: a) ethical issues related to assigning glosses to signs, b) reconciling the variety of gloss notation systems used by various labs, and c) challenges in glossing signs that do not have one-to-one translation equivalents.

#### 4.3.1. Glossing & community involvement

First, who gets to decide the glosses for the signs, the researcher(s) or the language community? Hochgesang et al. (2010) offer a framework for addressing this issue (but see also Dudis et al. (2009) for another perspective). In particular, it is important to be transparent throughout the research process and involve the community of language users at each juncture. To give an example, consider the ASL signs EAT and FOOD. It may not be clear at this point if there is any systematic way in which the form of these signs differs when they appear in citation form (though it is certainly clear with context), the community of ASL users may have opinions about which English word is a better unique identifier, and it is these intuitions researchers should be attentive to and consider when assigning glosses to lexical items. Discussion on how to best allow community participation is on-going.

#### 4.3.2. Different glosses, different labs

The second issue concerns consistency in glossing. We will discuss this issue as it relates to both within-lab and across-lab concerns. The difficulties of utilizing a glossing system for a signed language are well established (Pizzuto and Pietrandrea (2001) and see also our discussion in §4.2.). One way to maintain gloss consistency within a group is by using an available dictionary for as many signs as possible. The BTS group uses The American Sign Language Handshape Dictionary (Tennant and Brown, 1998) for glosses where possible, supplemented with an explicit handshape symbol at the beginning of each sign, as we discussed in §4.2.1.. As we have already mentioned, the development of an electronic link between corpus and database will even more greatly facilitate consistency within each group. In the United States, the issue of cross-lab inconsistency is a major concern. The diversity of sign language research labs, and the lack of national glossing standards, results in differences in glossing at the lexical level. Our database was specifically designed to allow for these differences and nevertheless permit cross-lab comparisons and eventual cross-corpus searches. In addition to the lab's primary gloss for a sign, each lab completes information about alternate glosses in a separate field in the database. If the database is queried, the displayed results will match either the main gloss or the alternate. This allows more flexibility in use as well as maximizing the ability to search the database. To continue the example from above, consider the sign for 'eat': One lab used the gloss EAT with the alternative gloss FOOD, while another lab did the opposite. In a cross-lab search, all of the relevant information is still retrieved with a single query.

#### 4.3.3. Signs without English translation equivalents

The third, and final, issue we will discuss here is the difficulty in assigning glosses to signs that do not have suitable translation equivalents between ASL and English. Thus far, our database does not include, classifiers , depicting signs (except for one) and other polycomponential forms, but in order for the database to reflect language use in ASL corpora, these are forms that will need to be included. Johnston (2011b) offers possible solutions, based on his work in developing an ID-glossed corpus for Auslan. He makes distinctions between signs that are fully lexical, those that are partially lexical and those that are non-lexical (e.g., gesture and emblems). For partially lexical items, or signs that are regular in form but for which the meaning is conditioned by the context of utterance, Johnston (2011b) suggests using some sort of indication of what type of partly lexical item it is (classifier, depicting signs, etc.), what handshape is used, and what does it mean in a particular context. To use Johnston's example, a partly lexical sign might be glossed DSH(F):describe-as-appropriate, where "DSH" indicates that it is a depicting sign with an "F" handshape. Another similar type of example from our database is pictured in Figure 5. Here, the UTX coder uses "DS" to indicate that the form is a type of depicting sign, "F" to indicate the handshape, and *"long-skinny"* to reflect the type of noun being classified. When this form is used in naturalistic signing, as opposed to citation form, the *"long-skinny"* designation can be replaced with whatever is indicated by the given context.

Figure 5: DS(F):long-skinny.

We will have to decide lab-internally, or as a group, what sorts of glosses we find most appropriate and fitting for this category of signs. One approach would be to adopt the glossing techniques of (Johnston, 2011b) wherein we distinguish between fully, partly, and non-lexical items. This would allow us, crucially, to capture enough about the form of the sign to encourage consistent application of a unique ID-gloss, but allows for flexible additions to the semantic content

## 5. Conclusions

Here we have shown that, despite challenges that persist in developing an ASL lexical database, the linking of transcripts within a corpus to such a database will aid in understanding crucial facts about the language. An ID-gloss database is essential for consistent and systematic annotation of sign language corpora, as Johnston (2010) has pointed out and as we have attempted to demonstrate. We provided preliminary results from a comparison between our database and an annotated corpus that did not have the benefit of an ID-gloss database. There are several logical next steps to consider in expanding the ID-gloss database and in strengthening its connection to ASL corpora. In closing, we will mention three of these steps. First, we should look more carefully at the source of the data and document the metadata. It is important to know who is signing, what is their language background, how old are they, where did they grow up and where do they live now. All of this information will help generate a more complete picture about how ASL is used and what differences exist between groups (e.g., regional groups, age groups etc). Second, we should take into account the genre of signing. Was a particular text from a presentation or an informal conversation? Was an adult directing signing at an infant or child? This will contribute valuable information that can lead to descriptions of distinct linguistic registers in ASL. Lastly, as we mentioned briefly in §4.3.3., we need to confront the matter of deixis, gesture, depicting verbs and other constructions that depend on signing space. Each of these questions will help grow our database, as well as allow for more accurately annotated corpora and thus strengthen the link between the the two.

## 6. Acknowledgements

The project described in this article was supported by Award Number R01DC009263 from the National Institute on Deafness and Other Communication Disorders. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health.

## 7. References

K. Alkoby, J.A. Hochgesang, G. Mirus, and P. Pascual Villanueva. 2010. Construction of an ID Gloss database. In *Proceedings from the* $10^{th}$ *meeting of the Theoretical Issues in Sign Language Research Conference*, West Lafayette, IN.

D. Chen Pichler, J.A. Hochgesang, D. Lillo-Martin, and R. Müller de Quadros. 2010. Conventions for sign and speech transcription of child bimodal bilingual corpora in ELAN. *Language, Interaction and Acquisition/Langage, Interaction et Acquisition*, 1(1):11–40.

P. Dudis, G. Mathur, and G. Mirus. 2009. Bringing a corpus in line with deaf communities. In *Proceedings from the Sign Language Corpora: Linguistic Issues workshop*, University College of London.

T. Hanke and J. Storz. 2008. iLex – a database tool for integrating sign langauge corpus lingustics and sign language lexicography. In O. Crasborn, E. Efthimiou, T. Hanke, E.D. Thoutenhoofd, and I. Zwisterlood, editors, *Proceedings of the Third Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora at the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 64–67, Marrakech, Morocco.

J.A. Hochgesang, P. Pascual Villanueva, G. Mathur, and D. Lillo-Martin. 2010. Building a database while considering research ethics in sign language communities. In *Proceedings of the* $4^{th}$ *Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies at Language Resources and Evaluation (LREC'10)*, pages 112–116, Valletta, Malta, May.

R. E. Johnson and S. K. Liddell. 2011. A segmental framework for representing signs phonetically. *Sign Language Studies*, 11(3):408–463.

T. Johnston. 2010. From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, 15:104–129.

T. Johnston, 2011a. *Auslan Corpus Annotation Guidelines*. Centre for Language Sciences, Department of Linguistics, Macquarie University, Sydney, Australia.

T. Johnston. 2011b. Lexical frequency in sign languages. *Journal of Deaf Studies and Deaf Education*, 36:1–31.

S. König, R. Konrad, G. Langer, and R. Nishio. 2010. How much top-down and bottom-up do we need to build a lemmatized corpus? In *Proceedings from the* $10^{th}$ *meeting of the Theoretical Issues in Sign Language Research Conference*, West Lafayette, IN.

B. MacWhinney. 2000. *The* CHILDES *Project: Tools for Analyzing Talk.* Lawrence Erlbaum Associates, Mahwah, NJ, $3^{rd}$ edition.

E. Pizzuto and P. Pietrandrea. 2001. The notation of signed texts: Open questions and indications for futher research. *Sign Language & Linguistics*, 4(1):29–45.

D. I. Slobin and N. Hoiting. 2002. The Berkely Transcription System for sign language research (BTS). In Gary Morgan and Bencie Woll, editors, *Directions in sign language acquisition*. John Benjamins, Amsterdam.

W. C. Stokoe, D. C. Casterline, and C. G. Croneberg. 1965. *A dictionary of American Sign Language on linguistic principles*. Linstok Press, Silver Spring, MD.

R. A. Tennant and M. G. Brown. 1998. *The American Sign Language Handshape Dictionary*. Gallaudet University Press, Washington.

# A Study on Qualification/Naming Structures in Sign Languages

**Michael Filhol, Annelies Braffort**
LIMSI-CNRS
Campus d'Orsay bat 508, BP 133, F-91403 Orsay cx, France
michael.filhol@limsi.fr, annelies.braffort@limsi.fr

**Abstract**

In the prospect of animating virtual signers, this article addresses the issue of representing Sign, in particular on levels not restricted to the language lexicon. In order to choose and design a suitable model, we illustrate the main steps of our corpus-based methodology for linguistic structure identification and formal description with the example of a specific structure we have named "qualification/naming". We also discuss its similarity and difference with other Sign properties described in the literature such as compound signs. Consequently we explain our choice for a description model that does not separate lexicon and grammar in two disjoint levels for virtual signer input.

**Keywords:** Sign Language animation; grammar modelling; weak hand persistence; compound signs; corpus analysis

## 1. Introduction

With the purpose of formally representing Sign Language (SL) elements and rules to generate animations and automatically produce SL utterances via a virtual signer, we have worked on the DictaSign corpus to identify various grammatical rules. This paper presents our methodology and the result of what is the first cross-SL study of the project. In terms of SL processing, the outcome of such research will benefit Sign synthesis by specifying what should be performed from a grammatical rule. Also, we believe that it can assist grammatical annotation tasks by specifying surface cues to be caught by image processing software.

Among other linguistic structures, we have identified one that we called the "qualification/naming structure", which constitutes the main focus of this report. It has the interest to be a structure which surface form can also be found in compound lexical units.

First, we describe the methodology used for the cross-language corpus observation, then, we discuss the constraints that must be represented by our formalism, and we conclude on how to refine the current results.

## 2. Methodology

Two approaches are possible to determine a systematic rule between a semantic structure or relation and a surface (phonetic) production: start from either the semantic function or the surface form. The structure presented here was discovered using the latter, as follows.

We have selected gestural units composed of one-handed signs performed by the strong hand while the weak hand is kept activated immediately after the end of a two-handed sign. More precisely, we consider structures containing what Liddell named "fragment buoys". A fragment buoy is the final handshape of a sign that has just been performed which is then held in the signing space while other signing activity continues on the other hand (Liddell, 2003). In a fragment buoy, the signer uses the fragment or handshape of a previous sign S0 as a buoy because S0 is referred to by other signs interacting with it (Johnston, 2011).

Figure 1 shows an example in LSF. The two-handed sign S0 is the sign LINE. It is followed by three one-handed signs YELLOW, U and THREE. The intent is to specify a subway line, the line Yellow with the name U3. The sign LINE is clearly held up by the weak hand and remains tense throughout the following three one-handed signs.

Figure 1: LSF example: LINE YELLOW U THREE.

To annotate our corpus, we used labels close to those proposed in the Auslan annotation guidelines (Johnston, 2011). A fragment buoy is labelled FBUOY, followed by a colon and the IDgloss of the two-handed sign S0. In our example, that is FBUOY: LIGNE (for LINE).

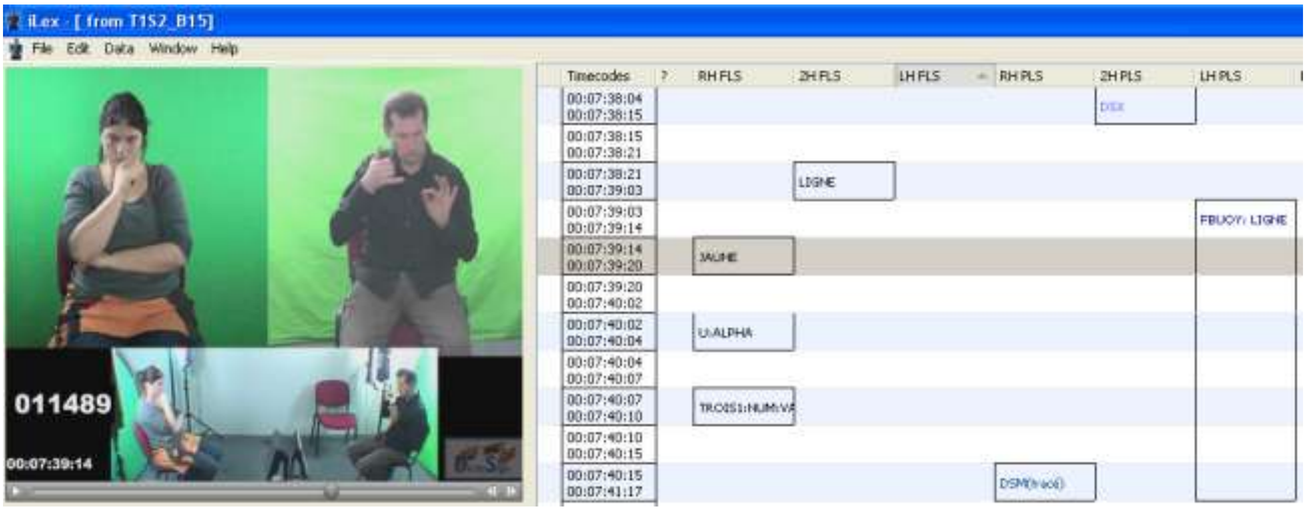



Figure 2: Annotation of our LSF example in iLex using FBUOY label.

Figure 2 shows a snapshot of the corresponding annotation in iLex (Hanke 2002; Hanke & Storz 2008). Time flows from the downwards. We use three tiers for the fully lexical signs (RH FLS, LH FLS and 2H FLS, for the activity of respectively the right hand, the left hand and the two hands), three tiers for the *partly lexical signs* (RH PLS, LH PLS and 2H PLS), and other tiers not detailed here. All the terminology is explained in (Johnston, 2011). Fully-lexical signs are what is often called conventionalised signs or standard signs and are identified with a ID-gloss that is the identifier of the entry in the sign lexicon database. Partly lexical signs include pointing signs, depicting signs, and buoys. See (Johnston, 2011) for a detailed explanation on how to identify and annotate these signs.
In Figure 2, the tier 2H FLS contains the ID-gloss LIGNE (for LINE), and while the LH PLS segment is labelled with FBUOY: LIGNE, the RH FLS tier contains successive segments with the ID-gloss JAUNE, U and THREE (for YELLOW, U and THREE).

T. Johnston suggests that if the activity on the weak hand is not meaningful, for example if it seems only to be the continuation of part of the previously articulated sign and to slowly relax to a neutral handshape or rest position, one must only annotate information for the strong hand. But in our annotation, we did consider the cases excluded by Johnston were there was no topological relationship between S0 and the following one-handed signs (thereby excluding things like classifier predicates, more semantically loaded and based on a lot more than a mere sequence).

Using our annotated part of the corpus (5 hrs of LSF dialogue), we have collected more than 500 occurrences of FBUOY segments and applied the following process.

(a) <u>Choice of target occurrences to collect from the corpus</u>
From these FBUOY segments, we had first noticed a large number of occurrences where the weak hand was held while the strong hand continued <u>on without the two being linked by any geometric or topological reason</u> (like pointing to the weak hand, or depicting a path holding the weak hand as a locative). This led us to define the "unrelated weak hand persistence" criterion as follows:
*A two-handed sign S0 is performed followed by one or more one-handed gestures while the final posture of the weak hand is held in place*.
Strong hand: |__ S0 __| |__ 1-handed signs __|
Weak hand: |__ S0 ______ held from S0 ________

(b) <u>From form to function in LSF</u>
We collected a minimum of 150 clear occurrences of the surface form described in (a), and found that all fitted either of the two categories below:

1. Qualification/naming: The one-handed utterance on the strong hand qualifies S0 like an adjective, or names it with a name-sign or finger-spells something to identify it. It can be a combination of those.
2. Conservation of activation: S0 is held because it is needed again after the one-handed sequence (S0 usually repeated then). This can be seen as a parenthesis in a discourse, during which S0 is to be kept "active".

(c) <u>From function to form in LSF, DGS and GSL</u>
The next step of the process was to submit this finding to the Greek and German teams and begin a cross-language verification process based on the LSF, DGS and GSL parts of the Dicta-Sign corpus. All languages were searched for occurrences of the qualification/naming semantic function above (b1), and the corresponding forms observed. The SLs were observed by local experts

separately and their feedback allowed us to suggest the following statement:
*When S0 is a 2-handed sign followed by one or more qualifying or naming 1-handed signs, the weak hand tends to be held strongly in its last S0 posture while the other signs are performed with the strong hand*.

Figures 4 and 5 shows examples extracted from DGS and GSL.

Figure 3: DGS example with TICKET 'rectangular object'.

Figure 4: GSL example with SALAD DELICIOUS.

This identified structure, to be animated in the hands of a virtual signer, must be formalised to enforce a temporary hand separation and synchronise them on a common timeline. The next section illustrates this process and raises a few linguistic questions.

## 3. Representation and discussion

Azalee is a representation model that allows specifying different parts of a signing activity independently, and that distributes them in time (Filhol, 2011; Filhol, 2012). It has two important properties, which makes it our choice to base our discussion to come:

1. Sign Language productions enrol several simultaneous parts, usually overlapping in time; Azalee defines 'time intervals' (TI), one for each separate part of the production, represented as a box in the diagrams below.
2. Sign Language productions are flexible in many ways, some of the variability is meaningless, some have an effect on the semantics; Azalee deals well with this aspect as it uses minimal sets of necessary constraints.

**Question 1:** Representation of the necessary and sufficient conditions
Fragment buoys are represented as follows (Figure 5), where "S0" is the eponymous TI for the initial two-handed sign, the qualifying/naming 1-handed sequence following S0 is composed of S1 and S2, and FBUOY represents the effect weak hand holding.

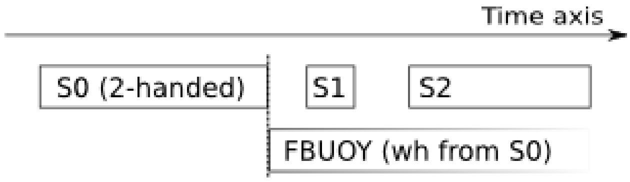



Figure 5: Time arrangement for weak hand persistence after S0

Pertaining to property no. 2 above: There is variability in the point where the FBUOY ends, but invariably signers hold it for a minimum of time. What is the necessary condition? Our model allows to constrain it to the longest commonly used time across signers, which does not force any animation to retract the hand past this boundary.

**Question 2:** Boundary between lexicon and syntactic structures
Another question appears when comparing this structure with compound signs. A compound sign is a lexical unit, whereas we deal with grammatical constructions not registrable as signs. We have noticed the presence of fragment buoy structures in the LSF lexicon database built during the Dicta-Sign project, where each entry corresponds to a given concept. The example shown in Figure 6 corresponds to the concept "relative". It is expressed in LSF with a compound composed of the signs FAMILY and PERSON. The weak hand is held from the sign FAMILY while the strong hand signs PERSON.

Figure 6: LSL expression of the concept "relative", which is signed FAMILY PERSON.

There is undoubtedly some similarity between the two constructions (phrase or lexical level). Though we would need quantitative measures on the start and end of the TIs to allow proper comparison of the dynamics and rhythm (they may differ by that only), this question already leads us to question the opacity and even the relevance of the boundary between lexicon and syntactic structures.

**Question 3:** Weak hand anticipation

If we invert the diagram above, we end up with a new phenomenon, analogous to what Johnson calls “weak hand anticipation” (Liddell & Johnson, 1986). Again, this has to do with lexical compound signs, i.e. signs composed of several lexical signs and including progressive, or in this case regressive, assimilation.

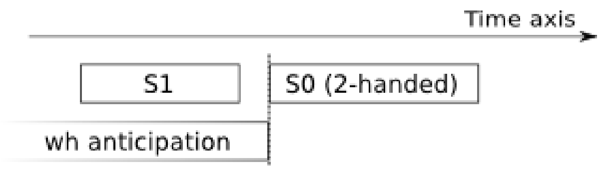



Figure 7: Time arrangement for weak hand anticipation after S0 (inverted fig. 5)

An interesting issue to raise at this point is to transpose question 2 on this inverted diagram. Indeed once again, weak hand anticipation is defined as a lexical property of compounds. But similarly to our observation in Q2, can we question this statement by finding any construction using wh-anticipation and still variable enough to be excluded from the lexicon?

In any case, we insist that Azalee be designed without assumption regarding these questions, both to ensure coverage of all structures and to provide Sign experts with a formalism to write down all possible approaches of a phenomenon. Indeed, only then can we efficiently debate over differences in representations and discover categories instead of having them assumed by the model. Given our observations above, this statement leads us strongly to advocate the use of a model with no immutable gap between lexicon and syntax.

## 4. Future work and conclusion

We have used more than 500 times the label FBUOY in our annotation, and we have not analysed all of them. A deeper and extensive analysis must now be conducted, in order to refine these first results on various aspects, and first of all, by verifying if there can be other semantic categories than the two presented in 2.b for this given surface form. We could use for example the same kind of approach that this used in (Nishio, 2009).

Then we must analyse other parts of the corpus that contain the qualification/naming semantic function that are not annotated with FBUOY. For example, we have to reply to the following questions:

- When another surface form is used (only one-handed signs, only two-handed signs, S0 being a one-handed sign and the following ones two-handed...), can we observe other frequent properties? We have hypothesised the following: “the shoulder line does not move during the sequence, and the time between S0 and the following signs is shorter than average”
- Is it possible that S0 is signed after the qualifying signs, and in which case?

This paper has presented a Sign linguistic structure for qualifying and naming 2-handed concepts. We have mentioned the unclear lexical vs. syntactic status of the productions using this structure, and explained the need for a representational model that does not make any strong division between those two levels of language. Future work awaits ahead in the study of more linguistic structures, always with the aim of full coverage by the description models.

## 5. Acknowledgements

The research leading to these results has received funding from the European Community, under grant agreement FP7-231135 for the Dicta-Sign project.

## 6. References

Dicta-Sign project website: http://www.dictasign.eu/

Dicta-Sign corpus portal: http://www.sign-lang.uni-hamburg.de/dicta-sign/portal/

Filhol, M. (2011). Formalism for multi-linear rules of Sign Language grammars. Formal and Experimental Advances in Sign language Theory" (FEAST) colloquium, Venice, Italy: http://virgo.unive.it/venicefeast/02_poster/21_P_Filhol.pdf

Filhol, M. (2012, tbp). Combining two synchronisation methods in a linguistic model to describe Sign Language, E. Efthimiou, G. Kouroupetroglou, S.-E. Fotinea (Eds.), Gesture and Sign Language in Human-Computer Interaction and Embodied Communication, Springer LNCS/LNAI 7206.

Hanke, T. (2002): iLex - A tool for sign language lexicography and corpus analysis. 3rd Int. Conf. on Language Resources and Evaluation (LREC). Canaria, Spain.

Hanke, T. & Storz, J. (2008): iLex – A database tool for integrating sign language corpus linguistics and sign language lexicography. Int. Conf. on Language Resources and Evaluation (LREC). 3rd Workshop on the Representation and Processing of Sign Languages (RPSL) . Marrakech – Morocco.

Johnston, T. (2001). Auslan annotation guidelines: http://www.auslan.org.au/video/upload/attachments/AuslanCorpusAnnotationGuidelines30November2011.pdf

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

Liddell, S. K. And Johnson, R. E. (1986). American Sign Language compound formation processes, lexicalisation and phonological remnants. Natural Language and Linguistics Theory, 4(4), pp.445-513.

Nishio, R. (2009). Corpus-based analysis of Weak Drop and Weak Prop in German Sign Language (DGS), “Sign Language Corpora: Linguistic Issues” workshop, University College London on 24-25 July, 2009

# Towards tagging of multi-sign lexemes and other multi-unit structures

**Thomas Hanke, Susanne König, Reiner Konrad, Gabriele Langer**

University of Hamburg
Institute of German Sign Language and Communication of the Deaf
Binderstr. 34, 20146 Hamburg, Germany
E-mail: {thomas.hanke, susanne.koenig, reiner.konrad, gabriele.langer}@sign-lang.uni-hamburg.de

**Abstract**

With the building of larger sign language corpora tagging, handling and analysing large amounts of data reach a new level of complexity. Efficiency and interpersonal consistency in tagging are relevant issues as well as procedures and structures to identify and tag relevant linguistic units and structures beyond and above the manual sign level. We present and discuss problems and possible solution approaches (focussing on the working environment of iLex) of how to deal with multi-unit structures and more specifically multi-sign lexemes in annotation and lexicon building.

**Keywords:** sign language tagging, multi-word expressions, annotation tools

## 1. Multi-sign expressions as lexicon entities

For many sign languages, compounds and idiomatic expressions are attested to occur. Becker (2003) differentiates between proper compounds and loan compounds in DGS (German sign language), with the first group being rather rare. Johnston & Ferrara (to appear) report that multi-sign idiomatic expressions are rare as well.[1] However, some of these units may have not been discovered yet since empirical studies on these topics require large amounts of data, which are only now becoming available with the development and accessibility of large sign language corpora. In addition, often there are no clear-cut distinctions between the three groups mentioned.

Whatever the exact definitions for these phenomena are, they have something in common: When these patterns appear, they are different from just the signs they consist of. There may be restrictions on the use of these patterns not be expected from how their components can be used, and the meaning might be different from the composition of meanings of their building blocks, or they might disambiguate POS attributions to their parts. This means they have to be considered part of the language's lexicon. Once multi-unit structures are stored in the lexical database, they can also be attributed with all kinds of lexicographic annotation such as regional use or syntactic restrictions. This would of course also be the place to further characterise the construction, e.g. what kinds of variation does it allow.

## 2. Tagging multi-sign expressions

In today's coding conventions (e.g. Johnston 2011, but also including our own), these multi-unit structures are not really dealt with in a way that allows to mark, access, list and describe these units as entities of their own right. Usually, only their building blocks are made visible in the annotation. Occurrences of these structures need to be retrieved by executing searches. This is unsatisfactory not only in the context of complex patterns difficult to search for, but also from the point of view of lemma revision. Searches will retrieve all occurrences of the patterns, and the information whether the pattern is actually used in the special (e.g. idiomatic) sense or in the literal that is sign-by-sign sense is not stored anywhere. We therefore look for a possibility to clearly identify tokens of multi-sign structures while maintaining the tagging of the constituents.

A first approach would be to have a separate tier tagging those time stretches where multi-sign structures occur with labels such as "idiom" or "compound". To find the multi-sign instances, one would then set up the search not only containing the sign pattern, but also the extra label. Nesting structures would require multiple extra tiers which is uncomfortable but still manageable. This approach, however, does not generalise to other forms of multi-unit structures such as multi-channel signs (e.g. a sign with an obligatory lexicalised mouth gesture) or discontinuous structures such as sandwich verbs or resumed holds: A simple tag in a separate tier would either include all co-occurring events in a tier or none.

An alternative actually in use for spoken languages multi-word expressions is to store them as the pre-terminal level in a treebank. However, today there is no annotation system for sign languages featuring treebanks.

What we currently envision is to add this lowest level of syntax trees to the tiers & tags model of iLex[2]. As we would not expect more than two levels above the basic token tags, these extra levels would be projected onto the iLex annotation grid display by framing those tags

[1] Their claim is for idiomatic expressions in Auslan, but the same seems to be the case for DGS.

[2] iLex is the transcription environment we use which features a lexical database closely integrated with the annotation scores, cf. Hanke (2002) and Hanke & Storz (2008).

constituting a multi-unit structure. Token structures compatible with the multi-unit structure as stored in the lexicon can be claimed an instance of that structure by dragging the lexical item onto one of the existing tokens.

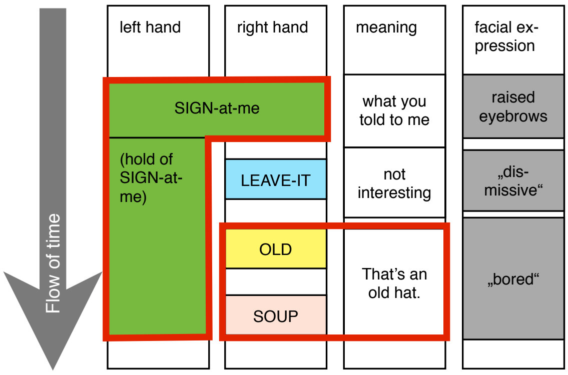



Figure 1: Idiomatic expression and hold structure

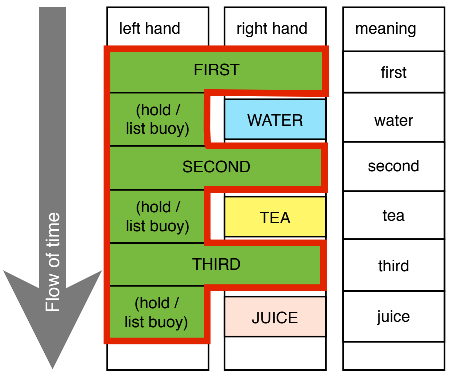



Figure 2: Multi-item list structure

## 3. Multi-sign lexemes in the lexical database

Naturally, this approach requires the lexicon structure in iLex to be extended in order to cover multi-sign multi-channel structures within the lexical database. Currently, a simplex sign is described as either one- or two-handed, with an optional code for mouthing or mouth gesture that may be copied to the mouth tier but is not considered part of the token. Complex signs are either simultaneous or sequential compounds or blends of two simplex signs. More complex structures cannot be appropriately handled in the implemented lexicon model. The idea is to allow any kind of element (simple signs, nonmanuals etc.) to be arranged in a structure expressing time relations such as “precedes” or “precedes immediately”. To the user, these structures would appear as miniature transcripts without concrete timestamps.
An extension allowing multi-channel signs would also cover obligatory facial actions for lexemes in a much more transparent way than the current solution.

## 4. 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.

## 5. References

Becker, C. (2003). *Verfahren der Lexikonerweiterung in der Deutschen Gebärdensprache*. Hamburg: Signum.

Hanke, T. (2002): iLex. A tool for Sign Language Lexicography and Corpus Analysis. In González Rodriguez, M. & Paz Suarez Araujo, C. (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.

Hanke, T., Storz, J. (2008). iLex - A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In Crasborn, O., Efthimiou, E., Hanke, T., Thoutenhoofd, E., Zwitserlood, I. (Eds.), *LREC 2008 Workshop Proceedings. W 25: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*. Paris: ELRA, pp. 64-67. Available online at http://www.lrec-conf.org/proceedings/lrec2008/workshops/W25_Proceedings.pdf

Johnston, T. (2011). *Auslan Corpus Annotation Guidelines. 30. November 2011*. Available online at http://www.auslan.org.au/video/upload/attachments/AuslanCorpusAnnotationGuidelines30November2011.pdf

Johnston, T., Ferrara, L. (in press). Lexicalization in Signed Languages. When is an Idiom not an Idiom ? Online Proceedings of UK-CLA Meetings [Proceedings of the 3rd UK Cognitive Linguistics Conference, University of Hertfordshire: 6-8 July 2010.] Available online at http://mq.academia.edu/TrevorJohnston/Papers/902380/Lexicalization_in_signed_languages_when_is_an_idiom_not_an_idiom

# Where Does a Sign Start and End? Segmentation of Continuous Signing

**Thomas Hanke, Silke Matthes, Anja Regen, Satu Worseck**
Institute of German Sign Language and Communication of the Deaf,
University of Hamburg
{thomas.hanke,silke.matthes,anja.regen,satu.worseck}@sign-lang.uni-hamburg.de

**Abstract**

We present the rules how to segment continuous signing into individual sign tokens as used in the projects Dicta-Sign and DGS Corpus and compare this approach to others. We then report on experiments applying the rules to high-speed video.

**Keywords:** segmentation

## 1. Introduction

Segmentation in the sense of tokenisation usually is one of the first steps in any sign language transcription work as it is the prerequisite to lemmatisation which in our view is at the very heart of sign language annotation. There are two basic approaches how to segment continuous signing into individual signs:

- A sign starts where the preceding one ends (i.e. fluent signing means there are no gaps between signs)
- Transitional movements between signs do not count as part of either sign. Therefore, usually there are gaps between two signs during which the articulators move from the end of one sign to the beginning of the next.

Johnston (2011:38-39) favours the first approach where time intervals not tagged indicate periods of no signing activity.[1] We have traditionally followed the second approach. In the context of the DGS Corpus and the Dicta-Sign project that approach offers advantages for the subsequent processing: First of all, variation between tokens is much lower than if the transition would be part of the sign. Secondly, a token tag represents only that part of the signal that is described by HamNoSys, which allows for more straightforward processing in the context of recognition and animation of continuous sign language. Boundaries between sign and transition also make it possible to separate sub-sign analysis from movement properties of the transitions. Obviously, one has to deal with the ambiguity if non-tagged time intervals stand for transitions or non-activity. In the past, we used heuristics based on the duration of non-tagged intervals: Transitions tend to be short compared to natural or even deliberate pauses.[2] With image processing becoming available (cf. Dubot & Collet, this volume), the ambiguity can be resolved without any further manual tagging. Automatic detection of manual activity vs. non-activity provides rather robust results that combine with the manual tagging to tell transitions (outside token tags, inside automatically tagged manual activity) apart from non-activity (outside token tags, outside automatically produced tags).[3] We therefore do not share Johnston's concern that our approach would result in false results e.g. when calculating overlaps between manual and nonmanual prosody.

A major concern for us is data quality. Variation of ±2 frames (at 25fps) within and between experienced annotators was unexpectedly high. We therefore detailed our segmentation criteria as much as seemed to make sense.

## 2. Segmentation Rules in Dicta-Sign and the DGS Corpus Project

The approach chosen for Dicta-Sign and the DGS corpus project is to cut off transitional movements from the actual signs. This leaves the annotators with the task to decide where exactly a certain sign starts and where it ends.

While the general aim is a bottom-up (i.e. data-driven) approach for sign language annotation, a certain amount of top-down decisions seems unavoidable in such an approach. (We use our knowledge about the type to cut a token.)

For signs with an HMH structure in the sense of Liddell & Johnson (1989) (or PTP in the sense of Johnson & Liddell (2011)) the sign starts at the beginning of the initial hold, i.e. as soon as its handshape has been formed and is placed in the right orientation at the starting location of the sign. Likewise the sign ends at the end of the hold, i.e. just before the first change of one of the parameters.[4]

[1] He actually suggests leaving gaps of "at least one frame" between subsequent tags, but only for technical reasons inherent to ELAN, the transcription environment used.

[2] These heuristics are unable to determine the turn-final return to rest position but Johnston's approach shares this problem as the turn-final tokens include the transition into the sign and out of the sign whereas all others only include the transition into the sign.

[3] Of course, this solution is not without problems either. It provides false positives in the case of manual activities that are neither signing nor gesturing, but for example scratching oneself (that you may want to notate only if you assume some communicative intent) or manipulating physical objects, e.g. drinking. False negatives, such as subtle backchanneling, are compensated by manual tagging.

[4] In comparison, Crasborn & Zwitserlood (2008:5-6) cut after

For other structures more specific definitions are needed:
Sign starts:

- In cases where two signs share a hold (i.e. one sign ends in a hold, and by chance the next sign is beginning with a hold at exactly the same location with the same handshape and orientation), cut the hold in the middle. (Here it is obvious that there cannot be a gap between the two tags.)
- In case of signs without a specific starting location, look for a discontinuity in the movement (e.g. a sudden change in direction) between the end of the previous sign and the end of the target sign. That point is then the starting point.
- In case of a continuous movement from the beginning of a sign to the end of the next sign (e.g. DENKEN[5] DU[6] in lax signing), cut in the middle/at the peak of that movement. (This is then also the end of the previous sign, i.e. there is no gap in-between the two signs.)

Sign ends:

- If the sign finishes with a movement, then cut just before a change of movement direction.
- If there is no change of movement, a change of handshape or orientation marks the end of the sign.
- In case there is no change of handshape or orientation but a continuous movement from the previous to the following sign, the sign ends in the middle / at the peak of that movement (see above).

For two-handed signs, in principle the above criteria can be applied to both hands individually. However, for some cases this results in different timings for the two hands (which is possible to tag if two separate token tiers are used, but at the expense of more time needed for segmentation). When using one tier, and that also holds for cutting the video itself, which is what counts for image processing, a combined criterion has to be defined. The easiest and most consistent definition to cut both hands in parallel is to just concentrate on the dominant/active hand and ignore the other (i.e. following the above rules).
Nonmanual activity is not considered at all when segmenting unless there is no manual activity. In that case, start and end of the movement define the duration of the sign.

## 3. Agreement Measures

This detailed decision tree, however, did not increase intra- and inter-transcriber agreement substantially.
Annotators reported that they still followed their intuition and only applied the rules step by step when in doubt. So it seems that annotators' intuition is strong, but nevertheless not precise to the video frame or that even native signers of the same sign language differ in their intuitions. Brentari & Wilbur (2008) suggested that people might pay attention to different parts of the signing stream when segmenting, but their research did not explain why that should still be the case for annotators who are signers of the same language.
One of the obvious difficulties in finding the right point in time for cutting is that signing movement has to be reconstructed from the images in the video frames available. So one hypothesis was that this problem would become easier with higher frame rates. In an experiment, we asked annotators to apply the same rules to a video shot at 50fps, and in fact they reported that they were more confident in their decisions (although not faster). Agreement still was in the range of ±2 frames which now corresponded to only half the time jitter experienced before. While this convinced us to move all annotation work to 50fps videos (either shot natively or deinterlaced from 25fps at the expense of spatial resolution), we were still unsure how much our rules depended on the video's temporal resolution.

## 4. Compatibility with the Johnson/Liddell Phonetic Model

In a small-scale study aiming at improving avatar performance naturalness, we compared our segmentation with the approach proposed by Johnson & Liddell (2011), aiming at detecting the beginning and end of a sign by identifying its sequential structure (Hanke et al. 2011).
According to Johnson & Liddell signs may not only be analysed as consisting of simultaneously occurring parameters (hand configuration, placement...), but they also show a sequentially organised sublexical structure consisting of alternating postural and transitional phases. A detailed segmentation for each of the individual parameters involved reveals the varying timing of changes happening during a sign: the parameters are neither established all at the same time nor do they change simultaneously. A posture in Johnson & Liddell's sense refers to those moments where all the parameters are stable and momentarily aligned (which may even last for only one frame). The picture of the hand is stated to be clearer than during the rest of the sign, which might be due to a slowdown of the hand's movements. During a transition, changes may occur in several parameters at a time, however these changes do not necessarily coincide and parameters are not all in place at exactly the same moment.
It turned out that defining postural and transitional phases is by no means a straightforward task. The suggestion given by Johnson & Liddell to distinguish clear pictures of the hand from fuzzy ones was mostly not applicable for our data as it depends heavily on the cameras used (esp. frame rate and exposure time). Having used videos with a frame rate of 50fps (i.e. larger than the 30fps available for the Johnson & Liddell data), we had expected to be able to recognise distinct static phases. However, the more frames there are, the more

---

the hand moves away from the initial location, i.e. after the initial hold.

[5] The examples given in this paper are all from DGS. THINK: Index finger upwards, palm towards body, hand moving away from contact with right temple.

[6] YOU

details are visible. This holds especially for signs that – on a first glance – inherit a comparably long placement (e.g. INDEX pointing at something). Looking at these occurrences frame by frame reveals the almost nonstop minor movements happening "naturally". For signs with short static phases, however, not always being able to rely on pictures being fuzzy or clear causes similar problems, namely the lack of criteria to define a phase as static that only lasts for one frame. Furthermore, the short static phases of the individual parameters do not necessarily show an overlap in time (i.e. postures in Johnson & Liddell's sense). It becomes evident that certain thresholds would need to be applied, however reliability still is an issue for human annotators.
In cases where postures were easily identified, they suggested sign boundaries coinciding with those determined by applying our segmentation rules set, given again a tolerance of two frames. In a couple of instances where postures had to be postulated as described above, we had slightly larger differences between the two criteria. This does not come unexpected as we apply different weights on the different parameters constituting the sign.

## 5. Segmenting High-Speed Video

In order to determine how much our approach actually depends on the video's frame rate and whether at a higher frame rate our approach would provide the same results as our procedures following Johnson & Liddell, we did another experimental recording with two cameras. One was a standard HD camera capturing at 720p50 (spatial resolution of 1280x720, temporal resolution 50fps), the other one was a high-speed camera working at 1080p500 (spatial resolution of 1920x1080, temporal resolution 500fps[7]). Due to the physical size of the high-speed camera, camera viewpoints are substantially different.
With 500 frames per second and correspondingly short exposure times, motion blur in signing no longer is an issue: All frame images are clear.[8]
The signing recorded had a length of 23.4 seconds containing 47 tokens.
The 50fps movie was separately annotated by three different annotators, two hearing and one Deaf native sign language user. For comparison, the 500fps movie was annotated by annotator A. In each case the annotator did not see the annotation done by the two others in order to avoid any influences.
Regardless of the sign being performed one- or two-handed, the segmentation concentrates on the dominant hand only (which is the right hand for this informant).
Due to the fact that iLex, the transcription environment used in our projects, currently cannot cope with movies with a temporal resolutions higher than 100fps, we had to convert the 500fps movie to slow motion. The disadvantage for the annotators is that they cannot watch the movie in real speed.
In general, the 500fps did not change the picture. Unlike the move from 25fps to 50fps, the 500fps movies did not make the annotators' job easier. In fact, they complained about the time needed to check infinitesimally small movements.
In the rest of this section, we report on problem cases observed by the annotators.

Defining the starting location (PL) of a sign turned out to be difficult in cases with a change in movement direction. This is often not a straight change of direction, but includes a slight curve or rotating movement. In these cases the tagging of the different annotators in the 50fps clip varies:

### 5.1 HOCHHAUS (high-rise building)

*Beginning of the sign (5 frames difference):*
After the end of the preceding sign the hand makes a downward movement, during which the HC of HOCHHAUS is established. Annotator B and C (with C one frame after B) tagged the beginning of the sign where the lowest point seems to be reached and the movement direction changes. However, the hand does not move straight down and up again but performs a small curve movement towards the body while changing movement direction. The definition of one frame as a PL is therefore mainly a theoretical assumption. Furthermore, this means that the FA (facing) is not fully established which violates the first segmentation rule. It seems, however, that a change in movement direction functions as a strong indicator for segmentation and might overrule FA (this was also reported for other occurrences).
According to the tagging of annotator A, the sign begins five frames later. The annotator stated she felt not able to define a certain point of time in the movie as a PL and therefore set the cut when the hand started to move straight upwards and the FA was in place. However, when segmenting the 500fps movie her tag matched the tags of annotators B and C. She reported that in the 500fps movie there were longer sequences where hardly any movement was visible (i.e. the movement is slower), while in the 50fps movie there were distinct changes from frame to frame that made it difficult to decide for one specific frame as a starting location.
A further occurrence of HOCHHAUS was found in the data where the beginning of the sign is set less low in signing space which minimises the curve movement. The tagging of this token is almost the same for all annotators (one frame difference).

*End of the sign:*
During the upwards movement the HC changes in anticipation of the following sign. However, the exact end of the sign (i.e. the point of change for HC) is not

[7] Actually, the high-speed recording was done in stereo, but for the purpose of this paper, only the left channel was used.
[8] This is the reason why we could not compare segmenting the 500fps video with segmenting a copy of that video down-sampled to 50fps, as motion blur is an issue with regular 50fps recordings.

perceptible due to the fuzziness of the picture in the 50fps movie. For the 500fps movie the camera perspective does not allow recognising the change of HC.

Picture 1: Sign HOCHHAUS during upward movement (50fps movie)

## 5.2 GLAS (glass)

*Beginning of the sign (2 frames difference):*
Similar to the example above the change of movement direction from upwards to downwards involves a small curve movement of the hand (including a temporary change of FA), which was tagged at the assumed peek of the movement change by annotator B. Annotator A set the tag border two frames later while annotator C's tag is in the middle of the two others. Again, annotator A reported on difficulties defining a PL in the 50fps movie, but set a tag boundary matching annotator B's tag when segmenting the 500fps movie.

*End of the sign:*
This was identically tagged by all annotators.

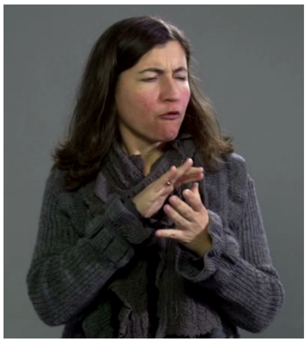
Picture 2: Sign GLAS at the beginning of the sign (500fps movie)

Our approach is not strictly bottom-up (i.e. data driven) as annotators use their knowledge about a sign type when deciding how to cut a certain token. In the following cases this led to differences in the segmentation (interestingly mainly between the Deaf and the hearing annotators):

## 5.3 FREUND (friend)

*Beginning of the sign (4 frames difference):*
Annotator C felt that the sign begins when the hands are closed and segmented the token accordingly. The tags of annotator A and B start four frames earlier as the annotators identified a discontinuity in the transition from the previous sign to the hands' contact in FREUND (confirmed by the 500fps movie). HC and FA are in place in about the middle of this movement (tag border) and the hand then moves straight down. Though only manual components were used to identify tag borders, it can be noted that the mouth pattern "freund" also begins before the hands' contact (see picture).

*End of the sign:*
This was tagged identically by all annotators. (The sign type shows a movement of both hands together, however this token only shows a contact of the hands, ending with a release and immediate transition to the following sign.)



Picture 3: Sign FREUND during the downwards movement (500fps movie)

## 5.4 URLAUB (holiday)

*Beginning of the sign (4 frames difference):*
Annotator C tagged the beginning of sign where the thumb makes contact with the body (analogue to the type form description). According to the tagging of annotator A and B the sign starts 4 frames earlier: Again HC and FA are in place in the middle of the movement from the end of the previous sign to the moment of contact (tag border) and the hand then moves straight towards the body. While the 500fps movie does not seem to provide any extra hints on how to segment the sign, annotator A in this case decided not to tag the movement towards the body, as she felt the movement was much too long to be part of the sign.

*End of the sign (3 frames difference):*
While the type description states finger wiggling, the actual token shows a simple closure of the fingers (except index finger, presumably because of the following sign "ME"). In the 50fps movie the tags for annotator A and B end at the same point of time, while annotator C's tag is three frames longer, including those parts of the closing movement of the hand where the fingers are not yet bent. In the 500fps movie annotator A also includes part of this movement into the sign (see picture below). According to her, the movement looked smoother than in the 50fps movie and was therefore regarded as part of the sign.

Picture 4: Sign URLAUB, end of the tag (500fps movie)

### 5.5 AUTOFAHREN (driving a car for a longer time)

*Beginning of the sign (3 frames difference):*

Annotator A and C tagged the beginning of the sign where the hands move away from the body. The tag from annotator B starts three frames earlier as it includes the preceding movement towards the body. (In the preceding sign the hand makes contact with the body. The hand then moves away from the body while forming the sign's HC which is in place at the end of the movement path (tag border) and then moves back towards the body.) However, annotators A and C see the backward movement as a transitional movement as a car is typically (and certainly in the given context) moving forward and therefore the sign should start with a movement away from the body. Additionally, the forward movement seems to be more emphasised than the backward movements. (This holds for the assumed transition as well as the intra-sign movements.) The 500fps movie does not provide any extra hints, as multiple minimal movements complicate the decision where to cut.

*End of the sign:*

This was tagged identically by all annotators.

Picture 5: Sign AUTOFAHREN during forward movement (500fps movie)

## 6. Conclusions

The annotation of the experimental high-speed recordings gives interesting insights on reasons why annotators disagree. Often these are related to how they judge personal contextual variation. This means that we cannot expect better agreement by further sharpening the criteria for segmentation, but have to tolerate some variation if we mix bottom-up and top-down (here pre-existing knowledge about the sign type's prototypical movement) processing. If we are to ignore small variation in segmentation, this renders agreement measures such as kappa even more inappropriate for sign language tokenisation and lemmatisation.[9]

Higher frame rates do reveal detail not visible in 50fps video, but do not lead to different segmentation in general.

Interestingly, annotators report that identifying the end of a sign is easier than to identifying the beginning. While this is a point in favour of Johnston's approach who just leaves out this step and thereby saves time in segmentation, the approach described here combines well with sub-sign phonetic encoding and will profit from automatic segmentation as introduced by Dicta-Sign.

## 7. Acknowledgements

This publication has been produced in the context of the joint research funding (DGS Corpus) 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.

The research leading to these results has also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 231135 (Dicta-Sign).

## 8. References

Brentari, D., Wilbur, R. (2008). A cross-linguistic study of word segmentation in three sign languages. In R. M. Quadros (ed.), Sign Languages: spinning and unraveling the past, present and future. TISLR9, forty five papers and three posters from the 9th. Theoretical Issues in Sign Language Research Conference, Florianopolis, Brazil, December 2006. Editora Arara Azul. Petrópolis/RJ. Brazil. Available online at http://www.editora-arara-azul.com.br/ebooks/catalogo/4.pdf

Crasborn, O., Zwitserlood, I. (2008). Annotation of the video data in the Corpus NGT. Nijmegen: Radboud University. Available online at http://www.ru.nl/aspx/download.aspx?File=/contents/pages/515436/corpusngt_annotationconventions.pdf

Hanke, T., Matthes, S., Regen, A., Storz, J., Worseck, S., Elliott, R., Glauert, J., Kennaway, R. (2011). Using Timing Information to Improve the Performance of Avatars. Second International Workshop on Sign Language Translation and Avatar Technology (SLTAT), Dundee, U.K., 23 October 2011. Available

[9] For a more promising approach, we refer the interested reader to Lücking et al. (to appear).

online at http://vhg.cmp.uea.ac.uk/demo/SLTAT2011Dundee/11.pdf

Johnson, R.E., Liddell, S.K. (2011). A Segmental Framework for Representing Signs Phonetically. Sign Language Studies 11(3), pp. 408-463.

Johnston, T. (2011). *Auslan Corpus Annotation Guidelines. 30. November 2011*. Available online at http://www.auslan.org.au/video/upload/attachments/AuslanCorpusAnnotationGuidelines30November2011.pdf

Liddell, S.K., Johnson, R.E. (1989). American Sign Language. The phonological base. In: Sign Language Studies 18, 64, pp. 195-277.

Lücking, A., Ptock, S., Bergmann, K. (to appear). Staccato: Segmentation Agreement Calculator according to Thomann. In: E. Efthimiou, G. Kouroupetroglou, S.-E. Fotinea (Eds.), Gesture and Sign Language in Human-Computer Interaction and Embodied Communication. Berlin/Heidelberg: Springer.

# Experiences Collecting Motion Capture Data on Continuous Signing

**Tommi Jantunen[1], Birgitta Burger[2], Danny De Weerdt[1], Irja Seilola[1], Tuija Wainio[1]**
[1]Department of Languages (Sign Language Centre)
[2]Department of Music (Finnish Centre of Excellence in Interdisciplinary Music Research)
P.O. Box 35, FI-40014 University of Jyväskylä, Finland
E-mail: {tommi.j.jantunen, birgitta.burger, danny.deweerdt, irja.r.seilola, tuija.wainio}@jyu.fi

**Abstract**

This paper describes some of the experiences the authors have had collecting continuous motion capture data on Finnish Sign Language in the motion capture laboratory of the Department of Music at the University of Jyväskylä, Finland. Monologue and dialogue data have been recorded with an eight-camera optical motion capture system by tracking, at a frame rate of 120 Hz, the three-dimensional locations of small ball-shaped reflective markers attached to the signer's hands, arms, head, and torso. The main question from the point of view of data recording concerns marker placement, while the main themes discussed concerning data processing include gap-filling (i.e. the process of interpolating the information of missing frames on the basis of surrounding frames) and the importing of data into ELAN for subsequent segmentation (e.g. into signs and sentences). The paper will also demonstrate how the authors have analyzed the continuous motion capture data from the kinematic perspective.

**Keywords:** motion capture, mocap, sign language, continuous signing, kinematic analysis

## 1. Introduction

The term *motion capture* (mocap) refers to the process in which a person's bodily movements are recorded and transformed into a digital format for further processing and analysis. The recording is normally done with infrared cameras that track the three-dimensional locations of reflective markers attached to the different parts of the person's body. The recording results in a numerical coordinate matrix that can be used as a source data for analysing the movements of the body and its parts from a kinematic perspective. Alternatively, the results of the recording can be used to build animated models of the moving person.

In sign language research, mocap data is generally considered to be the most accurate type of data available for signal-wise oriented, i.e. phonetic research. However, limitations in the availability and accessibility of the necessary technology have probably caused the number of studies taking advantage of it to remain relatively low. Examples of early studies exploiting mocap data are Wilbur (1990) and Wilcox (1992), who investigated stressed sign production and the kinematics of fingerspelling, respectively. More recent examples include Tyrone et al. (2010) and Duarte and Gibet (2010a). Of these, the former focused on variation in the hands' movements towards and away from the body, while the latter investigated variation in the kinematic characteristics of intersign transitions.

The data of most mocap studies into sign language have consisted of only relatively small sets of *isolated* expressions such as single signs (Wilbur, 1990), short fingerspelled sequences (Wilcox, 1992), or (carrier) phrases (Tyrone et al., 2010). The collection and exploitation of *continuous* mocap data, i.e. durationally longer discourse-type data, has been marginal (cf. Duarte & Gibet, 2010a). This is probably due to the fact that recording, processing, and analysing such data is extremely time consuming. However, such an endeavour is often worth the effort, mainly because of the inherent multifunctionality of such data. Continuous mocap data can be used not only in traditional sign-related phonetic studies (for an overview, see Duarte & Gibet 2010b) but, when accompanied with video, also as (supporting) corpora in studies that investigate sign language from various other, e.g. syntactic and discourse, perspectives.

The aim of this paper is to share some of the experiences the authors have had collecting continuous mocap data on Finnish Sign Language (FinSL) for the purpose of general phonetic and especially syntactic analysis. Our focus will be on issues that we consider to be crucial for the success of this type of mocap data collection, but which at the same time are also important for mocap-related work on sign languages in general. The topics covered include the issue of marker placement, the process of gap-filling, and the importation of mocap data into ELAN with video (Section 2).[1] We will also demonstrate how we performed kinematic analysis on our continuous mocap data (Section 3).

## 2. Collecting Continuous Mocap Data on Finnish Sign Language

We have been involved in collecting continuous mocap data on FinSL since the autumn of 2010. All the recordings have taken place in the motion capture laboratory of the Department of Music at the University of Jyväskylä, Finland. The laboratory hosts an eight-camera optical motion capture system (Qualisys ProReflex MCU120). The cameras have recorded the

[1] http://www.lat-mpi.eu/tools/elan/

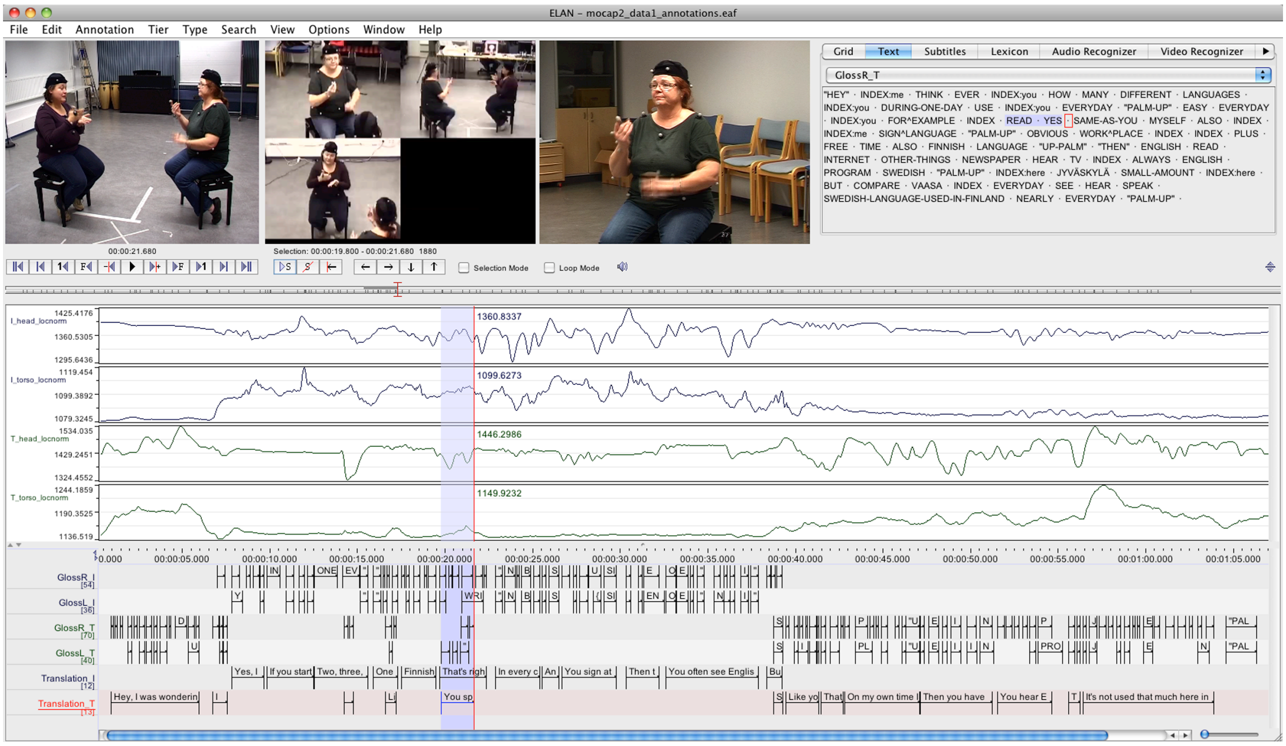


Figure 1: ELAN screenshot showing annotated video and motion capture data (lengths of location vectors describing the movement of the head and torso of both signers) recorded in a dialogue situation.

motion of the signer by tracking the three-dimensional locations of small ball-shaped reflective markers attached to the signer's hands, arms, head, and torso (see Section 2.1). The frame rate of the cameras has been 120 Hz, which is comparable to the frame rate of 100 Hz used in most modern sign language-related mocap work (e.g. Duarte & Gibet, 2010ab; Tyrone et al., 2010).

In addition to the mocap cameras, the laboratory also has a set of digital video cameras that are synchronizable with the motion capture system. In the data collection, the video cameras have recorded the signer from different angles and provided crucial supporting material for the later process of segmenting the quantitative mocap data into identifiable and processable chunks (e.g. into signs and sentences). In our work – as also in the work of Duarte & Gibet (2010ab) – the segmentation process has been done in ELAN, into which both the mocap data and the video data have been imported (see Figure 1). In general, ELAN has been a valuable tool for combining and controlling data obtained from conceptually different sources, and it also includes functions that allow the researcher to do simple numerical analyses with the data (Crasborn & Sloetjes, 2008). However, in our work, most of the actual analysis of the data has been done in Matlab using the *MoCap Toolbox* developed by Toiviainen & Burger (2011).[2]

In the following, we discuss some of the key issues in the data collection process. The discussion is carried out within two main themes that correspond to the two main phases of mocap data gathering: data recording (2.1) and data processing (2.2). The discussion is illustrated with examples from the continuous mocap data collected both in monologue and dialogue situations.

## 2.1 Data Recording

One of the most important questions in mocap data recording concerns marker positions: where to attach the reflective markers, and why? The issue is important because the location of markers affects their visibility in the system: covered markers are not recorded. Furthermore, markers that are placed inappropriately might make it difficult for the signer to properly articulate signs. Marker positions 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 (Toiviainen & Burger, 2011: 43, 46). Such processes are needed if one wishes to investigate, for example, the motion of the centroid of a certain joint or the kinetic energy of body parts.

Figure 2 shows the basic marker setup that we have used in our recordings. The total number of markers in the setup is twenty. The head has four markers at the level of the forehead; each arm and hand have seven markers in the main joint positions (the shoulder, elbow, ulnar and radial wrist joint, the most proximal joint of the index finger, and the tips of the index finger and thumb); and the upper torso has two markers, one in the middle of the chest (clavicle) and one on the back (C7).

[2] http://www.jyu.fi/music/coe/materials/mocaptoolbox/

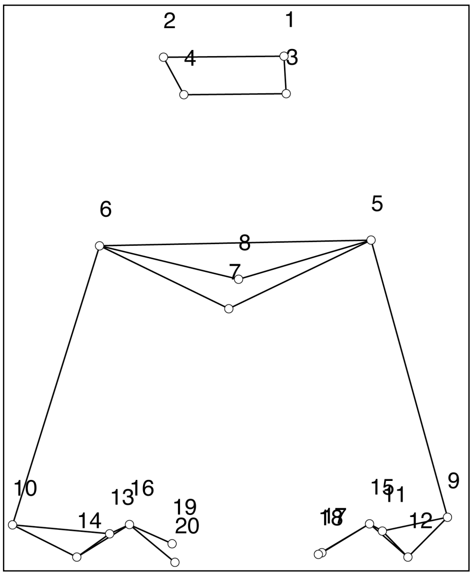


Figure 2: The basic marker setup in our mocap data showing basic marker connections and marker numbers.

The marker positions in our basic setup (Figure 2) have been decided so that the markers are maximally visible and identifiable to the system (our signers sit during the recording), and maximally processable (e.g. in data transformations), and so that they capture the main global movements of the hands, arms, upper torso, and head. The various local rotational movements of the wrist and index finger of both hands are also captured by the setup. The index finger has been preferred over other fingers because it is the finger that is most responsible for controlling and maintaining the rhythm and speed of signing (Ojala, 2011). The tip of the index finger has also been the reference point in other mocap-related studies (e.g. Wilbur, 1990; Wilcox, 1992).

We have deliberately wanted to keep the number of markers attached to the hand and fingers low because markers attached to these locations can easily impede the proper articulation of signs. This negative effect has been documented even with our present setup, which includes only three markers on the hand and fingers. Our signers have reported that especially the articulation of signs involving contact of the index finger with the body has occasionally been unnatural.

In comparison with other modern mocap studies, the total number of markers in our basic setup is relatively low: Tyrone et al. (2010) used thirty markers (7 on each arm, 7 on the head, and 9 on the torso) and Duarte & Gibet (2010b) – whose additional goal is to use the data to create animated signing figures, avatars – employed ninety-eight markers (43 facial markers, 43 body markers, and 12 hand markers, with 6 on each hand). The main difference between our basic setup and these other setups lies in the number of torso and head markers. In the recording of our dialogue data, we have experimented with adding more markers precisely in these areas. An example of an extended marker setup in which the number of markers per signer is thirty-one is illustrated in Figure 3.

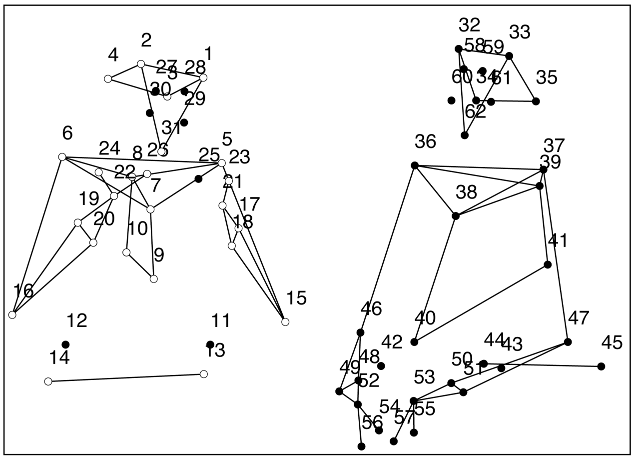


Figure 3: Our extended marker setup for dialogue (see Figure 1) showing the most important marker connections and marker numbers.

In this extended setup (Figure 3), the number of markers has been increased by five on the facial area (both brows, both cheeks, and the chin) and by six on the lower torso area (abdomen, T10, and altogether 4 hip markers), the rest of the marker locations corresponding to those in the basic setup. The main advantage of this extended setup is that, while keeping the data comparable with those recorded with the basic setup, it also captures the rigid lower torso movements. Also the movements of the head are now captured in more detail.

However, in general, the capturing of facial movements with the setup presented in Figure 3 proved not to be very succesful, as the facial markers did not remain visible to the system all the time. This was probably caused by the relatively small size of the reflective sticker tapes that we had to use on the face in place of the ball-shaped markers; the markers were easily covered by the hands articulating on the facial area. Also the markers attached to the lower torso area were not always picked up by the system. We suspect that this was caused by the fact that the signers were sitting during the recording and occasionally their shirts covered especially the hip markers. In the future, the obstruction of markers can be avoided by asking the signers to wear "mocap jackets" that are made from stiff fabric and thus keep the marker positions maximally visible.

In general, our experience is that a higher number of markers does not automatically produce better data. However, the choice of the number of markers is ultimately dictated by the specific goals of the mocap recording (cf. animation in Duarte & Gibet, 2010b). For the purpose of collecting continuous mocap data on FinSL for general kinematic use, we have found that our basic marker setup (Figure 2) supplemented by abdomen

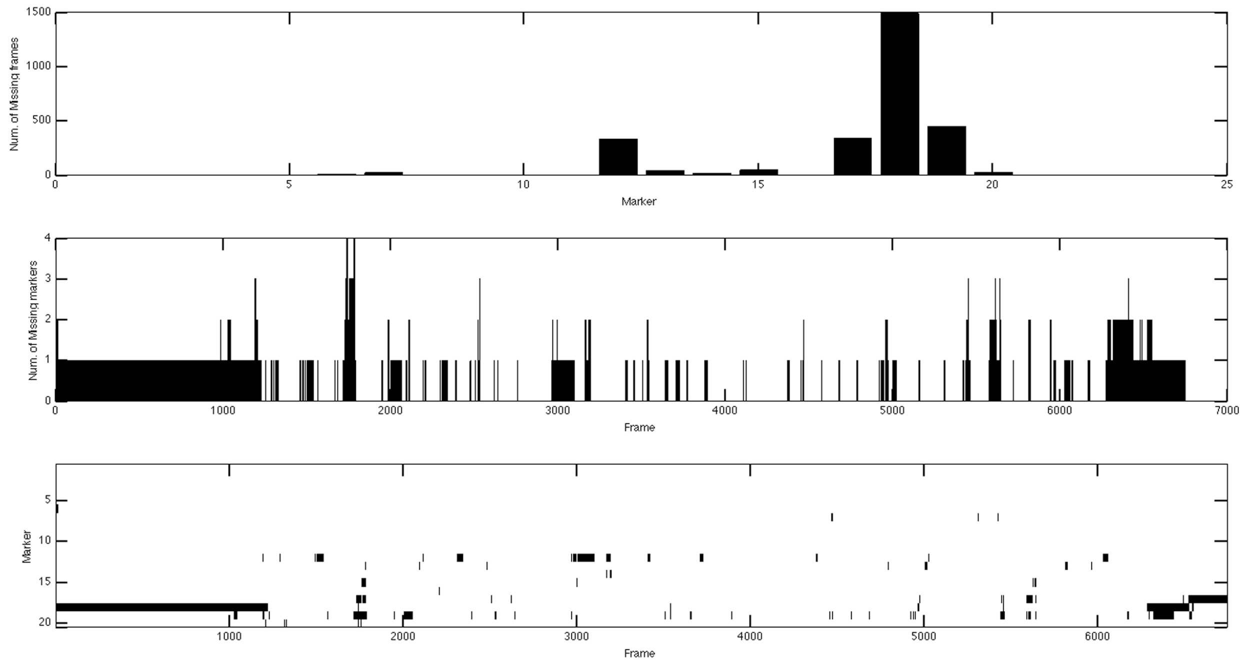



Figure 4: Three diagrams showing the original gaps in the 50 second-long monologue data. The top diagram illustrates the number of missing frames per marker; the middle diagram illustrates the number of missing markers per frame; and the bottom diagram illustrates for which frame(s) which markers were not recorded.

and T10 markers (see Figure 3) is sufficient, both from the point of view of the movements that it captures and the processing it allows.

## 2.2 Processing of the Data

### 2.2.1. Gap-filling

After the mocap data has been recorded, there are several essential steps that one has to take with it in order to make it into an analysable format. One important step is the process of gap-filling that takes place after the marker locations in the data have been assigned distinctive identities (i.e. the markers have been labelled). Gap-filling means searching for the empty frames that almost always occur during the recording and interpolating the missing data on the basis of the information in the surrounding frames. Normally this is a fairly automatic and reliable operation but it may also produce false results (e.g. when the gap is relatively long) which the researcher needs to take into account when assessing the validity of the results.

Figure 4 shows the original gaps in the approximately 50 second-long monologue data used in Section 3 to demonstrate how continuous mocap data in general can be used in kinematic analyses. The diagrams in the figure are the plots created from the output of the function *mcfillgaps* of the Matlab-based *MoCap Toolbox* used in the analysis of the data. The diagrams show that there have been slight problems in the visibility of markers attached to both thumbs (markers number 17 and 19 in the top diagram) and the one attached to the ulnar side of the wrist of the nondominant hand (number 12). More serious visibility problems have occurred with the nondominant hand index finger marker (number 18), especially immediately after the beginning of the recording and at the end of the recording (see the bottom diagram); at the beginning the first three frames were recorded (this cannot be seen from the bottom diagram of Figure 4 because of the scaling) but then there is a gap of about 1200 frames. The gap resulted from the fact that the hands and fingers were turned in such a way that the system could not see the markers.

The gap-filling algorithm of the *mcgapfill* function is able to successfully interpolate the missing data for most gaps shown in Figure 4 because of their relatively short duration. However, the longer duration of the gap of the nondominant hand index finger marker (18) at the beginning of the data cannot be handled properly by the default use of the gap-filling algorithm. This is demonstrated in Figure 5, which shows the locations and connections of the markers in frame number 1132 of the data. This frame occurs a few moments before the end (frame 1209) of the long initial gap of marker 18, i.e. just before the moment the nondominant hand starts to move up from the resting position towards the place of articulation of the FinSL sign WINTER. The linear interpolation of the gap between frames 3 and 1209 results in a slow and regular rising of the nondominant hand index finger marker to reach the position in the upper torso area when the marker was detected again. In order to overcome this problem, we used the *maxfill* parameter of the *mcfillgaps* function of the *MoCap Toolbox*. The parameter specifies the maximal length of gaps to be filled; longer gaps are not processed (see Toiviainen & Burger, 2011: 76). A more accurate though more time consuming way would be not to fill such gaps linearly, but to take the surrounding markers into account; in this case, the nondominant hand index finger marker would only start moving when the other hand/arm markers move and adapt its speed accordingly.

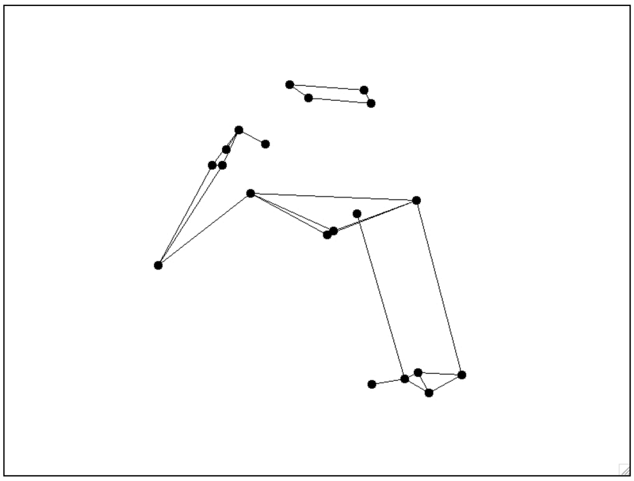


Figure 5: Unsuccessful gap-filling process.

As the gap-filling process alters the data and this can lead to undesired outcomes, it is sometimes tempting not to apply it to the data at all. However, in order to make continuous mocap data maximally analysable from the kinematic perspective, some gap-filling is normally required. The important thing is to check the outcome of the gap-filling process. The most convenient way that we have found to do this has been to create an animated stick figure model of the signer(s) on the basis of the processed data (cf. Figures 2, 3, and 5). These types of animations are easily constructed with the functions of the *MoCap Toolbox*. The animations also have other advantages. For example, our experience is that they are the easiest way to see whether marker identities are assigned correctly and markers are tracked properly, even before the gap-filling procedure. Particularly from the perspective of sign language-related mocap work, stick figure animations are also the best way to ensure that the recorded numerical data actually represents activity that is identifiable as signing, i.e. makes sense linguistically.

#### 2.2.2. Importing Mocap Data into ELAN

In order to make the continuous mocap data usable for linguistic analyses, it needs to be imported with the video material into a data management program in which it can be segmented into more processable chunks. In our work, we have managed the data in ELAN. The process we have used in importing the data into ELAN involves several steps. As there are no standard guidelines for this type of work, we will now describe these steps.

The first step is the cropping of the data. The mocap recording with our system results in a three-dimensional numerical coordinate matrix in .tsv format. With our basic setup of twenty markers, this matrix consists of sixty columns, i.e. three for each recorded marker (x, y, and z dimensions). However, ELAN (ver. 4.1.2) is able to process additional data files that include a maximum of twenty columns of numerical information (and of these at least one column must include timecodes). Consequently, in order to make the mocap data importable into ELAN, the data first needs to be cropped, i.e. unwanted or otherwise redundant marker columns need to be removed from the matrix.

In our work, we did the cropping by opening the original (gap-filled) matrix in Matlab and copy-pasting the desired columns of marker coordinates onto an empty Excel spreadsheet. Before the columns are copied into Excel, their information can be further processed in Matlab with the *MoCap Toolbox*. An example of some simple processing that we have normally done for the data at this point is the calculation of velocity and acceleration vectors and their Euclidean Norms (i.e. magnitudes, or lengths) for the marker location data. When all the relevant columns are copied onto the Excel sheet, one gets a reduced yet also augmented version of the original data that can contain information, for example, on the three-dimensional locations of the tip of index finger and chest (C7) markers (3+3 columns) as well as on the velocity and acceleration of these markers in all three dimensions (2x(3+3) columns).

The second step in the process of importing continuous mocap data into ELAN is the generation and addition of timecodes and frame numbers onto the Excel sheet containing the cropped (and usually augmented) data. In order for any additional data file to be processable in ELAN, it must include the timecode information in one column. Such information – or frame number information – is not automatically exported by our mocap system and we have used a specific JavaScript code to generate it. Once the incremental timecode and frame number information is generated for all the frames of the data, it is added into the first two columns of the Excel sheet. Note that because timecodes and frame numbers require two columns on the Excel sheet, the maximum number of mocap data columns that can be copied onto the sheet from Matlab is eighteen.

After the Excel sheet containing timecodes, frame numbers and the relevant mocap data is completed, it is saved as a file in .csv format. This is the format ELAN uses to process additional data files.

The third and final step in getting the mocap data into ELAN is the actual data import process. First, the video recorded with the mocap data is imported; the video has been synchronised with the mocap data by our mocap recording system but we have found that minor editing (cropping) work with the video is still often required to make its length correspond to that of the mocap data. The primary video used in ELAN is added through the normal process of creating a new ELAN annotation file. Additional videos (as in Figure 1) may be imported through ELAN's *Linked Files* function (the Linked Media Files tab), found in the *Edit* menu.

The .csv data file created in Excel is also imported into ELAN through the *Linked Files* function (the Linked Secondary Files tab). Note that the number of files to be added is not limited to one. The twenty-column limitation of one file is thus compensated for here with the possibility of working with several twenty-column .csv files.

After the addition of data file(s), the columns containing the numerical information in the file(s) need to be configured for ELAN. This is done by control clicking anywhere in ELAN's Timeseries Viewer, which contains the (still empty) trackpanel(s) and, from the menu that appears, choosing the *Configure Tracks* option. This opens up a dialogue box that displays the maximum of twenty data columns included in the .csv file and several options of how they can be configured to be shown as tracks (i.e. linegraphs) in the trackpanels of ELAN's main screen. In addition to specifying the data in columns, it is crucial to define the timecode column that ELAN uses to synchronise the data with video(s) and annotations.

An example of the end result of the import procedure is presented in Figure 1. The figure also shows how the data has been segmented into signs by following the annotation layout of the "Corpus NGT" (Crasborn & Zwitserlood, 2008). The completed annotation makes it possible to use ELAN's *Extract Data* function (accessible through control clicking the Timeseries Viewer) to automatically generate annotation cells corresponding to the durations of signs and to display the initial and final frame number of each sign in these cells on the basis of the information in the underlying .csv file. This frame number information is needed for successful analysis of specific signs and their sequences (e.g. sentences) in Matlab with the *MoCap Toolbox*.

## 3. Analysing the Data

Continuous mocap data can be used for a variety of purposes. The most straightforward use of the data is to exploit it to support the annotation of sign language corpora. The information concerning the motion of the hands visualised in ELAN has undeniable value for the segmentation of continuous signing into signs and other linguistic units. The changes in the direction of the movement of the hands and other articulators such as the head and torso, which often mark linguistic boundaries, can be hard to notice by looking only at the video, but they are easily detected by looking at the graphs representing the changes in the three-dimensional locations of the markers.

However, the real value of the continuous mocap data lies in its use for the kinematic analysis of signing and the linguistic units contained in it (e.g. signs and sentences). In the following, we give examples of these types of analyses with one set of our monologue data. The data comprises a story lasting about 50 seconds in FinSL describing a wintertime cycling incident near Jyväskylä University. The data has been recorded with our basic marker setup with twenty markers.

### 3.1 Analyses Based on Location Data

In our ongoing work, we have used the three-dimensional marker location data to calculate the cumulative distances travelled by different markers during the production of different FinSL sentences (with the function *mccumdist* in the *MoCap Toolbox*; the focusing on these types of specific sequences in the continuous data has been enabled by the frame number information extracted in ELAN, see 2.2.2). Figure 6 illustrates the result of such a calculation for one FinSL sentence. The markers involved in the calculation were the dominant hand index finger tip marker and the front right head marker.

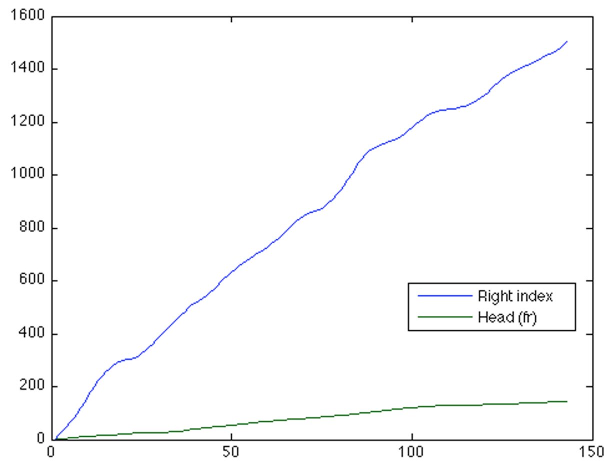



Figure 6: The cumulative distance travelled by the dominant hand (upper line in blue) and the head (lower line in green) in FinSL sentence ME HAVE-TO GO-TO UNIVERSITY 'I had to go to the university'. The distance is measured in millimeters per frame.

The diagram in Figure 6 shows that in the production of this particular sentence the tip of the index finger travelled a distance of about 1.5 meters. In the same amount of time, the distance travelled by the head was only about 0.2 meters. The difference is predictive for declarative FinSL sentences in general.

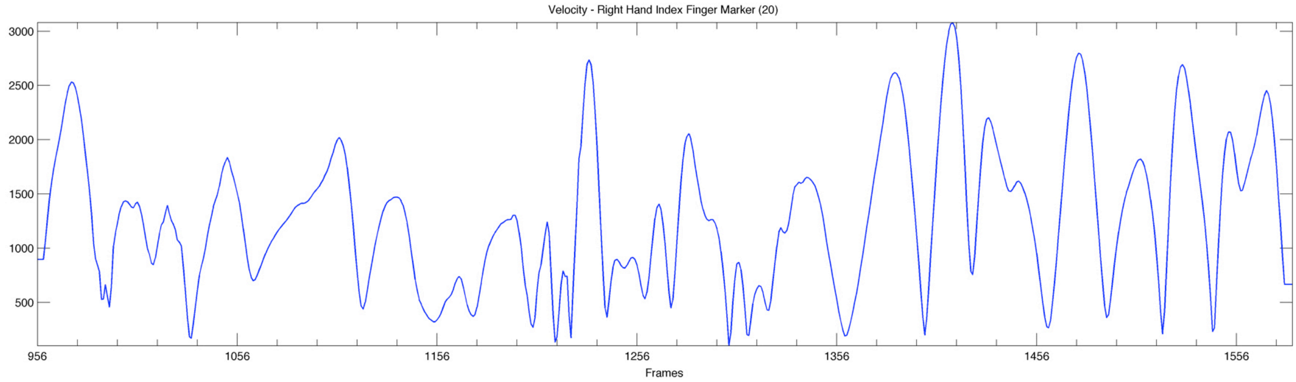



Figure 7: Velocity magnitude plot for the sequence of the first nine signs in the present monologue data. Velocity magnitude is measured in meters per second.

More generally, we have also used the marker location data to study correlations between movements produced with different articulators. The correlations have been calculated on the basis of both the three-dimensional location data and the Norm data, representing the variation in the length of the location vectors of markers (cf. Figure 1). Table 1 summarises some results of this work for Norm data on the monologue used in the present study. The articulators investigated are the dominant hand (operationalized as the centroid of the dominant hand ulnar and radial wrist marker), the head (the centroid of head markers), and the upper torso (the centroid of clavicle and C7 marker).

| *Articulators* | *R* | *Interpretation* |
|---|---|---|
| wrist-head | 0.217 | weak |
| wrist-torso | 0.520 | strong |
| head-torso | 0.764 | very strong |

Table 1: Correlations in the motion of three articulators in the monologue Norm data.

The results show that the motion of the hand follows the motion of the upper torso (correlation co-efficient=0.520) but not that of the head (0.217). The motion of the head closely follows the motion of the upper torso (0.764). The interplay of the articulators is largely explained by the anatomy and physiology of the human body.

### 3.2 Analyses of Velocity and Acceleration

We have also used our continuous mocap data to investigate the velocity and acceleration characteristics of signs and sentences. For this purpose, we have applied especially the Euclidean Norms of velocity and acceleration vectors calculated on the basis of the three-dimensional marker data (with *MoCap Toolbox* functions *mctimeder* and *mcnorm*). Some results of this investigation are shown in Figure 7, which presents the magnitude of the velocity (i.e. speed) of the index finger tip marker as a function of time during the first nine signs of the present data.

Figure 7 shows that the speed of the tip of the dominant hand index finger marker varies considerably in continuous signing. In general, moments of slowest speed in the plot are identified fairly accurately with the borders of sign strokes (Kita et al., 1998). The moments of highest speed, on the other hand, associate either with the middle phases of strokes or with transitions.

### 3.3 Analysis of Rhythm

Examples of more complex analyses with the continuous mocap data include analyses of the inherent rhythm of the motion of different articulators. In our work, we have focused especially on the rhythm of head movements in FinSL sentences. In our investigation of this phenomenon, we have defined the notion of rhythm as regularity and predictability in motion. From this perspective, we have used autocorrelation to study the periodicity and aperiodicity of head movements (with the *mcperiod* function in the *MoCap Toolbox*). The three diagrams in Figure 8 show some of the results of this investigation.

In Figure 8, the diagram of two consecutive transitive clauses illustrates how in these types of clauses the head normally moves from side to side in a fairly periodic manner with relatively low amplitude. This class of side-to-side head movements contrasts with those typically found in FinSL topic-comment structures (Jantunen, 2008) and in negative expressions. In topic-comment structures, the head movement is aperiodic, the break in the regularity of rhythm being caused by the tendency to keep the topic prosodically detached from the following comment. In negative expressions the movement of the head is again periodic. However, in comparison to prototypical transitive clauses, the amplitude of the side-to-side head movements in negative expressions is higher.

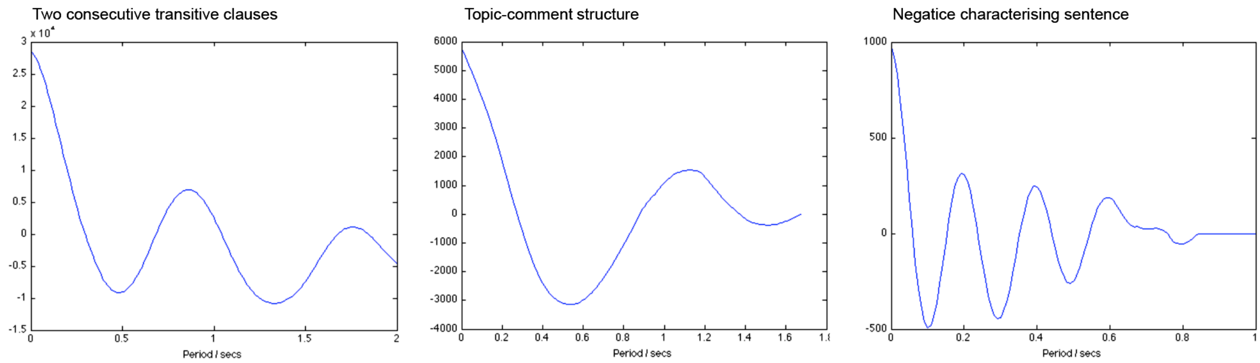



Figure 8: Descriptors of the autocorrelation function illustrating the periodicity of sideways head movement in three types of prototypical FinSL sentences.

## 4. Conclusion

This paper has described our experiences in collecting continuous mocap data on FinSL. The description has focused on several key issues in mocap data recording (marker placement), processing (gap-filling, importing data into ELAN) and analysis (the kinematic analysis of signs and sentences) and shown that mocap data collection is a complex process involving several steps and requiring expertise in different scientific fields. In the future, we would like to see more researchers collect more continuous mocap data and use it in sign language studies on all aspects of linguistics. Our own plans for the future include continuing both our data collection and the analyses demonstrated in the paper. We will also examine the possibility of adding our continuous mocap data to the FinSL corpus, preparations for which are currently being made.

## 5. Acknowledgements

The authors wish to thank Eleanor Underwood for checking the English of the paper. The financial support of the Academy of Finland under grants 134433 (Signs, Syllables, and Sentences; 3BatS) and 118616 & 141106 (Finnish Centre of Excellence in Interdisciplinary Music Research) is gratefully acknowledged.

## 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 at LREC 2008*. Paris: ELRA, pp. 39-43.

Crasborn, O. & Zwitserlood, I. (2008). Annotation of the Video Data in the "Corpus NGT". Dept. of Linguistics & Centre for Language Studies, Radboud University Nijmegen, The Netherlands. Online publ. http://hdl.handle.net/1839/00-0000-0000-000A-3F63-4.

Duarte, K. & Gibet, S. (2010a). Reading Between the Signs: How are Transitions Built in Signed Languages? Paper presented at TISLR 10, West Lafayette, IN, USA, Sept. 2010.

Duarte, K. & Gibet, S. (2010b). Heterogeneous Data Sources for Signed Language Analysis and Synthesis: The SignCom Project. In *Proceedings of the LREC 2010*. Paris: ELRA, pp. 461-468.

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, pp. 83-123.

Kita, S., van Gijn, I. & van der Hulst, H. (1998). Movement Phases in Signs and Co-speech Gestures, and Their Transcription by Human Coders. In I. Wachsmuth, M. Froelich (Eds.), *Gesture and Sign Language in Human-computer Interaction: Proceedings of International Gesture Workshop*. Berlin: Springer, pp. 23-35.

Ojala, S. (2011). *Towards an Integrative Information Society: Studies on Individuality in Speech and Sign.* TUCS Dissertations No 135. University of Turku, Finland.

Toiviainen, P. & Burger, B. (2011). *MoCap Toolbox Manual*. University of Jyväskylä, Finland.

Tyrone, M., Nam, H., Saltzman, E., Mathur, G. & Goldstein, L. (2010). Prosody and Movement in American Sign Language: A Task-Dynamics Approach. *Speech Prosody 2010*, 100957, pp. 1-4.

Wilbur, R. (1990). An Experimental Investigation of Stressed Sign Production. *International Journal of Sign Language*, 1, pp. 41-59.

Wilcox, S. (1992). *The Phonetics of Fingerspelling*. Amsterdam: John Benjamins.

# Towards Russian Sign Language Synthesizer: Lexical Level

**Alexey A. Karpov[1] and Miloš Železný[2]**

[1] St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Russia

[2] The University of West Bohemia (UWB), Pilsen, Czech Republic

E-mail: karpov@iias.spb.su, zelezny@kky.zcu.cz

**Abstract**

In this paper, we present a survey of existing Russian sign language electronic and printed resources and dictionaries. The problem of differences in dialects of Russian sign language used in various local communities of Russia and some other CIS countries is discussed in the paper. Also the first version of a computer system for synthesis of elements of Russian sign language (signed Russian and fingerspelling) is presented in the given paper. It is a universal multi-modal synthesizer both for Russian spoken language and signed Russian that is based on a model of animated 3D signing avatar. The proposed system inputs data in the text form and converts them into the audio-visual modality, synchronizing visual manual gestures and articulation with audio speech signal. Generated audio-visual signed Russian speech and spoken language is a fusion of dynamic gestures shown by the avatar's both hands, lip movements articulating words and auditory speech, so the multimodal output is available both for the deaf and hearing-able people.

**Keywords:** Russian sign language, Russian signed speech, fingerspelling, sign language resources, computer synthesizer

## 1. Introduction

At present, sign languages are national languages of human-to-human communication in the USA, Finland, Czech Republic, etc. In the Russian Federation, official status of the sign language is lower and it is not a national communication means yet.

One of the major problems connected with the Russian sign language (RSL), which is regularly used by 1.5-2 million deaf people both in Russia and in some CIS countries (Ukraine, Belarus, Kazakhstan, etc.), consists in geographical vast of the country and existence of various dialects of RSL, for example, in Moscow, St. Petersburg, Novosibirsk, Vladivostok, Minsk, etc. Differences in dialects originate from a history of earliest schools for the deaf: the first one was organized in Pavlovsk near St. Petersburg in early XIX century (it still works) with teachers from France, and the first school in Moscow was opened much later and teachers were from Germany.

It is clear that differences in various versions of the sign language cause high ambiguities for translation and interpretation. An expert analysis shows that less than half of signs for the same words are similar in different dialects of RSL, these are simple manual gestures showing real objects in the field of vision (for example, "I", "YOU", "HEAD", "NOSE", etc.), however, most of abstract concepts highly depend on practice and traditions of local communities.

## 2. Russian Sign Language Resources

There exist several multimedia electronic dictionaries of RSL (some of them are available on-line and the rest are distributed by media):

1) "Thematic dictionary of the Russian sign language" developed by the Moscow organization of All-Russian society of the deaf in 2006 (www.deafmos.ru/info.phtml?c=24&id=1059), it has above 3K commonly used gestures on 4 DVDs, several human demonstrators are presented (Figure 1a).
2) "Russian Sign Language Explanatory Dictionary RuSLED" has over 2.5K gestures with etymology of the signs (Voskresenskiy, Gulenko & Khakhalin, 2009) and contains video data recorded in 2002 by the Inter-regional Rehabilitation Center for deaf people in Pavlovsk (Figure 1b).
3) Electronic learning system "Russian sign language. Basic course" (http://istina.inion.ru/NIOT/rgy.htm) created by the "Truth Center" in Moscow in 2001, it has up to 2K gestures (Figure 1c).
4) Interactive on-line dictionary DigitGestus (www.digitgestus.com) was collected in the Novosibirsk region in the late 90s.
5) New database recorded by the Novosibirsk State Technical University in 2010-2011 with above 3.2K signs (Grif et al., 2011a). It has a good quality of video data and one demonstrator (Figure 1d).
6) On-line RSL dictionary of the European project "Spreadthesign" (www.spreadthesign.com/ru) for 15 different sign languages of the world, it includes over 4.5K gestures of RSL (Figure 1e).
7) On-line RSL dictionary collected by the Stanford University (www.stanford.edu/dept/lc/rsl/). It has some hundreds of signs recorded in the USA.
8) On-line dictionary "Surdoserver" (http://surdoserver.ru), it has the same video data as in (3) plus an additional dictionary on Information Technology topic, it also has a version for mobile devices.
9) On-line dictionary "Surdoportal" (http://vorb.ru/ps/) has a small-sized dictionary for several technical topics.

Among the printed illustrated dictionaries of RSL, the following ones should be mentioned: books authored by Geilman (1979; 2001) from St. Petersburg (see Figure 1f), written by Zaitseva (2000), Fradkina (2001) from Moscow and by Dimskis (2002) from Minsk, etc.

(a) (b) (c) (d) (e) (f)

Figure 1: Variety of signs for the word "ONE" (digit) in different dictionaries of RSL

These electronic and printed dictionaries represent lexicon of RSL and illustrate differences in RSL dialects. An expert analysis of these electronic and printed dictionaries shows that only 30-40% gestures are similar in the Moscow and Petersburg dialects, though in some cases gestures can be understood, the same situation in other regions as well. Even in the Moscow region there are some RSL sub-versions in local communities and dictionaries are not unified and normalized, for example, the sign shown in Figure 1c (Figure presents various realizations of signs for the word "ONE") taken from the electronic system "Russian sign language. Basic course", created under support of the All-Russian society of the deaf in Moscow, differs from the sign for this word from other dictionaries originated from the Moscow region. So, it is not possible to say about one normalized RSL and unified automatic system for RSL analysis and synthesis.

Grammatical structure of RSL is not sufficiently studied and formalized yet to say on fully automatic text-to-sign language translation. Last years, some studies on structure of RSL grammar are made by linguists in the Moscow State University (Kibrik & Prozorova, 2007; Prozorova, 2009), in the Novosibirsk State Technical University (Grif & Demyanenko, 2011b), as well as by some other researchers working inside and outside of Russia (Voskresenskiy, Gulenko & Khakhalin, 2009; Kimmelman, 2009a; Kimmelman, 2009b; Mjasoedova & Filippovich, 2010). Unfortunately, the current scientific level and essential differences in semantic-syntactic structure of written/spoken and sign languages do not allow to perform machine translation from Russian texts to Russian sign language and there are no any models for automatic translation yet. In order to create such a model, it is required to use "deep" semantic, pragmatic and situational analysis and parsing of written phrases, however, at present there exist only superficial semantic analyzers because of imperfection of algorithms, databases and ontology for Russian and general complexity in grammatical and morphological structure of the language.

Our main goal is to develop a computer sign language synthesis system for the St. Petersburg version of RSL. At the given stage of the research we have developed a model for computer synthesis of signed Russian and fingerspelling, where input text processing is much easier.

## 3. Computer Synthesizer

There is quite large community aiming at computer processing of sign languages in Europe organized around several European projects. However, there is a lack of computer systems for RSL processing including sign language analysis and synthesis. Some models for signed Russian synthesis have been developed in Novosibirsk (www.vesti.ru/doc.html?id=358385&p=35&cid=1) and in Minsk (http://ont.by/news/our_news/0062482). Both are based on compilation of video fragments for whole phrases from pre-recorded video databases with one human demonstrator. However, an essential disadvantage of such systems is that produced video stream cannot show continuous signing and it is only isolated signs synthesis, because all the lexicon items are video fragments, where a demonstrator shows each sign independently from other signs and every gesture starts and ends with a neutral position of the hands (usually both hands are in an initial position below the belt). In the case of real continuous signing people do not go through this initial position after each gesture, but only in the beginning and the end of the whole phrase. So, video dictionaries cannot serve as the basis for construction of continuous sign language and signed speech synthesizers for human-computer interaction.

Therefore, animated characters or virtual humans (avatars) are more adequate for this task. A database of animated 3D gestures allows compiling phrases in a sign language in the continuous manner keeping smooth transitions from sign to sign.

Since 2009, SPIIRAS Institute and the University of West Bohemia have been developing a multimodal text-to-sign language system for RSL. It is originally based on a 3D signing avatar for signed Czech speech and language (Krňoul et al., 2008), parameters of which and hand movements are controlled by the codes of Hamburg Notation System (Prillwitz et al., 1989; Hanke, 2004). The proposed system inputs data in the text form and converts them into the audio-visual modality, combining visual manual gestures and articulation with audio-visual speech, so the multimodal output is available both for deaf and hearing-able people. The main software components of the multimodal system are:

1) Text processing module that inputs and processes text phases and generates word labels, phonemic and visemic transcriptions, as well as a stream of inner codes of hands movements from the dictionary.
2) Text-to-speech system for spoken Russian that generates auditory speech signal corresponding to the text (Hoffmann et al., 2007).
3) Virtual 3D model of human's head with controlled lips articulation, mimics and facial expressions (Zelezny et al., 2006).
4) Bimodal audio-visual "talking head" that integrates the speech synthesizer and the virtual head model and synchronizes lip movements with synthesized auditory speech signal taking into account natural asynchrony between audio and visual speech modalities (Karpov et al., 2009).
5) Virtual 3D model of human's upper body (Kanis & Krňoul, 2008) with both hands, movements of which are controlled by the codes of HamNoSys notation.
6) Multimodal computer synthesizer that synchronizes and integrates all the components for automatic generation of auditory speech, visual speech (articulation) and manual gestures of signed speech and language.

Figure 2 shows general architecture of the multimodal system, its main components and interaction between them. In the proposed system, synchronization of audio-visual speech with manual signs is controlled using time stamps of start and end of spoken words generated by the auditory speech synthesizer. Since natural speech has a higher tempo than the corresponding manual gestures, then the signing avatar speaks and articulates isolated spoken words and the system waits for the following acoustic word until completion of the current gesticulation (if no any sign for a word in the system's dictionary it is spelled as a sequence of finger sings by the avatar's right hand). By this way continuous gesticulation of the whole phase is provided.

The sign language synthesizer based on high-quality virtual 3D avatars has a lot of merits:

- It allows a user to see generated visual data from different sides and viewing angles that leads to better understanding of spatial information, for example distance between the hands and the body or hands each from other.
- It is possible to add new items into the dictionary quite easily; it is an animated virtual human, so there is no requirement to record one human demonstrator in the same dress, haircut and make-up with similar lighting conditions and equipment.
- It can produce a continuous stream of visual gestures without transitions through a neutral position of hands and there are no seen borders between adjacent signs.
- It is possible to change one virtual avatar to another one and to create new models of human beings.
- It is able to show synthesized signed phrases on a screen with any required speed, slowing down or speeding up the visual stream.

The multimodal system is aimed not only for deaf, deaf-mute and hearing impaired people, but is useful for hearing people as well. It is a universal multimodal computer system for synthesis both Russian spoken language (audio-visual modality) and the sign language (visual modality). Generated audio-visual signed Russian speech and language is a fusion of dynamic gestures shown by the avatar's both hands (or only by the right hand in the case of Russian fingerspelling), lip movements articulating spoken words and acoustic speech. Many deaf people are able to read speech by lips and to understand phrases even without manual gestures. Acoustic spoken language is a natural speech modality for communication with hearing-able people. Avatar's lips articulation synchronized with audio stream helps to improve both intelligibility and naturalness of generated speech.

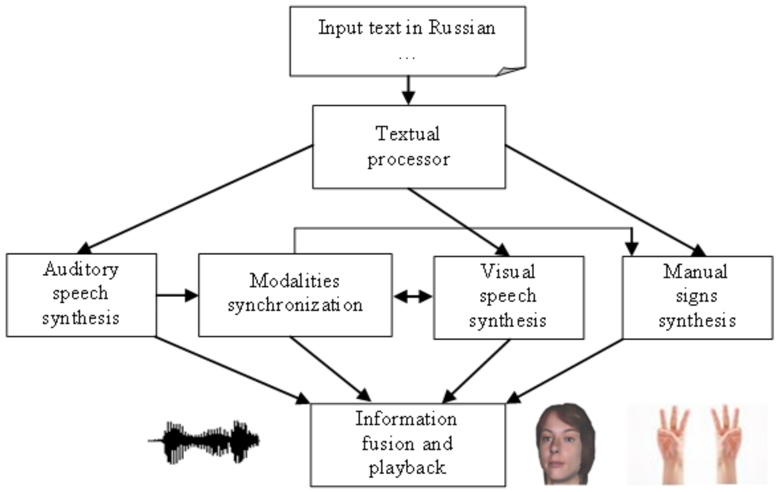



Figure 2: Architecture of the multimodal synthesizer for audio-visual spoken language and signed Russian

## 4. Conclusions

We presented the survey of RSL resources and the first version of the text-to-sign synthesizer for signed Russian and fingerspelling. It is the universal multimodal system for synthesis of Russian spoken language (audio-visual modality) and signed Russian (visual modality) aimed both for the deaf and hearing-able people.

Demonstrations of the multimodal synthesis system for Russian fingerspelling and elements of RSL are available on-line: www.spiiras.nw.ru/speech/demo/daktilrus.avi and www.spiiras.nw.ru/speech/demo/signlang.avi.

Qualitative user evaluation of the system was made in the end of 2011 with the help of some representatives of the All-Russian society of the deaf in St. Petersburg. They said on novelty and urgency of the system and positively estimated intelligibility and naturalness of lips articulation of the talking head and recognizability of manual gestures of the virtual avatar. At the same time they expressed some requirements for the future work, i.e. to use lexical items of RSL from the books and dictionaries (i.e., prepared by Gejlman) created in St. Petersburg region only and to enforce further research on text-to-sign language machine translation.

## 5. Acknowledgements

We are very grateful to Dr. Zdeněk Krňoul from the University of West Bohemia in Pilsen for provision of the 3D signing avatar and its parameters adjustment, as well as to Olga Novoselova, who is the head of the All-Russian society of the deaf in St. Petersburg, and her colleagues for assistance in evaluation of the synthesizer and suggestions on system's improvement.

This research is supported by the Ministry of Education and Science of Russia in the framework of the Federal Target Programme «R&D in Priority Fields of S&T Complex of Russia for 2007-2013» (contract No. 11.519.11.4025), by the grant of the President of Russia (project No. MK-1880.2012.8), and by the Ministry of Education of the Czech Republic, project No. ME08106.

## 6. References

Dimskis, L.S. (2002). *Learning sign language* (Изучаем жестовый язык), Moscow: Academia, 128 p.

Fradkina, R.N. (2001). *Speaking hands: Thematic dictionary of sign language of the deaf of Russia* (Говорящие руки: Тематический словарь жестового языка глухих России), Moscow, 598 p.

Gejlman, I.F. (1979). *Specific means of communication of the deaf. Fingerspelling and mimics* (Специфические средства общения глухих. Дактилология и мимика). 4 volumes. Leningrad.

Gejlman, I.F. (2001). *Get to know: Manual speech* (Знакомьтесь: Ручная речь), Moscow: Zagrej, 172 p.

Grif, M.G., Demyanenko, Y.A., Korolkova, O.O., Tsoy, Y.B. (2011a). Development of Computer Sign Language Translation Technology for Deaf People. In *Proc. IFOST-2011*, Harbin, China, pp. 674--677.

Grif, M.G., Demyanenko, Y.A. (2011b). Russian text analysis in a system of text to Russian sign language translation (Анализ русского текста в системе компьютерного сурдоперевода на русский разговорный жестовый язык). *NSTU Proceedings*. Novosibirsk, Russia, 63 (1), pp. 83--90.

Hanke, T. (2004). HamNoSys - representing sign language data in language resources and language processing contexts. In *Proc. LREC-2004*, Lisbon, Portugal, pp. 1--6.

Hoffmann, R., Jokisch, O., Lobanov, B., Tsirulnik, L., Shpilewsky, E., Piurkowska, B., Ronzhin, A., Karpov, A. (2007). Slavonic TTS and SST Conversion for Let's Fly Dialogue System. In *Proc. SPECOM-2007*, Moscow, Russia, pp. 729--733.

Kanis, J., Krňoul, Z. (2008). Interactive HamNoSys Notation Editor for Signed Speech Annotation. In *Proc. LREC-2008*, Morocco, pp. 88--93

Karpov, A., Tsirulnik, L., Krnoul, Z., Ronzhin, A., Lobanov, B., Zelezny, M. (2009). Audio-Visual Speech Asynchrony Modeling in a Talking Head. In *Proc. Interspeech-2009*, Brighton, UK, pp. 2911--2914

Kibrik, A.A., Prozorova E.V. (2007). Referencial choice in Russian sign language (Референциальный выбор в русском жестовом языке). In *Proc. Dialog-2007*, Moscow, Russia, pp. 220--230.

Kimmelman, V. (2009a). Parts of speech in Russian Sign Language: The role of iconicity and economy. *Sign Language & Linguistics*, 12(2). pp. 161--186.

Kimmelman, V. (2009b). Reflexive pronouns in Russian Sign Language and Sign Language of the Netherlands // MA thesis, Universiteit van Amsterdam.

Krňoul, Z., Kanis, J., Železný, M., Müller, L. Czech Text-to-Sign Speech Synthesizer (2008). In *Proc. MLMI Workshop*, Brno, Czech Republic, Springer LNCS 4892, pp. 180--191

Mjasoedova, M.A., Filippovich, Yu.N. (2010). A system for automated translation of signed speech (Система автоматизированного перевода жестовой речи). In *Proc. CLAIM Workshop*, Moscow: MSTU, Russia.

Prillwitz S., et al. (1989). *HamNoSys. Hamburg Notation System for Sign Languages. An introductory guide.* (International Studies on Sign Language and Communication of the Deaf; Hamburg: Signum, 46 p.

Prozorova, E.V. (2007). Russian sign language as the object of linguistic research (Российский жестовый язык как предмет лингвистического исследования). *Voprosy yazykoznaniya*, 1, pp. 44--61.

Voskresenskiy, A., Gulenko, I., Khakhalin, G. (2009). RuSLED Dictionary as Tool for Semantic Study // *Computational Linguistics and Intellectual Technologies*, Moscow, Russia, pp. 64--69.

Zajtseva, G.L. (2000). *Signed speech. Fingerspelling* (Жестовая речь. Дактилология), Moscow: Vlados, 192 p.

Zelezny, M., Krnoul, Z., Cisar, P., Matousek, J. (2006). Design, implementation and evaluation of the Czech realistic audio-visual speech synthesis // *Signal Processing,* 86(12), pp. 3657—3673.

# From form to function. A database approach to handle lexicon building and spotting token forms in sign languages

**Reiner Konrad, Thomas Hanke, Susanne König, Gabriele Langer, Silke Matthes, Rie Nishio, Anja Regen**
Institute of German Sign Language (IDGS), University of Hamburg
{reiner.konrad, thomas.hanke, susanne.koenig, gabriele.langer, silke.matthes, rie.nishio, anja.regen}
@sign-lang.uni-hamburg.de

**Abstract**

Using a database with type entries that are linked to token tags in transcripts has the advantage that consistency in lemmatising is not depending on ID-glosses. In iLex types are organised in different levels. The type hierarchy allows for analysing form, iconic value, and conventionalised meanings of a sign (sub-types). Tokens can be linked either to types or sub-types.
We expanded this structure for modelling sign inflection and modification as well as phonological variation. Differences between token and type form are grouped by features, called qualifiers, and specified by feature values (vocabularies). Built-in qualifiers allow for spotting the form difference when lemmatising. This facilitates lemma revision and helps to get a clear picture of how inflection, modification, or phonological variation is distributed among lexical signs. This is also a strong indicator for further POS tagging. In the long term this approach will extend the lexical database from citation-form closer to full-form.
The paper will explain the type hierarchy and introduce the qualifiers used up-to-date. Further on the handling and how the data are displayed will be illustrated. As we report work in progress in the context of the DGS corpus project, the modelling is far from complete.

**Keywords:** lemmatisation, type hierarchy, citation form modification, variation, qualifiers, full-form lexical database

## 1. Background

The aim of pre-processing language data in corpus linguistics is lemmatisation. Counting and sorting of words or word forms, part-of-speech tagging, further annotation and analysis rely on machine-readable, lemmatised corpora. Reliability as one of the quality criteria of empirical science depends on how consistently tokens are matched to lexical types.

Whereas written texts of languages with a written tradition are pre-processed more or less automatically, spoken texts have to be written down beforehand. To build a corpus in an oral language with no written tradition, one has to choose an appropriate writing and/or notation system. This is the case for sign languages that have no written tradition. Phonographic notation systems such as HamNoSys or SignWriting were developed to write down the form of a sign and are part of a transcription system. But as the International Phonetic Alphabet (IPA) in spoken languages, inventory and conventions of notation systems are not helpful for lemmatising. Therefore a coding system in the sense of Hulst & Channon (2010) is needed that allows for computerized sorting, counting and comparing of signs.

For coding tokens of sign types, glossing is the most widespread practice. Glosses are written words from the surrounding spoken language or from the researcher's language. Their meaning usually covers one of the lexical meanings of the sign. They are "relatively crude and simplistic" translations (Johnston, 2009: 91). However, a gloss neither represents the contextual meaning of a sign nor does it give any information about the sign form. There are two main reasons why glosses made their way in sign language linguistics: First, glosses are a mnemonic aid. For those having some knowledge of the respective sign language, glosses can be used as a hint to recall the sign. Second, with glosses one can communicate with ease about signs. Using glosses for literal or free translation is misleading. In corpus linguistics we are likely to deal with thousands of signs and sign variants so that the first reason is bound to fail as "it is often very difficult to know with certainty which sign form is actually being referred to by a particular gloss" (Johnston, 2009: 91). The only way to achieve consistent token-type matching is to use glosses as ID-glosses, as unique identifiers of a sign (Johnston, 2010a). This means that glosses function as if they were identifiers. Working with annotation tools like ELAN[1], where tags are not linked to a lexical database, this seems the only way to build reliable lemmatised corpora.

The reason why we developed iLex (Hanke, 2002; Hanke & Storz 2008), an integrated lexical database for sign languages, is to handle large numbers of lexical types and their tokens in a consistent way. Glosses are helpful to represent sign types in the two ways mentioned above, but they are not used as identifiers in iLex. They are one value of a lexical entry amongst others as e.g. the citation form, written in HamNoSys. The identifier of each entry is a numeric code created by the database itself that guarantees its uniqueness and allows for restrictions such as one cannot create a new type entry with an already existing gloss string. In combination with token tiers that only allow for tags whose value is a type ID, the software supports the transcriber in being consistent. This support is essential in a multi-user environment, especially for quality assurance. Lemmatisation does not rely on glosses as

[1] EUDICO Linguistic Annotator; latest version and documentation are online available at: http://www.lat-mpi.eu/tools/elan/.

free text annotations, but is executed by linking tokens to type entries in a unique way.[2]

## 2. Data base approach and type hierarchy

With no comprehensive dictionary or lexical resource at hand, token-type matching is hard to achieve. As a bottom-up approach in transcribing each token form is far from realistic, the only way out is to build up a lexical database in parallel to segmenting and lemmatising signed utterances. This means that the transcriber constantly has to switch between top-down driven type matching and bottom-up driven adding new type entries (Konrad & Langer, 2009; König et al. 2010). In spoken languages, lexemes are conventionalised form-meaning pairs. Instantiations (tokens) of lexemes can have different forms according to a limited set of inflected forms. Applying this lexicological and morphological model to sign languages, one has to deal with two issues:

1. Due to its iconic aspects a sign can cover a far wider range of meanings than words. It can be combined with different mouthings to express meanings that are not necessarily semantically related. Leaving aside metaphorical use and homophone calques[3], usually all these meanings are related to the same underlying image. In König et al. (2008) we refer to this process of productively combining signs and mouthings as the "iconic-combinatorial procedure".
2. Until now, no complete descriptive grammar of any sign language exists. It is an open question whether one can define complete form paradigms for different sign classes. It is one of the research directions in sign language corpus linguistics to validate assumptions on part-of-speech classification and e.g. verb modification.[4]

### 2.1 Type hierarchy: types and sub-types (double glossing)

Our approach to face the first issue was to take the iconicity of signs into account. Identifying lexemes by comparing token forms and meanings following the rule "same form (paradigm), same (lexical) meaning → same lexeme" does not fit the needs of sign languages. In many cases this would result in mapping the spoken language lexicon onto the sign lexicon. In changing the rule into "same form, same iconic value (+ same image producing technique[5]) → same lexeme", things look quite different. In accord with Ebbinghaus & Hessmann, their assumption that signs and words (perceptible as mouthings) contextualise each other mutually and their postulation that "[i]nformation about regular collocations with nonmanually produced units should be part of the lexicographic description of the manual lexicon of a sign language" (Ebbinghaus & Hessmann 2001: 134), we distinguish between conventional and productive sign-mouthing combinations. This procedure is operationalised by double glossing[6] and implemented as a type hierarchy in iLex. Type entries in the table "types" are linked to the table "levels" which defines type dependencies, and what kind of type information can be added. Level-3 types (in the following called types) can be parents of several level-1 types (children; in the following called sub-types). Sub-types cannot be created without a reference to a type (parent). Sub-types are conventionalised sign-mouthing combination with a lexicalised meaning. In most cases the meaning corresponds to the meaning of the mouthed word.[7] Sub-types can only be subsumed to types if they share the same underlying image and the same citation form. This information is stored in the type entry and is valid for all sub-types. Meaning is not entered in the type, but in the sub-type entry. In contrast to so-called productive signs created on the spot, corresponding to "partly-lexical signs", each type, corresponding to "full-lexical signs" (Johnston 2010a) must have at least one lexical meaning, so that each type has to be parent of at least one sub-type. If the form of a token can be identified as an instantiation of the type's citation form, and if the iconic value of the type is valid for its use in context, but the contextual meaning of the token does not correspond to the lexical meaning of a sub-type, this token will be matched directly to the type. In many cases such tokens are productive sign-mouthing combinations covering a wide range of meanings. Matching tokens either to types or sub-types helps to sort "regular collocations with nonmanually produced units" from occasional collocations. Grouping subtypes into types allows for identifying different conventionalised readings of a sign (polyseme[8]) and prevents from mapping the spoken lexicon into the sign lexicon.

[2] Cf. Johnston's "note on the use of an integrated lexical database with ELAN" (Johnston 2011: 16-17).

[3] The DGS sign for 'Enkel' (grandchild) is the same as for 'Engel' (angel) because of the similarity of sound in spoken German. The mouthing of the German words is the same. This phenomenon is not restricted to DGS, e.g. in ASL you will find HUNGRY/HUNGARY.

[4] Cf. Johnston's (2010b: 141) findings on spatial modification of verbs in Auslan that support Liddell's (2003) analysis of indicating and depicting verbs in ASL.

[5] See Langer (2005) for a detailed description, König et al. (2008) for a short version.

[6] See Konrad (2011b: 145-155) for an extensive discussion; see also König et al. (2008), König et al. (2010).

[7] It happens that tokens are articulated without the corresponding mouthing. If the contextual meaning of the sign fits to the lexical meaning, this token will also be matched to the sub-type.

[8] Note that on the one hand in sign language we have to deal with far-reaching lexical ambiguity which is more context sensitive than in spoken languages, on the other hand iconicity is a valid criterion to group related meanings and distinguish lexemes (s. König et al. 2008), which is not applicable to spoken language.

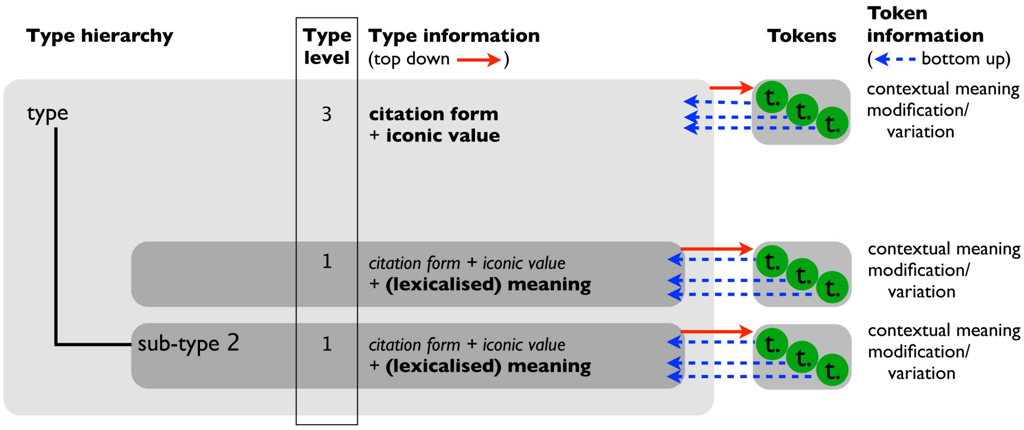



Figure 1: Type hierarchy, type information and token information

## 2.2 Type hierarchy and qualified types

The second issue regards the signs' potential for variation and modification in context. Differences between token form and citation form can either indicate variation that will be assigned to the phonological level or morphosyntactic patterns. When matching tokens to types, the transcriber has to compare token form and type citation form. Instead of deferring the documentation of token form differences to a second annotation pass (Johnston, 2010a: 116-117) where these differences are annotated in several tiers (orientation, citation modification, or variation tier)[9], we annotate this information to the token tag in the process of lemmatisation. These annotations are one of the main criteria to check whether the token-type matching is correct during the process of lemma revision. As in iLex tokens are linked to a type, all tokens of a type can be listed and sorted by token information such as form difference.[10]

Since 2009 we are modifying the type hierarchy in iLex in order to group different form features. Each type and sub-type can have several qualified types. Qualified types[11] are combinations of types with qualifiers. These qualifiers are form features that can have several feature values (see below). Instead of annotating the form difference to tokens, the transcriber can refer a token to an existing qualified type. This makes lemma revision easier because the tokens of one type are not only pre-sorted by conventional and productive use of signs but also by form features. Figure 2 shows the expanded type hierarchy, in figure 3 the structure is exemplified by parts of the subtree belonging to the type glossed DA1 (there).

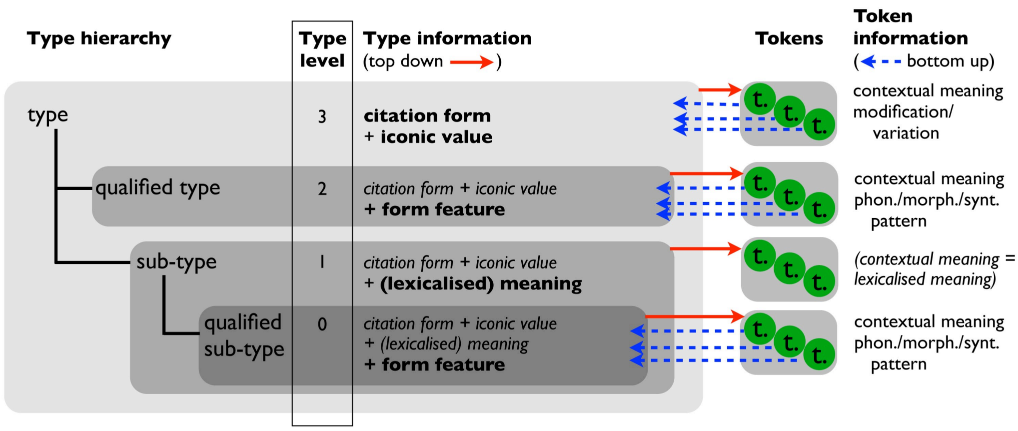



Figure 2: Expanded type hierarchy with qualified types

9 Cf. Johnston 2011: 53-70: "Secondary processing".

10 The process of lemma revision in iLex is described in Konrad (2011a pp. 93-96); see also König et al. (2010 and Konrad & Langer (2009).

11 In the following all what is said about qualified types is also valid for qualified sub-types.

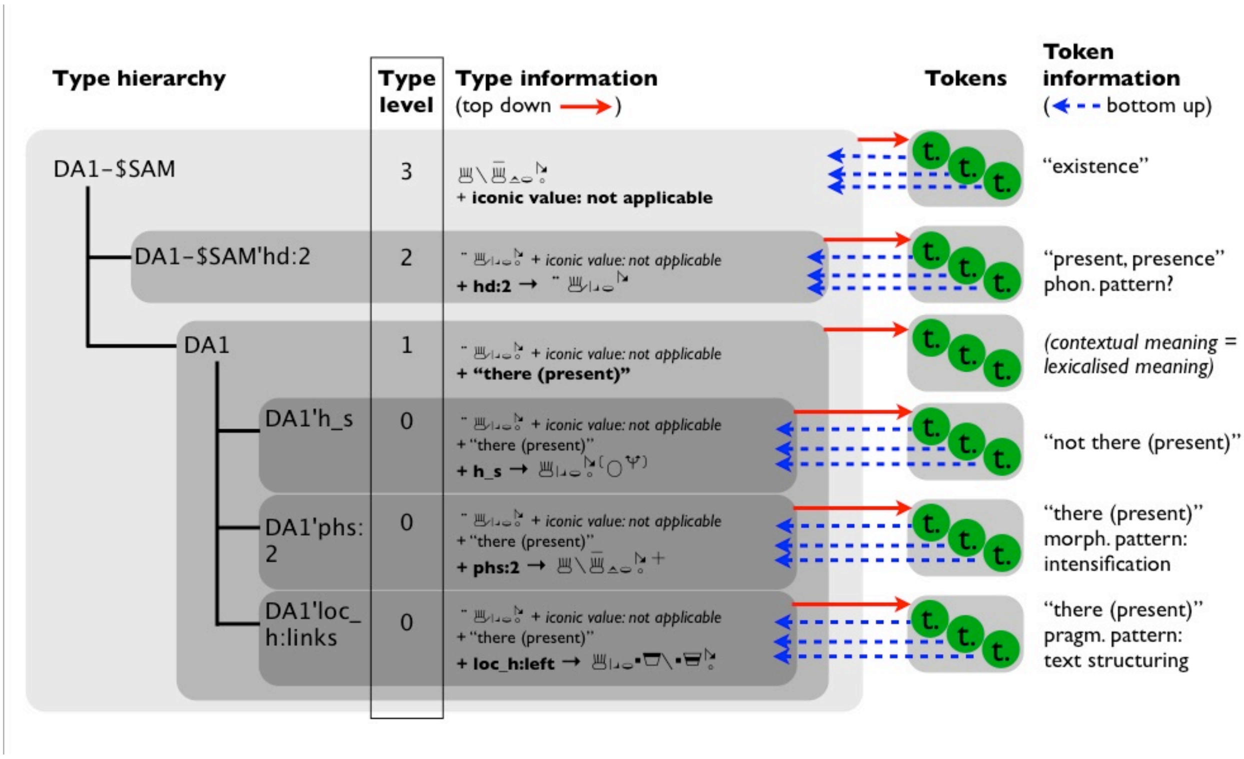



Figure 3: Expanded type hierarchy with example

In the following, the inventory of qualifiers and the handling in iLex will be described.

## 3. Inventory of qualifiers and feature values

As we report on work in progress, the qualifier list and the corresponding feature values are far from complete. Until now they cover some morphological patterns like inflection (location, direction, source and goal), where the form-function relation is known. Other form features like "phases" of movement (i.e. different kinds of repetition) are yet deliberately unspecified for a certain function. At the end of the lemmatisation process the distribution and frequency of these features over types and sub-types, in combination with context information, will show what phonological, morphological, and syntactic change each sign can undergo. The instantiations of qualifiers are a strong indicator for part-of-speech tagging. In addition, they will allow us to move from a citation-form lexical database closer to full-form.

The following table lists the qualifiers that are already used in iLex to specify a type or a sub-type. For most of the qualifiers feature values are pre-defined and implemented as vocabularies. The aim of these values is to get a coarser division of token form characteristics than it would be by transcribing the token form, e.g. using HamNoSys. Closed vocabularies also have values for tokens that need to be discussed (*unclear*) or that are candidates for a new feature value to be added (*leftover*).

When creating a new qualified type the qualifier code is added automatically to the type/sub-type gloss. This telling gloss suffix makes it easy to understand the modification of the sign form (see explanation to figure 5 below). In general, the form of a qualified type, like the citation form of the type, is transcribed in HamNoSys. For some qualifiers this HamNoSys string can also be adapted automatically. To code tokens that show more than one form feature qualifiers can be combined so that the vocabularies can be kept concise (see below table 1, figure 4, and 5).

Except *head shaking*[12], all features refer to manual parameters. Feature values of *number of hands* include one- and two-handed and should only be used for symmetrical signs. Two-handed symmetrical signs are further qualified by movement (reversed, anti-cyclic).

The modification of the citation form by adding the nondominant hand so that a one-handed sign becomes an asymmetrical two-handed sign, or dropping the nondominant hand of an asymmetrical sign is divided into four qualifiers: *hold* and *hold resume* to identify sequences where the nondominant hand is part of a previous sign, *continued* to indicate that starting from a simultaneous sign construction the articulation of the sign of the nondominant hand is stretched over two or more signs in the dominant hand, and *base* when the nondominant hand is added (weak prop) and its iconic value can be analysed as a substitutive or manipulative image producing technique (see Langer 2005; König et al. 2008). Therefore, several feature values are provided for the basic and frequently used handshapes B-hand,

[12] This qualifier is used when the meaning of a sign is negated only by headshaking. The negation is limited to the sign and does not affect the whole phrase. It is not used when the negation is expressed by a manual form feature like *alpha negation*. In these cases headshake is additional and will be annotated in the gesture tier.

C-hand, and fist. In contrast, weak hand drop is coded by “'bas:none”.

Spatial modification is divided into movement and location. Location in turn is grouped into the use of a real location of a referent or an action (qualifier *location horizontal, sagittal, vertical*) and the metaphorical location e.g. when the signing space is used to contrast two topics and therefore signs are located either on the left or on the right side (Johnston, 1991: 10-11). The text structuring and pragmatic function of location will be coded separately (qualifier *location text structure horizontal, sagittal, vertical*).[13]

| **Covering …** | **Qualifier** | **Code** | **Feature** | **Vocabulary** |
|---|---|---|---|---|
| Number of hands | number of hands | hd | Number of hands | closed |
| Nondominant hand | hold | h | hold | closed |
| | hold resume | hres | | |
| | continued | cont | | |
| | base | bas | base | closed |
| Location | location horizontal | loc_h | location horizontal | closed |
| | location sagittal | loc_s | location sagittal | closed |
| | location vertical | loc_v | location vertical | closed |
| | location on body | bodyloc | location on body | open |
| | location text structure horizontal | loc_ts_h | location ts horizontal | closed |
| | location text structure sagittal | loc_ts_s | location ts sagittal | closed |
| | location text structure vertical | loc_ts_v | location ts vertical | closed |
| Movement … | source | src_h | source+goal | closed |
| | source | src_v | location ts vertical | closed |
| | goal | gol_h | source+goal | closed |
| | goal | gol_v | location ts vertical | closed |
| Movement | phases | phs | phases | closed |
| | reverse | rev | | |
| | offset direction | offdir | offset direction | closed |
| | alpha negation | alph | | |
| Handshape | assimilation | assim | | |
| Nonmanuals | head shaking | h_s | | |
| Sub-system manual alphabet | fa one-handed | 1 | alphabet | open |
| | fa two-handed | 2 | alphabet | open |
| | fa tracing | sk | alphabet | open |
| | fa tracing on hand | skh | alphabet | open |
| | fa ligature | lig | | |
| Sub-system numbers | quantity | q | quantity | closed |
| | number | n | number | closed |
| | m out of n | of | quantity | closed |
| | detour | dt | | |

Table 1: Inventory of qualifiers

[13] Of course, metaphorical use of location to express temporal aspects like locating signs on a horizontal, vertical, or sagittal time line will also be covered by a separate location qualifier.

Real location is coded by three features following the three dimensions in space. *Location horizontal* has five values (head, throat, upper chest, chest, belly), *location vertical* has also five values (left, diagonal left, front, diagonal right, right), and *location sagittal* has three values (close, near, far). These values can be added using a diagram that displays the spatial relations. In the same way the features for the text structuring use of location follow the three dimensions, but their vocabularies are smaller (high and low for the horizontal plane, left and right for the vertical plane, and front and back for the sagittal plane). Another qualifier helps to code all tokens that were modified by a specific body location (*location on body*). Due to anatomical facts and the more or less specific use of body parts the corresponding feature list can be quite large and is implemented as an open vocabulary.

For movement modification, following the well-known inflection of directional verbs, the qualifiers *source* and *goal* are used. Each of them are coded by two features with respect to the horizontal and vertical dimension (*source h, source v, goal h, goal v*). The horizontal plane is divided into left, right, middle, and signer, whereas for the vertical plane the vocabulary of the feature *location text structure vertical* can be used. In addition the sweeping and the zigzag movement that are morphological features of the distributional aspect of some verbs are coded separately. These movement modifications also involve change in palm orientation and/or direction of the fingers.

The qualifier *phases* covers repetition of movement. It turned out that for signs which already have repeated or repeated circular movement in their citation forms it was not sufficient only to label one up to three repetitions. Further on reversed movement is covered by a separate qualifier (*reverse*) just as repeated movement with simultaneous change in direction (*offset direction*). The combination of these three features allows for coding different movement patterns. Finally, in DGS some verbs can be modified by changing their movement as if the hand would trace the Greek letter alpha (α) in the air. The semantic function of this movement pattern is negation and will be covered by the qualifier *alpha negation*.

In order to get a clear picture of the variety of signs for manual alphabet and numbers, we have defined several qualifiers. For the manual alphabet (*alphabet*) we differentiate between one-handed (*fa one-handed*) and two-handed signs (*fa two-handed*), tracing signs in the air (*fa tracing*) or on the nondominant hand (*fa tracing on hand*). If the fingerspelled letters are connected by a slight movement, this feature can be added by choosing the qualifier fa ligature. Signs for numbers are covered by the qualifier *number*. For number incorporation the qualifier *quantity* is provided. It can be combined with the feature *detour* for movements with an additional slightly curved path. If the nondominant hand shows a quantity from two to five and serves as a list for the dominant hand pointing to any finger but not the one representing the maximal number, a separate feature *m out of n* is used.[14]

## 4. Handling: Attributing qualifiers and data retrieval

The main task in the process of lemmatisation is to find the right type a token should be matched to. In comparing type citation form and token form the transcriber should be able to document his findings in a quick and easy way. The simplest way is to mark the token that there is a form difference. In the second pass of lemma revision where all tokens of one type are checked, this piece of information is relevant. A more efficient way for lemma revision is to note the salient feature in which the token is different from the citation form, e.g. in HamNoSys. Sorting all tokens of one type according to these annotations helps to get a quick overview and to find tokens with the same kind of modification or variation. This is what we did before implementing qualifiers in iLex and what we still will do when the qualifier and its feature values do not cover all the token features. So lemma revision is not a singular pass, but has to be done several times.

The reason why we moved to qualifiers and qualified types is a practical one. Instead of annotating several times the same salient form feature to  tokens and in a second pass grouping these tokens together, the transcriber can do this in the first instance of the lemmatising process. After linking the token tag to a type, e.g. by dragging an item from the type list and dropping it over the token tier (Konrad & Langer, 2009), one can use the context menu (right mouse click) to display type, sub-types, and qualified types of the chosen item. Figure 4 shows the type and all the qualified types of the sub-type glossed as DA1.[15]

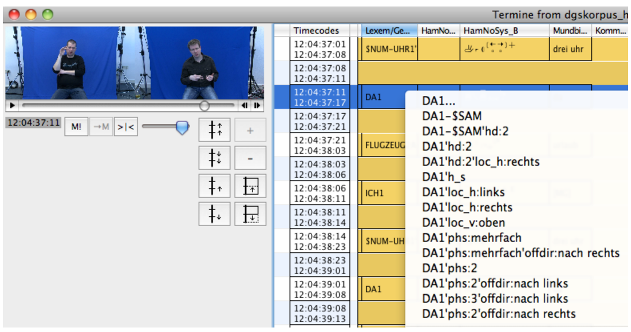



Figure 4: Context menu displaying type and qualified types of the sub-type DA1

If the token form does not match any of the already existing qualified types, the transcriber can either annotate the form difference or create a new qualified type by choosing “DA…” on the head of the listed items.

[14] See Liddell 2003, Johnston 2011: list buoys.

[15] To see all the sub-types of the type glossed as DA1-$SAM, one first links the token to the type and then re-uses the context menu.

Note that even though the database creates new type entries, this procedure is always bottom-up. Qualified types are derived from already existing types. The type hierarchy is used here to simplify token annotation. As with lexicon building, in the beginning one has to invest more in creating new qualified types, but once they are there, they will be listed and can easily be picked to spot token form features.
Instead of running searches over a multitude of tiers to find tokens that match the search criteria, the type hierarchy allows for having all spotted token form features of one type at a glance. Figure 5 shows the existing qualified types of the sub-type DA1. There are tokens with headshake ('h_s), two-handed articulation ('hd:2), different location ('loc_h:links (left)), combination of two-handed and location feature ('hd:2'loc_h:rechts (right)), repetition ('phs:2, 'phs:mehr-fach (multiple)), and combination of repeated movement with simultaneous change in direction (e.g. 'phs:2'offdir:-nach links (to the left)). Double-clicking on one of the items opens the qualified type entry where all tokens are displayed.[16]

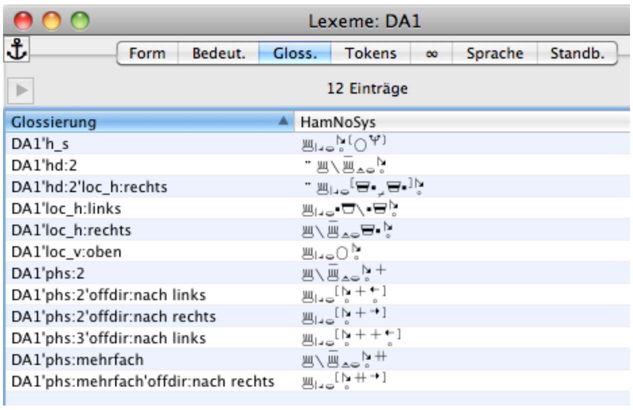



Figure 5: Qualified types of sub-type DA1

## 5. 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.

## 6. References

Crasborn, O., Zwitserlood, I. (2008). *Annotation of the video data in the Corpus NGT.* Nijmegen: Radboud University. [Online available: http://www.ru.nl/aspx/download.aspx?File=/contents/pages/515436/corpusngt_annotationconventions.pdf].

Hanke, T. (2002): iLex. A tool for Sign Language Lexicography and Corpus Analysis. In González Rodriguez, M. & Paz Suarez Araujo, C. (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.

Hanke, T., Storz, J. (2008). iLex - A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In Crasborn, O., Efthimiou, E., Hanke, T., Thoutenhoofd, E.D., Zwitserlood, I. (Eds.), *LREC 2008 Workshop Proceedings. W 25: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora.* Paris: ELRA, pp. 64-67. [Online available: http://www.lrec-conf.org/proceedings/lrec2008/workshops/W25_Proceedings.pdf].

Hulst, H. van der, Channon, R. (2010). Notation systems. In Brentari, D. (Ed.), *Sign Languages.* Cambridge University Press, pp. 151-172.

Johnston, T. (1991). Transcription and Glossing of Sign Language Texts: Examples from AUSLAN (Australian Sign Language). In *International Journal of Sign Linguistics*, 2(1), pp. 3-28.

Johnston, T. (2009). Creating a Corpus of Auslan within an Australian National Corpus. In Haugh, M., Burridge, K., Mulder, J., Peters, P. (Eds.), *Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus*. Somerville, MA: Cascadilla Proceedings Project, pp. 87-95.

Johnston, T. (2010a). From archive to corpus: Transcription and annotation in the creation of signed language corpora. In *International Journal of Corpus Linguistics*, 15(1), pp. 106-131.

Johnston, T. (2010b). Adding value to, and extracting of value from, a signed language corpus through secondary processing: implications for annotation schemas and corpus creation. In Dreuw, P., Efthimiou, E., Hanke, T., Johnston, T., Martínez Ruiz, G., Schembri, A. (Eds.), *LREC 2010 Workshop Proceedings. 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pp. 137-142. [Online available: http://www.lrec-conf.org/proceedings/lrec2010/workshops/W13.pdf].

Johnston, T. (2011). *Auslan Corpus Annotation Guidelines.* [Last updated 30 November 2011. Online available: http://www.auslan.org.au/video/upload/attachments/AuslanCorpusAnnotationGuidelines30November2011.pdf].

König, S., Konrad, R., Langer, G., Nishio, R. (2010). How Much Top-Down and Bottom-Up do We Need to Build a Lemmatized Corpus? Poster presented at the *Theoretical Issues in Sign Language Research Conference (TISLR 10), Sept 30 - Oct 2, 2010 at Purdue University, Indiana, USA*. [Online available: http://www.purdue.edu/tislr10/pdfs/Konig%20Konrad%20Langer%20Nishio.pdf].

König, S., Konrad, R., Langer, G. (2008). What's in a sign? Theoretical Lessions 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. (2011a). *Die Erstellung von Fachgebärdenlexika am Institut für Deutsche Gebärdensprache (IDGS) der Universität Hamburg (1993-2010).*

[16] By changing the list statement more information, e.g. the number of tokens in this view, can be displayed.

Corrected Version. [Online available: http://www.sign-lang.uni-hamburg.de/projekte/mfl/konrad_2011_fachgeblexika.pdf]
Konrad, R. (2011b). *Die lexikalische Struktur der Deutschen Gebärdensprache im Spiegel empirischer Fachgebärdenlexikographie. Zur Integration der Ikonizität in ein korpusbasiertes Lexikonmodell.* Tübingen: Narr.
Konrad, R., Langer, G. (2009). Synergies between transcription and lexical database building: The case of German Sign Language (DGS). In Mahlberg, M. / González-Díaz, V., Smith, C. (Eds.), *Proceedings of the Corpus Linguistics Conference (CL2009). University of Liverpool, UK, 20-23 July 2009.* [Online available: http://ucrel.lancs.ac.uk/publications/cl2009/346_FullPaper.doc].
Langer, G. (2005). Bilderzeugungstechniken in der Deutschen Gebärdensprache. In *Das Zeichen.* 70, pp. 254-270.
Liddell, S.K. (2003). *Grammar, gesture and meaning in American Sign Language.* Cambridge: Cambridge University Press.

# Integrating corpora and dictionaries: problems and perspectives, with particular respect to the treatment of sign language

**Jette H. Kristoffersen, Thomas Troelsgård**
Centre for Sign Language, University College Capital
Ejbyvej 35, DK-2740 Skovlunde, Denmark
E-mail: jehk@ucc.dk, ttro@ucc.dk

**Abstract**

In this paper, we will discuss different possibilities for integration of corpus data with dictionary data, mainly seen from a lexicographic point of view and in a sign language context. For about 25 years a text corpus has been considered a useful, if not necessary tool for editing dictionaries of written and spoken languages. Corpora are equally useful to sign language lexicographers, but sign language corpora have not become accessible until recent years. Nowadays corpora exist, or are being developed, for several sign languages, and we have the possibility of editing new, truly corpus-based sign language dictionaries, and of developing interfaces that integrate corpus and dictionary data. After a brief look at three existing integrated interfaces, one for German, one for Danish, and one for Danish Sign Language, we point out some of the problems that should be considered when making an integrated interface, and, finally, we briefly outline the future perspectives of integrated sign language corpus-dictionary interfaces.

**Keywords:** sign language, integrated interfaces, lexicography

## 1. Introduction

In this paper, we will discuss different possibilities for integration of corpus data with dictionary data, mainly seen from a lexicographic point of view and in a sign language context.

Since at least the 1990's, a text corpus has been considered a useful, if not necessary tool for editing dictionaries of written/spoken languages. A corpus can provide the lexicographer with:

- frequency lists (e.g. used in connection with lemma selection or ordering of homonyms)
- examples of language use (e.g. used as evidence of particular word senses, or for example sentences (directly or adapted))
- frequent co-occurrences (e.g. used for describing multi-word expressions or valency patterns)

We know of no sign language dictionary that is truly corpus-based, or even edited with extensive use of the tools provided by a corpus, probably due to the fact that larger, fully annotated sign language corpora is a relatively new phenomenon. An example will be the new German Sign Language-German dictionary which is part of the DGS Corpus Project and will be "the first comprehensive corpus-based dictionary of DGS" (DGS-Corpus, no date [online]).

For corpus projects of languages with an established written form, the lemmatisation of tokens is typically based on an existing dictionary. For sign language corpora this is equally appropriate, as argued in Johnston (2008), but the execution is impeded by the inevitable and – at least in the nearest future – manpower-consuming task of tokenising the corpus texts sign by sign. Furthermore, this approach presupposes the existence of a dictionary or lemma list composed using consistent lemmatisation principles and a consistent identification of the lemmas, e.g. through unique glosses or numbers. For a number of sign languages no such dictionary exists, and building a corpus, would imply the simultaneous building of a dictionary, which would make the process even more time-consuming.

As language resources dictionaries and corpora are both valuable tools for many types of users, and combining the two in one interface, or linking between dictionary and corpus interfaces could, if it is done in a clear and preferably intuitive understandable way, afford a synergetic enhancement of the resources.

## 2. Examples of interfaces that integrate corpus and dictionary content

In this section we will take a closer look at some existing interfaces that integrate corpus and dictionary content. The German DWDS (Digitales Wörterbuch der deutschen Sprache, 'Digital Dictionary of the German Language') (DWDS, no date [online]) is an example of a combinded corpus-dictionary product where, as Asmussen puts it: "Corpus and dictionary are not formally interlinked, they appear side by side, accessible through a joint interface", (Asmussen, 2012). A standard word lookup in DWDS presents the user with six sub-windows, see Figure 1. The standard lookup shows (letters refer to the labels Figure 1): the result of a lookup in the DWDS dictionary (A), the result of a lookup in OpenThesaurus[1](B), a tag cloud ("Wortprofil") based on the DWDS corpus (C), the result of a lookup in an etymological dictionary (D), and, finally, two concordances (E and F), one drawn from the basic DWDS corpus, and one drawn from a newspaper corpus (Die Zeit) . The standard view of a search result can be changed by adding or removing included resources, and

[1] OpenThesaurus is an open source thesaurus project that was initiated in connection with the development of the OpenOffice software. Information about the project can be found at http://www.openthesaurus.de

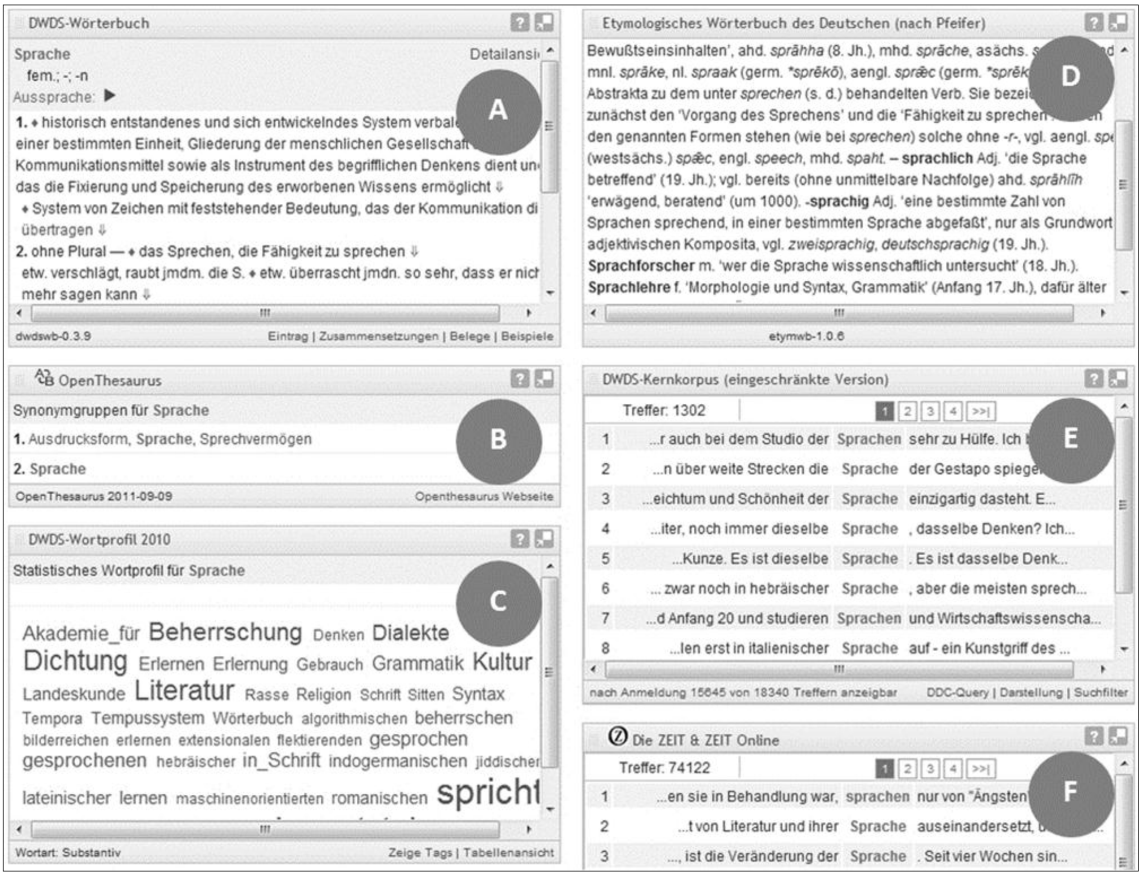



Figure 1: Partial screen dump of the standard view of a lookup of *Sprache* ('language') in DWDS.


Figure 2: A dictionary lookup at Ordnet.dk, with links to corpus searches.

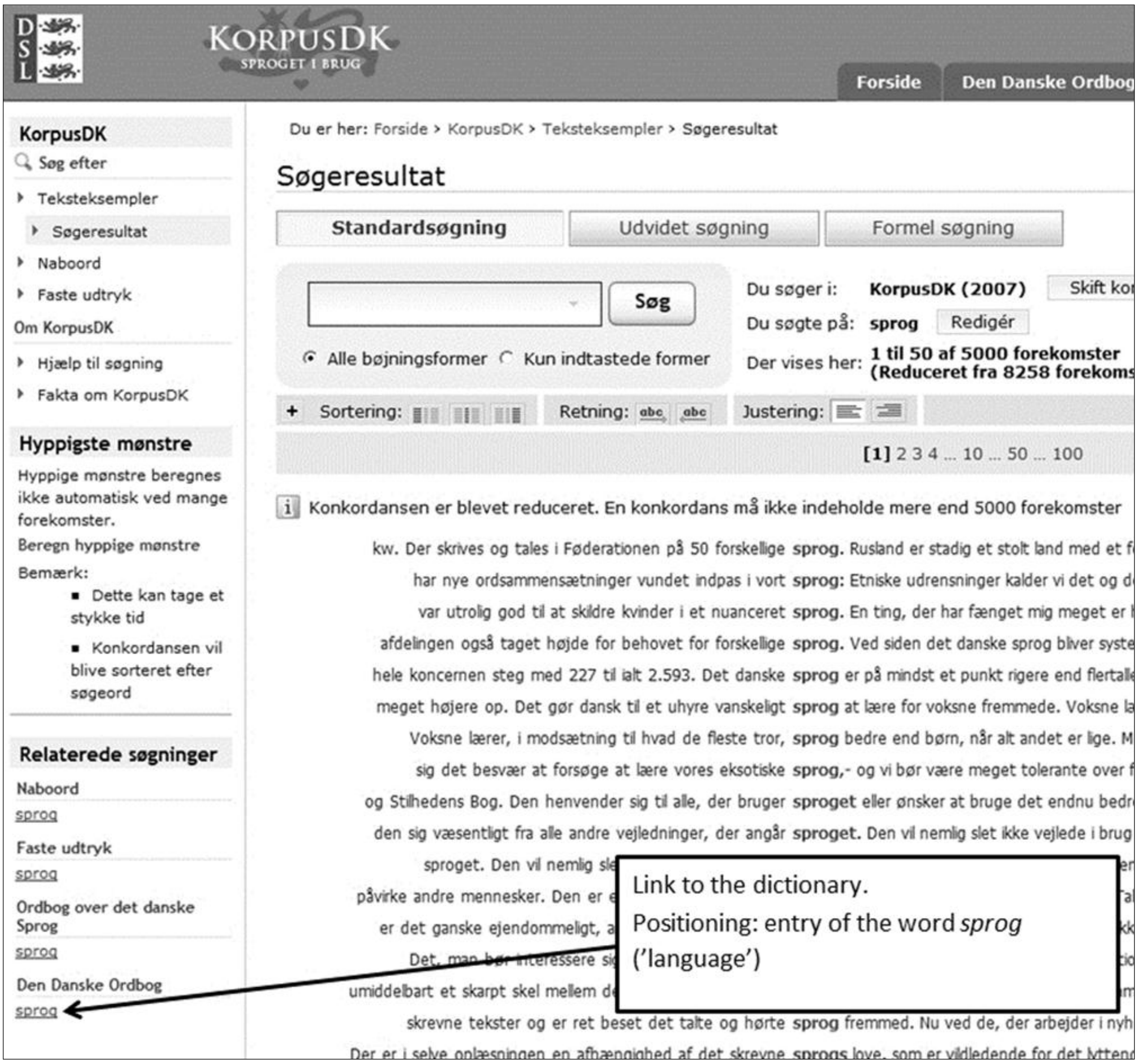



Figure 3: Result of a corpus search at Ordnet.dk, with link to the dictionary.

by expanding or collapsing the views of the sub-windows individually.

Another example is the Ordnet.dk (no date [online]) website, which deals with Danish. This site also has no formal interlinking, but if differs considerably from the DWDS site, as its approach aims an interconnection of several resources rather than a simultaneous access. Thus, there is no universal search facility, but a word search in one of the two included dictionaries (a contemporary and a historic) provides the user with the dictionary content as well as with links to relevant corpus searches (on the key word and on selected collocations), see Figure 2. Similarly, a corpus search result is accompanied by lookup links to the two dictionaries, see Figure 3. In addition to this, all three resources include a link to a list of the most frequent co-occurences of the word. See Trap-Jensen (2010) for more information on the Ordnet.dk website.

Where the two examples mentioned above are not formally interlinked, the last example, the Danish Sign Language Dictionary (no date [online]) is. The weakness of this dictionary, on the other hand, is that its corpus is what Asmussen (2012) refers to as a "quasi-corpus", in this particular case, a corpus build entirely of adapted sentences, namely the usage examples of the dictionary. Furthermore, half of these sentences are derived from video recordings of natural signing, while the remainder have been constructed by native signers. The integration of the two resources is quite basic and somewhat similar to the one used in Ordnet.dk; in the dictionary, there are links from each sign entry to a concordance view of the all the occurrences of the sign in the collection of example sentences, see Figure 4. In the other direction, the individual signs in the sentences of a concordance are linked to the corresponding sign entries in the dictionary, see Figure 5. This feature is added in order to present additional examples of the use of a sign to the user, and although the corpus is not a "real" corpus, the dictionary site still serves as an example of how corpus (or corpus-like) data can be integrated into a sign language dictionary. A discussion of what type of sentence was considered the most suitable for uncommented use as example sentences in the Danish Sign Language Dictionary can be found in Kristoffersen & Troelsgård (2010). A more detailed description of the dictionary can be found in Kristoffersen & Troelsgård (2012).

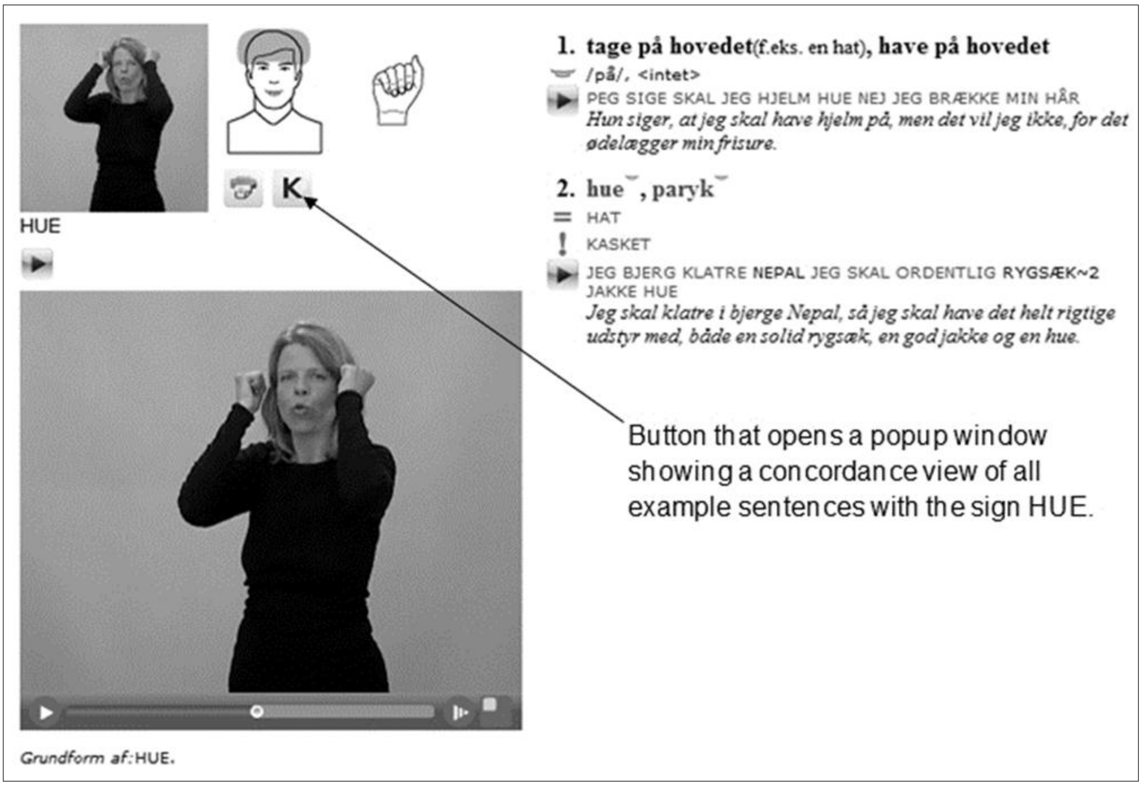



Figure 4: The Danish Sign Language Dictionary. Entry for the sign HUE (‘cap’), with a link to a concordance view of all example sentences containing this sign.


Figure 5: The Danish Sign Language Dictionary. Concordance view of example sentences containing the sign HUE (‘cap’). Glosses in the concordance lines are linked to the appropriate sign entries (if they exist).

## 3. Considerations regarding the integration of corpora and dictionaries

Language use is described differently in a dictionary and in a corpus; whereas the dictionary data are the result of an editing process and often adapted to a specific purpose, the corpus data, be it text examples or co-occurrence statistics etc., are "raw" and have to be interpreted by the user. The difference between the two resources is somewhat comparable to that between an encyclopaedia and an internet search; the former typically being more well-arranged and reliable, while the latter often provides more information, and more updated information, the downside being that it is presented as lots of co-ordinate results with no quality guarantee. How "dangerous" it is to present corpus data to the user depends partly on the nature of the corpus texts, partly on how trained the user is in the use of the corpus.

Corpora of languages with a written form typically have written texts as their main source; digital text, e.g. from newspapers or from the internet is easy to obtain, and you will relatively unproblematically be able to build a corpus – apart from legal issues and an expected margin of error in connection with the tokenisation of the corpus texts. Building corpora of spoken or signed languages, on the other hand, requires a manual or, at best, semiautomatic tokenisation process in order to become searchable. Mainly for this reason such corpora are typically smaller than corpora of written text.

Large corpora of a written language are often composed of different types of text, balanced in order to obtain a broad and adequate picture of the language use. For corpora of signed and spoken language, such a balancing will probably always be a major challenge; as you are dealing with non-written language, the only "authentic" text types available would be rather special ones like recordings of radio or television broadcasts, or recordings of speeches and conversations, which are rarely performed spontaneously. Thus, existing sign language corpora mainly consist of elicited data or recordings made for linguistic purposes in a more or less unnatural context. Many corpora contain non-edited language, allowing for ungrammatical language use, misspellings (for written language), and, especially for spoken/signed texts, elliptic utterances.

As a result of the above mentioned impediments for building a corpus, a corpus user could find him or herself dealing with corpus that is of limited size, with more or less unnatural text types, and containing ungrammatical sentences. For e.g. a lexicographer, this would be a minor problem, as he or she would look at the source critically, but for an inexperienced corpus user, and even more so for a user of an integrated corpus-dictionary interface, who isn't necessarily aware of the complex character of the resource, it could be quite problematic to extract the needed information, cf. the discussion in Asmussen (2012). For this reason, an integrated corpus-dictionary interface should always, at least ideally, provide the user with the necessary prerequisites for using the corpus in a meaningful way, e.g. by informing about the corpus sources, and by clearly indicating if the user is presented with edited or non-edited text.

A corpus-dictionary pair of a specific language can be more or less closely related, or coherent, so to say. Thus, the ideal prerequisite for an extensive integration of a corpus and a dictionary is a situation where the lexicon and definitions of the dictionary are based on the corpus which, in its turn, is linked token by token to the dictionary. On the other hand, as exemplified in Asmussen (2012), if the two resources are not based on the same source texts, an integration can lead to situations where the user is presented with a confusing result, e.g. if a specific word sense is predominant in the corpus, but absent in the dictionary (or the other way round).

## 4. Future perspectives of integrated sign language resources

The integration of sign language corpora and dictionaries is an obvious field for development in the future. The crucial formal interlinking between corpus and dictionary, which for written language corpora is often insufficient or missing, is typically an innate feature of a sign language corpus project, as a tokenisation is needed in order to make the corpus searchable, and as the tokenisation, in its turn, requires a dictionary or lexicon.

The synergy that rises from joining a corpus and a dictionary could even be enhanced by including e.g. grammatical information, or links to external resources. There is however a risk that the user is overwhelmed by the amount of diverse data and possibilities presented in the interface.

Another future challenge could be the development of means of accessing sign language corpora that are more appropriate than traditional text-based concordance views.

Several sign languages are now documented, or in the process of being documented, through corpora, e.g. Australian, British, Dutch, German, New Zealand and Swedish Sign Language, and, hopefully, we can look forward to seeing some innovative projects integrating corpus and dictionary data in the future.

## 5. References

Asmussen, J.(2012). Combined products: Dictionary and corpus. In R.H. Gouws et al., (Eds.) *Dictionaries. An International Encyclopedia of Lexicography* [Suppl. Volume of the Handbooks of Linguistics and Communication Science series: *Recent developments with special focus on computational lexicography*]. Berlin: Mouton de Gruyter. Forthcoming.

Danish Sign Language Dictionary (no date) [online]. http://www.tegnsprog.dk/ (Accessed 30 March 2012).

DGS-Corpus (no date) [online]. http://www.sign-lang.uni-hamburg.de/dgs-korpus/ (Accessed 30 March 2012).

DWDS (no date) [online]. http://www.dwds.de/ (Accessed 30 March 2012).

Johnston, T. (2008). Corpus Linguistics and Signed Languages: No Lemmata, No Corpus. In O. Crasborn et al., (Eds.), *Proceedings of the Third Workshop on the Representation and Processing of Sign Languages: 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.

Kristoffersen, J.H., Troelsgård, T. (2010). Compiling a Sign Language Dictionary. Some of the Problems faced by the Sign Language Lexicographer. In M. Mertzani (Ed.), *Sign Language Teaching and Learning; 14 Papers from the 1st Symposium in Applied Sign Linguistics, 24-26 September 2009*. Bristol: University of Bristol Centre for Deaf Studies.

Kristoffersen, J.H., Troelsgård, T. (2012). The Electronic Lexicographical Treatment of Sign Languages: The Danish Sign Language Dictionary. In S. Granger & M. Paquot (Eds.), *Electronic Lexicography*. Oxford: Oxford University Press. Forthcoming.

Ordnet.dk (no date) [online].
http://ordnet.dk/ (Accessed 30 March 2012).

Trap-Jensen, L. (2010). Access to Multiple Lexical Resources at a Stroke: Integrating Dictionary, Corpus and Wordnet Data. 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*. (Cahiers du CENTAL 2010). Louvain-la-Neuve: Presses universitaires de Louvain.

# A Colorful First Glance at Data on Regional Variation Extracted from the DGS-Corpus: With a Focus on Procedures

**Gabriele Langer**

University of Hamburg
Institute of German Sign Language and Communication of the Deaf
Binderstr. 34, 20146 Hamburg, Germany
E-mail: gabriele.langer@sign-lang.uni-hamburg.de

**Abstract**

In this work in progress procedures for analyzing and displaying distributional patterns of sign variants have been developed and tested on data for color signs elicited by the DGS Corpus Project. The data for this preliminary study were elicited as isolated signs and have been made accessible through spot annotations in iLex. The annotations had not been lemma revised but nevertheless revealed some interesting insights. Several color signs exhibited a high degree of variation. The distributional maps showed that a number of signs were mainly used in certain regions and thus provided evidence on dialectal differences within DGS. The relevant information necessary to generate distributional maps have been directly extracted via SQL-statements from the corpus and fed into R. The approach is data driven. The distributional maps show either the distribution of one sign form (variant) or of several different variants in relation to each other. Analyses of regional distribution as displayed by the distributional maps may support the annotation and lemma revision process and are a valuable basis for a lexicographical description of signs and their use as needed for compiling dictionary entries. A refined procedure to take multiple regional influences on informants into account for analysis is proposed.

**Keywords:** generation of distributional maps from corpus data, regional variation in DGS, color signs, data-driven approach

## 1. Introduction

Within the DGS Corpus Project about 1160 hours of footage with an estimated 540 hours of signed activity have been collected. 330 informants in 13 German regions were filmed in pairs. This material will constitute a general corpus of German Sign Language (DGS) after it has been made accessible through annotation. The next stage of the project is dedicated to annotation and transcription of the raw data. At a later stage the first corpus-based general dictionary of DGS–German will be produced based on the data documented in the corpus.

One of the project's aims is to document lexical variation including regional variation. Information on regional variation is an interesting and useful piece of information on signs that should be included in dictionary entries wherever possible. Within the project, procedures need to be proposed, tested and established to extract and present information on regional distribution from the corpus data efficiently as it is needed to support the compilation of dictionary entries. Even though the prerequisite for the analysis of many sociolinguistic variables are provided for in the metadata gathered, these kinds of general studies on variation are not part of the DGS Corpus Project itself. Within the project, only variation of individual signs is analyzed as far as this information is needed for the compilation of a dictionary entry such as the sign's regional distribution or sign use restricted to certain age groups.

Since annotation is currently in progress, analyses on regional distribution of signs from the corpus cannot be based on large amounts of empirical data yet and therefore can only be preliminary. To gain practical experience in dealing with widespread variation spot annotations of color signs filmed during the task *elicitation of isolated signs* are being used as a testing ground for analysis procedures.

## 2. Elicitation Method

One of the two elicitation tasks specifically aimed at eliciting regional variation is the *elicitation of isolated signs* (cf. Nishio et al., 2010). The goal was to elicit signs for a small number of selected concepts from a large number of informants. In this task concepts that were known to exhibit a high variation in DGS were presented as written words, some of them also in combination with a picture. Informants were asked to produce their signs for these concepts. Eleven colors (red, blue, yellow, green, orange, purple, pink, brown, black, white, gray) were presented on the screen as unicolor plane without written references to the concepts. Informants were asked to name these colors.

## 3. Sample Size

One informant of each pair (i.e. 165 informants) was asked for his/her color signs in the task *elicitation of isolated signs*. For preliminary analysis raw data from 156 informants of 12 regions available were transcribed resulting in 2052 tokens for colors. This included the tokens from the spot transcription[1] of the *isolated signs task* and tokens that have already been annotated within other parts of the corpus material. The movies from the

[1] Spot transcriptions for this study were made by Nele Groß, Ilona Hofmann, Lutz König and Gabriele Langer. Technical support was provided by Sven Wagner.

last region (Leipzig) and a few movies from other regions had not been available for transcription at the time and could therefore not be included. Even though the sample size is rather large it is still too small to gain a clear picture of regional distribution for all variants, especially since other factors like schooling might have a greater influence on variant use than the actual place of living. However, the preliminary results show some interesting tendencies of regional distribution. Within the DGS Corpus project a web-based feedback function (technical term: voting) is planned and in the future will provide further information to be included in the analyses of regional distribution of signs.

## 4. Annotation

The data of this study have been annotated in a very basic way with the transcription tool and integrated database of iLex (Hanke, 2002; Hanke & Storz 2008). Spot annotations have been carried out to identify different form variants for color signs. All variants have been described by separate type entries regardless of whether they would be considered phonological or lexical variants. Forms e.g. with a clearly extended thumb constituted new type entries in iLex whereas small deviations of form that have been known to occur frequently with certain handshapes (such as small differences of thumb position or more or less spreading or bending of fingers) or that seemed to be either idiosyncratic or accidental did not constitute new type entries. Instead these minor differences were noted with the token (i.e. in the token tag) as form deviations from the citation form of the type. When the number of tokens with the same deviation within a type entry is increasing they can be re-categorized at a later stage of the annotation process called lemma revision (cf. Konrad, 2011 pp. 93-96; König et al. 2010). Also, some kinds of variation that have led to separate entries in one case (such as one-handed vs. two-handed) have been subsumed under one entry in other cases with qualifications or token deviations noted.[2] This is to say, the data is still somewhat messy as it has yet to undergo the lemma revision process. Therefore the categorizations of this study are preliminary. It is expected that some form types will be merged into one while others (for example BLACK1) may be separated in two or more types on grounds of the distributional data of form variation so far considered as minor. For this preliminary analysis of regional distribution all variants have been annotated and analyzed separately focusing on the variants with the highest number of tokens (9 tokens or more) and leaving out variants with a lower number of tokens. The point of this preliminary study is to show that even with corpus data that is not completely consistent yet analyses of distribution can provide some useful insights that may even support the decision-making process of re-categorizing the data.

## 5. Analysis of Distribution

### 5.1 Regionality of Informants

One requirement for the selection of informants was their rootedness within a given region. Only lifelong or at least long-term residents of a region were accepted as informants. Preferably the informants should have grown up and currently have their permanent residence within the region. A residency of at least ten years within the region was also accepted. Metadata of the informants include the place of living, the place of growing up, the school they attended and all other places the informants had been living at for a longer period of time.

Three informants who had recently moved away were nevertheless accepted for their original region. In this case the current place of living did not coincide with the prominent regional linguistic affiliation of the informant. For these informants their last residence within the original region has been used for the preliminary analysis of regional distribution.

### 5.2 Displaying Regional Distribution

This first preliminary study is based on the place of residence of the informants. The distribution of the most frequent color signs (9 tokens or more) was matched onto the map of Germany with a resolution at the county level. For this each informant's place of living was matched to the corresponding county and the county coding (corresponding to the GADM dataset for Germany[3]) was stored as metadata to the informant within iLex. By an SQL query all county codes with an attested sign use for a certain sign were extracted from iLex. All counties with attested sign use were then colored to show the regional distribution of the sign in question. The data exported from iLex were fed into the statistical analysis program *R* using the packages *maps* and *sp* and the GADM dataset for Germany to produce the maps.

The maps displaying the attested use of a specific sign are a result of the described procedure and directly driven by the data from the corpus, combining metadata (place of living) and annotation data. Maps can either

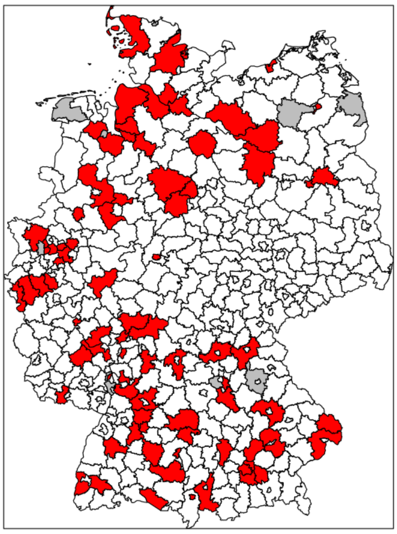
[2] In the DGS Corpus Project the iLex database and working environment is used for annotations. The database contains large amounts of annotated data and type entries from previous projects. Each project had used somewhat different annotation rules. Annotation guidelines, structures and procedures for the DGS Corpus Project are still being developed. To draw on type entries from previous projects is a huge advantage but also constitutes a challenge for the consistency of rule application. While the number of hands had often constituted new type entries in the past the number of hands are now being annotated by qualifier structures implemented in iLex (see Konrad et al. 2012, this issue). This is the reason why for some color signs there still exist separate entries for one-handed and two-handed variants while for others this kind of variation is already marked by qualifiers within the same type. Re-categorizing old entries and tokens following new annotation rules and structures will take some time and effort and will happen step by step as new rules are being developed and implemented and more and more sign entries go through the lemma revision process.

[3] http://www.gadm.org/

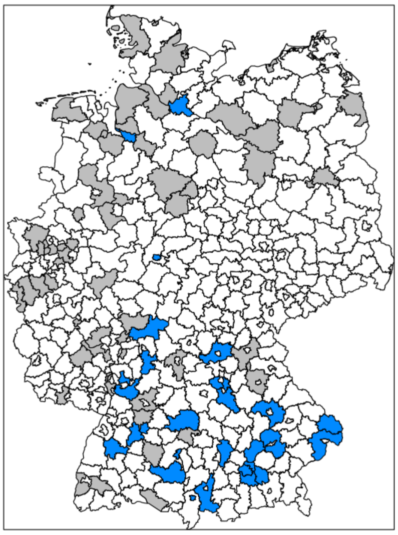
show the distribution of one particular sign as in map 1 (RED1) and map 2 (BLUE3) or the distribution of two (map 3) or a number of variants in relation to each other (maps 4 and 5).

In the future the described procedure can be implemented to automatically produce distributional maps of selected signs on command and thus provide a quick overview to support lemma revision and the compilation of dictionary entries.

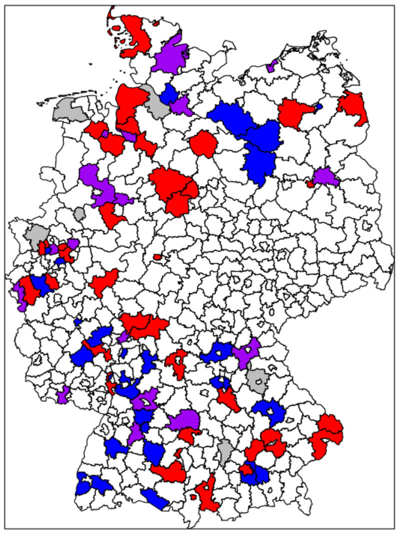
### 5.3 Distributional Maps



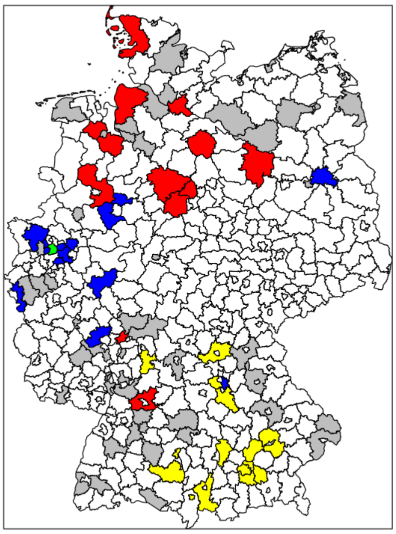
Map 1:
RED1



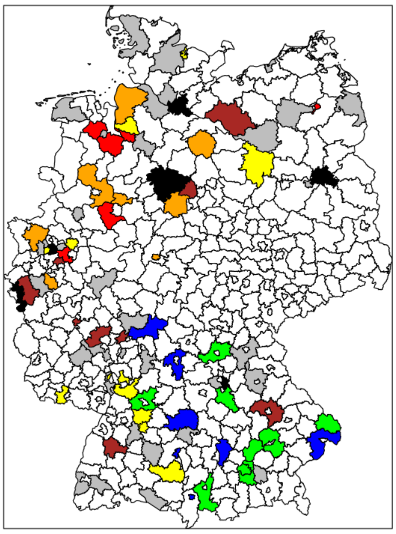
Map 2:
BLUE3



Map 3:
BLACK1 (red),
BLACK2 (blue)



Map 4:
GREEN2 (blue),
GREEN3 (red),
GREEN9A (yellow)



Map 5:
six variants for *brown*
(glosses and HamNoSys for these variants are listed on the right)

The maps are based on data of 156 informants from 90 counties. Counties without informants are colored white, counties with informants but no attested sign use are colored grey.
Maps 1 and 2 show the distribution of a single variant. Map 3 contrasts the use of two variants while map 4 contrasts the use of three variants. In both maps overlapping areas of use are marked by the corresponding mixed color, e.g. areas of overlapping use for BLACK1 (red) and BLACK2 (blue) are colored purple (map 3).

Map 5 shows the distribution of the following six variants for the color *brown*:
BROWN2A (red),
BROWN029 (orange),
BROWN7 (yellow),
BROWN8 (brown),
BROWN9 (blue),
BROWN4 (green),
(overlapping areas of use: black)

### 5.4 Results

For this study 2052 tokens of color signs from 156 different informants of 90 counties in 12 regions were annotated and matched to 256 types. For 117 of these types only one token was found, 45 types had 9 tokens or more. Only these more frequently used types were analyzed for regional distribution. They accounted for 75%

of all tokens. (For an overview on the numbers of types and tokens see table 1).

| mouthing: purple*: lila; purple**: violett; pink*: rosa; pink**: pink | number of types (variants) | number of to-kens | types with one token | types with 9 or more tokens (A) | number of to-kens of (A) | For (A): % of all tokens |
|---|---|---|---|---|---|---|
| blue | 23 | 173 | 8 | 6 | 138 | 80% |
| brown | 34 | 161 | 16 | 7 | 93 | 58% |
| yellow | 32 | 192 | 18 | 5 | 152 | 79% |
| grey | 47 | 169 | 19 | 5 | 72 | 43% |
| green | 39 | 182 | 19 | 4 | 98 | 54% |
| purple* | 23 | 174 | 5 | 3 | 126 | 72% |
| purple** | 2 | 7 | 0 | 0 | | |
| orange | 21 | 177 | 12 | 5 | 153 | 86% |
| pink* | 26 | 160 | 10 | 3 | 107 | 67% |
| pink** | 6 | 11 | 3 | 0 | 0 | 0% |
| red | 4 | 163 | 1 | 1 | 154 | 94% |
| black | 4 | 310 | 0 | 2 | 298 | 96% |
| white | 13 | 167 | 5 | 4 | 148 | 89% |
| beige | 1 | 5 | 0 | 0 | | |
| turquoise | 1 | 1 | 1 | 0 | | |
| | 276 | 2052 | 117 | 45 | 1539 | 75% |

Table 1: Number of types and tokens for colors

Results of this preliminary study show that there is a lot of variation in color signs in DGS. Even though the data still has to undergo the lemma revision process it nevertheless can already be used to visualize tendencies of distribution. Five examples of distributional maps for selected color signs are included in this paper. The maps show that RED1 (map 1) is used all over Germany (as far as data was available for these areas) while BLUE3 (map 2) is primarily used in Southern Germany. BLACK1 and BLACK2 (map 3) both seem to be used in all areas of Germany. The overlap areas of attested use are marked by the corresponding mixed color (in this case purple as the mixture of red for BLACK1 and blue for BLACK2). A closer investigation of the form deviations of BLACK1 may bear interesting results as a variant with slightly spread and bent fingers appears to be used in Southern parts of Germany. Map 4 is an example of a very clear regional distribution of three lexical variants for *green* (GREEN2, GREEN3 and GREEN9A). Map 5 shows the distribution of 6 variants for *brown*. Here overlap areas are colored black. Maps 3, 4 and 5 all indicate that there might be a distinct dialectal area in Southern Germany while dialectal areas in other parts of Germany cannot be seen as clearly from these few analyses. It will be very interesting to look at signs from other domains and also from the data elicitation region of Leipzig to get a clearer picture of dialectal regions of DGS in Germany.

### 5.5 Limitations of the preliminary study

This preliminary study has a number of limitations. The analyzed sample does not include data from all regions and informants yet. The informants filmed at Leipzig (from an area covering the Southern part of former East Germany) are not included. Also in other data collection tasks further tokens of color signs will occur that have not been transcribed yet. More data is needed to stabilize the findings and to fill the gaps.

All annotations for this preliminary study have to undergo lemma revision. Within this review process some variants will probably be divided into different subvariants. For example, the deviation information of the tokens of BLACK1 indicate that there may be at least one subvariant that is consistently used in the south. Other forms (especially forms with only one or few tokens) might be re-categorized as deviations of other variants thus reducing the number of variants for the associated color. This is to say that the results presented in this paper indicate tendencies but are to be received with caution and not to be taken as final results.

The chosen geographical display of regional distribution has also some limitations. Berlin has been treated as one area (county), but for historical reasons should be divided into an Eastern and Western part to be able to analyze effects of the division of Berlin from the 1960s to the 1980s on sign distribution in that area. Some recent changes of administrative areas (counties) are not included in the GADM dataset and one county is completely missing. For future implementation of this procedure a more complete and up to date dataset has to be used.

The number of tokens or the number of different informants per sign and county respectively are not displayed on the distributional maps yet. Including this information would show the central areas of use more clearly. Improved versions of distributional maps should also indicate overlap areas more clearly.

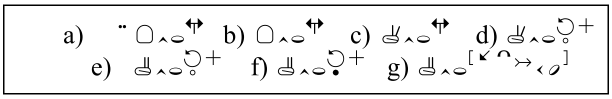
Other regional influences than the place of living should be taken into account. See section 6.3 for a suggested approach to this issue.

Sociolinguistic variables other than region should also be investigated and put into relation to regional factors as it was done in other projects on sociolinguistic variation in signed languages (cf. for example Lucas et al. 2001; McKee & McKee, 2011; McKee et al., 2008; Schembri et al., 2009). As this is not part of the DGS Corpus Project, this issue awaits further research.

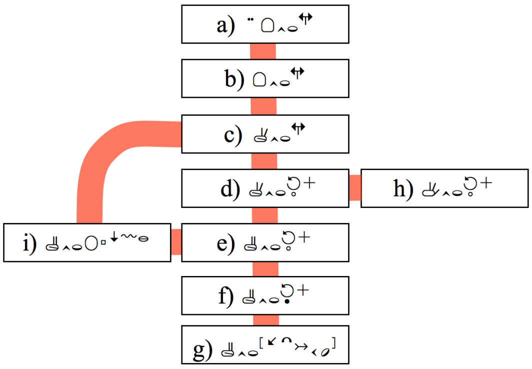
## 6. Issues of Procedure and Research

### 6.1 Lexical and phonological variation

In the annotation and analysis of variants, usually lexical variation (phonologically unrelated forms, that is, distinct signs) is distinguished from phonological variation (phonologically related forms of the same basic sign). Two similar sign forms are generally treated as phonological variants (also called subvariants) rather than lexical variants when they differ in only one parameter from each other, such as handshape or movement or place of articulation (cf. for example Lucas et al., 2001 p. 180; Johnston, 2003 p. 349; McKee & McKee, 2011 p. 502; Hollman & Sutrop 2010 p. 141). However, the distinction between phonological and lexical variation is not always as clear as it might seem on first glance. Sometimes there exist chains of sign forms where each sign form differs from its neighbors only slightly in one formational feature, so that direct neighbor(s) in the chain would usually be considered phonological variants of each other, while the signs at the distant ends of such chains may not have much in common with each other and would usually not be analyzed as phonological but rather as lexical variants of each other (see example 1 for a chain of partly similar forms used for *blue* in the DGS Corpus data).



Example 1:
A chain of partly similar forms used for *blue*



Example 2:
Partly similar forms used for *blue* (branching chain)

Examples 1a and 1g seem to be totally unrelated sign forms and differ with respect to number of hands, handshape and movement. However, in between these two signs other forms exist where each sign in the chain differs from its neighboring signs with respect to only one formational feature: a to b: number of hands, b to c: handshape, c to d: movement, d to e: handshape, e to f: size of movement, f to g: shape of movement (arc instead of full circle with an additional change of orientation making the arc anatomically more comfortable). Even if for this reason 1g would be considered distinct from the other forms, the same point could be made focusing on a and f. To complicate things further, chains may also branch off and possibly reconnect (see example 2).
This example shows that distinguishing phonological from lexical variants cannot be based on the formational similarity of the sign forms alone. König et al. (2008, p. 394) suggest to take into account the underlying image and the image producing technique of signs when determining whether two similar forms are phonological variants of the same sign (based on the same underlying image, produced by the same technique) or independent lexical variants (different underlying images and/or techniques). This can be helpful when dealing with iconic signs, but it cannot be applied when the signs in question either lack iconicity or when their underlying image cannot be determined, as it is the case for many color signs in DGS.
In the case of this study one-handed and symmetrical two-handed productions were often treated as the same sign (example 1a and b), as well as certain differences in the spreading of fingers that often occur in signs with specific handshapes (example 1d and 1e, also flat and slightly spread fingers for the B-handshape in BLACK1). Frequency of occurrence can be taken as an additional criterion for grouping tokens into separate entries. Frequently attested forms were treated as separate entries while others that had only one or a few tokens used by only one or few signers were either interpreted as idiosyncratic deviations of another form (for example 1f was interpreted as instantiations of 1e with the deviation of an enlarged movement) or they have been omitted in the overall analysis because their number of tokens was too small.
The analysis of regional distribution of very similar forms may reveal whether they are different phonological variants of the same sign used in the same region or two dialectal variants used in different regions. Thus data-driven distributional maps as introduced in this paper may aid the annotation process itself by providing clues for categorizing or re-categorizing certain form variants into one or separate entries of the lexical database used as a basis for annotations. For the lexicographical description of individual signs these analyses are also very helpful. Phonological variants with the same distribution might better be treated in one common sign entry in the dictionary covering these forms and describing the range of the variation while it would be more user-friendly to produce two separate sign entries for dialectal variants. Distributional maps can also support practical lexicographic work for identifying and describing the use of individual signs and some smoothed-out version of the maps could even be included as a visual hint on the distribution of the given sign.

## 6.2 Multiple Regional Influences

Depending on where DGS was acquired the place of growing up or living might not be the strongest or the only regional influence on the signing of a particular informant. For example, it was reported for many sign languages that residential schools have a strong influence on the signs a signer uses (cf. for example Lucas et al., 2001; Schembri at al., 2009; Schermer, 2003). Studies on regional variation of spoken languages usually only include informants who have lived all of their lives in one place/area. For sign languages it is rather unlikely that a sufficient number of such signers can be found and recruited. Therefore also signers with a long but not a livelong residence in the specific area are accepted as informants, even though their signing may show influences of different regions.

When several geographical data have been collected on each informant it could be attempted – provided the sample size is large enough – to take different geographical influences on a particular informant into account for an analysis of the distribution of a certain sign. This could be done by comparing the different geographical regions attributed to the informant to the overall regional distribution of the given sign form by other informants and identifying the most plausible regional influence for the given signer and sign. In the next section (6.3) a procedure for this kind of analysis is outlined. This type of analysis will become especially useful when dealing not only with corpus data but also with data collected through the public feedback gathered at a later stage of the project.

The public web-based feedback function will supplement the data from the corpus. Within this feedback function members of the sign language community are asked to participate and answer questions on the signs presented there. The feedback will include information such as whether the participant knows and/or uses a particular sign or not. One has to register in order to participate. Registration will include some geographical information about the participants such as place/region of living and possibly other geographical information like place/region of schooling or place/region of growing up. It is expected that a number of participants have been living in several different regions and that each of these may have influenced their signing and their knowledge of signs.

### 6.3 Dealing with multiple regional influences: proposed procedure

Here an analytic procedure is outlined of how to take multiple potential regional influences on one informant into account for regional analysis of a particular sign. This outline is meant as a contribution open for discussion as it is work in progress and has not yet been implemented or tried out. The idea is that a particular informant may have several regions that potentially influence his or her signing, for example region of growing up, region of schooling, region where his or her deaf parents come from, different regions of long-term residence, long-term stay abroad and so on. In this paper these regions are called potential regional influences (PRI). All PRIs of an informant have to be known and matched to a geographical area. They also have to be categorized for their kind (e.g. permanent residence, place of schooling, place of growing up and so on). Provided enough data is available from many other informants using the same sign it should be possible to identify the most probable regional influence (MPRI) of the given PRIs for the use of this particular sign by comparing the PRIs to the attested regional distribution of the sign.

The analysis procedure can be described as follows:

Step 1: As basis for the comparison all areas of interest (for example all counties of Germany[4]) are given a value for the sign in question – depending on how many tokens of the sign from how many different informants are attested and attributed to this area. I will call this set of values for each area a-values. All PRIs of all informants are to be taken into account for this a-value calculation for a particular sign. When one informant has three PRIs attributed to him/her and uses a certain sign, then this contributes to the a-value of all three PRIs (e.g. counties). Areas with many tokens from many different informants receive a high a-value (e.g. 4), areas with few tokens from only few different informants receive a middle a-value (e.g. 3), areas with tokens by only one informant receive an a-value of 2 and areas that have no tokens but are neighboring a high or middle score area receive a low a-value (e.g. 1).[5] All other areas receive the a-value of 0. All areas with an a-value above 1 are called attested areas, all areas with the value 1 are called neighboring areas. Threshold values need to be defined for this categorization as high or middle score attested area. The threshold values can be adapted to the number of overall tokens of the sign.

Step 2: The a-values are taken as basis to determine the most probable PRIs for all informants and their tokens. Now all PRIs of each informant in question are compared to the a-values of the areas and the most probable area of influence for this sign may be determined by the following rules:

- a) The PRI area that has the highest corresponding a-value is the most probable influence for the use of the sign in question.
- b) When two or more PRI areas have the same corresponding a-value, the PRI area with the highest priority on a priority list (see below) is chosen as the most probable.
- c) When no PRI area has a corresponding a-value above 2, then the PRI area with the highest priority on a priority list (see below) is chosen as the most probable.

In order to resolve cases where two or more PRIs have the same value (see above case b and c) a priority list has to be defined that ranks the kinds of geographical areas (for example: area of growing up is favored over area of only two years of residence). This list ensures that for each sign and informant exactly one area of the PRIs can be chosen as the most probable even if there are only few tokens available or if none of the PRIs of the particular informant overlaps with the PRIs of other informants.

Once the most probable area (MPRI) has been determined for a given sign and informant of his or her PRI areas, all tokens of this sign by this informant are attributed to the determined MPRI.

---

[4] As we do not have data from all counties it might prove more useful to broaden the granularity from counties to larger areas such as districts. In this case the procedure can be adapted accordingly.

[5] In addition, PRIs of informants with a livelong residence at one place and therefore only one PRI should rank higher than the PRIs of informants with several PRIs.

Step 3: The values of all areas (e.g. counties) are again determined. This is done on the basis of all identified most probable areas (MPRIs) only. This new set of values for all areas will be called b-values.
Step 4: The results of step 3 can be displayed on a map using different shades of colors for high, middle and low b-value areas.
The described procedure will consolidate the areas of attested sign use and filter out most accidental singular occurrences. Another advantage of this procedure is that competing signs for the same concept used by the same informant can be taken into account and analyzed separately. Other studies have used only the first response of an informant to a lexical elicitation task for analysis because it was considered “the signer’s default, spontaneous usage” (McKee & McKee, 2011 p. 499). However, it is likely that within a corpus of spontaneous signing one informant uses several competing variants without one variant being more spontaneous than the other. Each of these sign variants might be traced back to different PRIs by the described procedure.
Another idea is to take the results of this procedure (b-values) and automatically fill gaps between attested areas so that the result is one large area of use on the map rather than several isolated colored counties. This could be done on the basis of nearness of neighboring areas surrounded by attested areas. For this completion procedure competing forms (different regional variants used for the same concept) should be taken into account: When a presumed area of use is to be extended to a non-attested area on the basis of geographical nearness this should only be done when this area is not attested for another competing sign.

### 6.4 Lexicographical Perspective

In sign language variation studies regional distribution of lexical variants usually has been dealt with by taking sites or predefined regions as a starting point and collecting data to determine which signs are used for certain concepts there. Then results can be compared with regard to number of variants and subvariants and the overlap of use in the different regions can be investigated. Regions have been defined on grounds of presumed or known differences within the language communities, small pilot studies or presumed or known influences of different locations of residential schools. The point here is, that usually the analysis looks at predefined regions and the use of signs therein.
In this study, the direction of focus has been turned around to facilitate a lexicographical perspective on regional distribution. The individual sign is the starting point of the analysis and the target of investigation is where exactly this particular sign is being used. This can be done without relying on predefined larger dialectal areas. The corpus data can speak for itself. It reveals the relevant areas of use for each sign through distributional maps produced directly from the corpus. This type of information is useful when writing a lexicographical description of signs in dictionary entries.

### 6.5 Dialectal Regions

The geographical boundaries between areas of use of different regional lexical variants for the same concept are called isoglosses. Corresponding isoglosses of several sets of signs with similar distributional patterns can be taken as indications of boundaries of dialectal regions. This is not only the case for lexical variants but also for all kinds of linguistic variables that display comparable patterns of regional distribution. Distributional maps cannot only be produced for the distribution of lexical variants but also for the distribution of other kinds of variation. The same procedure used here for the analysis of occurrences of signs can be adapted to occurrences of other phenomena coded and annotated in the corpus data.

### 6.6 Implications for Research on Color Signs

The elicitation of colors in the task *elicitation of isolated signs* was designed to gain data on lexical variation across regions, it was not intended to for a study on basic color terms in DGS. With the exception of one color-blind informant all informants were able to spontaneously give their color signs, some of them showed more than one variant (which were all included in the study). In few cases informants were unsure about the color presented, in three cases informants misinterpreted orange for beige. This might be due to the selection of the particular color as stimulus, lightning conditions at the site or the vision of the informants. The very high number of tokens for black (cf. table 1) can be explained by the elicitation setting. A black screen was used to elicit the color black and at the end of the task a black screen appeared to signal the end of a task in the same way as in other tasks. Most informants reacted to this black screen showing their sign for black again. Only in few cases an informant used the same manual sign form with different mouthings to name different colors. The most commonly used sign was RED1, which was used by almost all signers across the country with very few exceptions. For black (2 main variants), purple (3 main variants) and white (4 main variants) only few stable variants were found while a high number of variants were found for grey, green, brown and yellow. Some signs were used for more than one color.
There does not exist one single set of color signs for DGS as a whole. The observed high variation and complex distributional patterns of signs for colors in DGS might present a challenge for the research on basic color terms at the present state of research. Several combinations of regional variants that overlap to various degrees have to be taken into account for future studies on color signs.

## 7. Conclusion

The preliminary analysis of regional distribution of color signs from the DGS Corpus is one example of the many ways an annotated corpus can be utilized. Maps showing the regional distributions of tokens of sign variants can be generated directly from the annotations stored in a database together with lexical entries and relevant geographical data (metadata) on informants, as it is done in the iLex database and working environment. The visualization of the data on a geographical map provides a quick overview on regional distribution and can thus support the annotation and lemma revision processes as well as be a valuable tool for describing signs and their use in dictionary entries. Naturally, the results of such visualizations depend on the quality and consistency of the annotations and the existence of relevant geographical metadata on informants. First analyses of the signs for colors confirms the expectation that in DGS there is a high degree of variation in color signs and that a certain extent of these variants can be shown to be regional variants.

## 8. 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.

## 9. References

Hanke, T. (2002). iLex. A tool for Sign Language Lexicography and Corpus Analysis. In González Rodriguez, M. & Paz Suarez Araujo, C. (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.

Hanke, T., Storz, J. (2008). iLex - A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In Crasborn, O., Efthimiou, E., Hanke, T., Thoutenhoofd, E.D., Zwitserlood, I. (Eds.), *LREC 2008 Workshop Proceedings. W 25: 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*. Paris: ELRA, pp. 64-67. [Online available: http://www.lrec-conf.org/proceedings/lrec2008/workshops/W25_Proceedings.pdf].

Hollman, L., Sutrop, U. (2010). Basic Color Terms in Estonian Sign Language. In *Sign Language Studies*, 11(2), pp. 130-157.

Johnston, T.A. (2003). Language Standardization and Signed Language Dictionaries. In *Sign Language Studies*, 3(4), pp. 431-468.

König, S., Konrad, R., Langer, G. (2008). What's in a sign? Theoretical Lessions from Practical Sign Language Lexicography. In J. Quer, (Ed.), *Signs of the Time. Selected Papers from TISLR 2004*. Hamburg: Signum, pp. 379-404.

König, S., Konrad, R., Langer, G., Nishio, R. (2010). How Much Top-Down and Bottom-Up do We Need to Build a Lemmatized Corpus? Poster presented at the *Theoretical Issues in Sign Language Research Conference (TISLR 10), Sept 30 - Oct 2, 2010 at Purdue University, Indiana, USA*. [Online available: http://www.purdue.edu/tislr10/pdfs/Konig Konrad Langer Nishio.pdf].

Konrad, R. (2011). *Die Erstellung von Fachgebärdenlexika am Institut für Deutsche Gebärdensprache (IDGS) der Universität Hamburg (1993-2010).* Corrected Version. [Online available: http://www.sign-lang.uni-hamburg.de/projekte/mfl/konrad_2011_fachgeblexika.pdf].

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; this issue.

Lucas, C., Bayley, R., Valli, C. (2001). *Sociolinguistic Variation in American Sign Language*. Washington, D.C.: Gallaudet Univ. Press.

McKee, R., McKee, D. (2011). Old Signs, New Signs, Whose Signs? Sociolinguistic Variation in the NZSL Lexicon. In *Sign Language Studies*, 11(4), pp. 485-527.

McKee, D., McKee, R., Major, G. (2008). Sociolinguistic Variation in NZSL Numerals. In de Quadros, R.M. (Ed.), *Theoretical Issues in Sign Language Research 9. Sign Languages: spinning and unraveling the past, present and future. TISLR9, forty five papers and three posters from the 9th Theoretical Issues in Sign Language Research Conference, Florianopolis, Brazil, December 2006*, pp. 296-313.

Nishio, R., Sung-Eun 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 22.-23.05.2010*, pp. 178-185. [Online available: http://www.sign-lang.uni-hamburg.de/lrec2010/lrec_cslt_01.pdf].

Schembri, A., McKee, D., McKee, R., Pivac, S., Johnston, T., Goswell, D. (2009). Phonological variation and change in Australian and New Zealand Sign Languages: The location variable. In *Language Variation and Change*, 21, pp. 193-231.

Schermer, T. (2003). From Variant to Standard: An Overview of the Standardization Process of the Lexicon of the Netherlands over Two Decades. In *Sign Language Studies*, 3(4), pp. 469-485.

# CUNY American Sign Language Motion-Capture Corpus: First Release

**Pengfei Lu**
The City University of New York (CUNY)
Computer Science Program, Graduate Center
365 Fifth Ave, New York, NY 10016 USA
E-mail: pengfei.lu@qc.cuny.edu

**Matt Huenerfauth**
The City University of New York (CUNY)
Computer Science Department, Queens College
65-30 Kissena Blvd, Flushing, NY 11367 USA
E-mail: matt@cs.qc.cuny.edu

**Abstract**

We are in the middle of a 5-year study to collect, annotate, and analyze an ASL motion-capture corpus of multi-sentential discourse. Now we are ready to release to the research community the first sub-portion of our corpus that has been checked for quality. This paper describes the recording and annotation procedure of our released corpus to enable researchers to determine if it would benefit their work. A focus of the collection process was the identification and use of prompting strategies for eliciting single-signer multi-sentential ASL discourse that maximizes the use of pronominal spatial reference yet minimizes the use of classifier predicates. The annotation of the corpus includes details about the establishment and use of pronominal spatial reference points in space. Using this data, we are seeking computational models of the referential use of signing space and of spatially inflected verb forms for use in American Sign Language (ASL) animations, which have accessibility applications for deaf users.

**Keywords:** American Sign Language, motion-capture, corpus, sign language, spatial reference.

## 1. Introduction

Software to generate American Sign Language (ASL) animations can provide benefits for the significant number of deaf individuals in the United States with relatively low written English literacy. Our research goal is to improve technologies for generating ASL animations through the collection and analysis of a motion-capture corpus of ASL multi-sentence discourse. Our intention is to provide the research community with a sufficient-quality corpus for their future study on ASL linguistic phenomena and to conduct our own analysis of this corpus using statistical modeling and machine learning techniques to create models useful for generating grammatically correct ASL animations.

Signers associate entities under discussion with 3D signing space locations, and signs whose paths or orientations depend on these locations pose a special challenge: They are time-consuming for users of scripting software to produce, and they are not included in the repertoire of most ASL generation/translation software. Our goal is to construct computational models of ASL that could be used to partially automate the work of human authors using scripting software or to underlie generation/translation systems.

Section 2 provides basic information and statistics about the released portion of our corpus. Section 3 describes the linguistics of spatial reference points and inflecting verbs. Section 4 describes the recruitment and prompting strategies we have used to elicit signing performances of the desired form. Section 5 describes the motion-capture equipment, motion-capture data recording, and post-processing. Section 6 describes the annotation process. Section 7 describes a sub-corpus we have collected of ASL inflecting verbs. Section 8 contains conclusions and future research plans.

## 2. The Released Corpus

After the three years of data collection, we have gathered 246 ASL unscripted multi-sentence single-signer passages from 9 native signers, each signer came to the lab for one recording session on a different day (Lu & Huenerfauth, 2010). While we have recorded and begun annotation on a total of 215 minutes of ASL motion-capture data thus far, we are ready to release to the research community the first sub-portion of our corpus that has been checked for quality.

This paper is the first announcement of this corpus release, which includes 98 passages performed by 3 native signers. The data includes Autodesk Motion Builder files of the motion-capture recording, BVH files (another commonly used file format for motion-capture data), high-resolution video recordings, and annotations for each passage. The annotations are in the form of SignStream™ files (Neidle et al., 2000) and plaintext files. Figure 1 shows two screenshots of videos of a signer in our corpus (a front view and a side view), a screenshot of a visualization in Autodesk MotionBuilder of the motion-capture data recorded (a virtual human body whose movements are driven by the data), and a visualization of the joint locations and orientations as recorded by the sensors (yellow dots on bottom right).

In addition to our primary corpus containing unscripted multi-sentential passages, we are also releasing a small sub-corpus containing several hundred instances of eight ASL inflected verbs (discussed in section 7). As this is our first corpus release, we welcome feedback from other researchers on how to best organize and release this corpus so that it is most useful. Future releases of our corpus may contain revisions of the data formats or annotation for this sub-corpus and additional passages not yet released. Our lab website contains details about accessing the corpus: http://latlab.cs.qc.cuny.edu/

## 3. Pronominal Spatial Reference Points and Inflected Verbs

Signers frequently associate entities under discussion with locations in the signing space involved in later pronominal reference and other purposes (Liddell, 2003;

Meier, 1990; Neidle et al., 2000). Various ASL constructions can be used to establish a spatial reference point (SRP) for some entity. While sign languages used around the world are not mutually intelligible, they do share certain key linguistic aspects – the use of spatial reference and verb inflection. All of these phenomena involve the use of the 3D space around the signer (often called the "signing space") to represent entities under discussion. During a conversation, signers often associate people, concepts, or other entities under discussion with 3D locations around their bodies. For example, by pointing at a location in the surrounding space at the beginning or at the end of a noun phrase mentioning a new entity, the human signer associates the entity referred to in the noun phrase with that location. Signers remember these spatial associations, and the movements of later signs in the performance may change based on these locations. When referring to one of these entities later in the conversation, a signer may use a pronoun sign (which also looks like a pointing gesture) aimed at the appropriate location in the signing space.

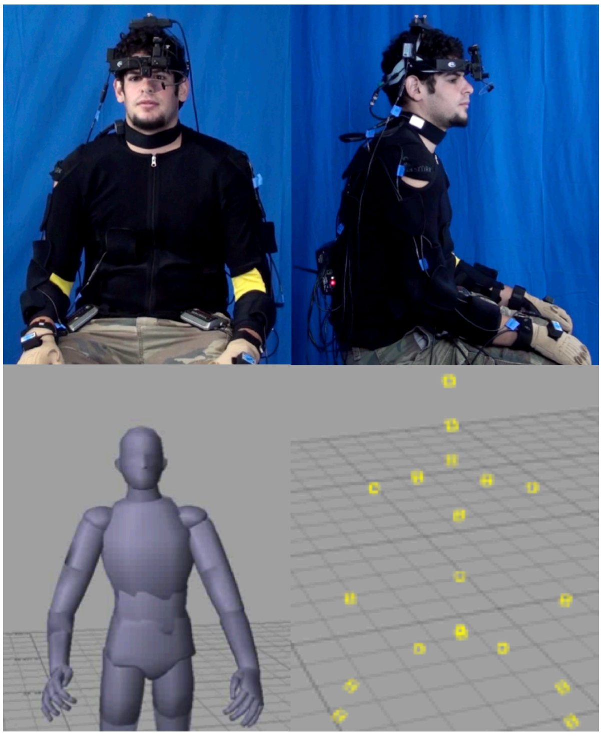


Figure 1: Screenshots of signer in the video recording and visualizations of the motion-capture data.

Some sign language verb signs also change their motion paths or hand orientation to indicate the 3D location where a spatial reference point has been established for their subject, object, or both (Liddell, 2003; Padden, 1988). Generally, the motion paths of these inflecting verbs change so that their direction goes from the subject to the object; however, their paths can be more complex than this. These verbs have been referred to by linguists as "inflecting verbs" (Padden, 1988), "indicating verbs" (Liddell, 2003), or "agreeing verbs" (Cormier, 2002). We call them as "inflecting verbs" in our research. Each inflecting verb has a standard motion path which is affected by the subject's and the object's 3D locations – producing a motion path that is unique to the specific verb, the specific 3D location of the subject, and the specific 3D location of the object.

In prior experimental studies, we determined that the use of spatially inflected verbs in an ASL animation significantly increased viewers' comprehension of the animations (Huenerfauth & Lu, 2012). However, most sign language animation generation software lacks sophisticated models of this phenomenon. A current focus of our research has been to develop computational models of how the motion-paths of inflecting ASL verbs change based on the 3D location in the signing space associated with the subject and/or object of the verb. During the construction of our ASL motion-capture corpus, in addition to the unscripted ASL passages in our released corpus, we also collected recordings of signers performing instances of spatially inflected verbs, and we also release some of those spatially inflected verb samples as a sub-corpus in this paper (discussed in section 7).

## 4. Recruitment and Elicitation

For the data recording sessions for our corpus, all instructions and interactions were conducted in ASL. Advertisements posted on Deaf community websites in New York City asked whether potential participants had grown up using ASL at home or whether they attended an ASL-based school as a young child. Of the 3 participants in the current corpus release: 3 grew up with parents who used ASL at home, 1 was married to someone deaf/Deaf, 3 used ASL as the primary language in their home, 3 used ASL at work, and 3 had attended a college where instruction was primarily in ASL. The signers were 3 men of ages 22-33 (mean age 25.7).

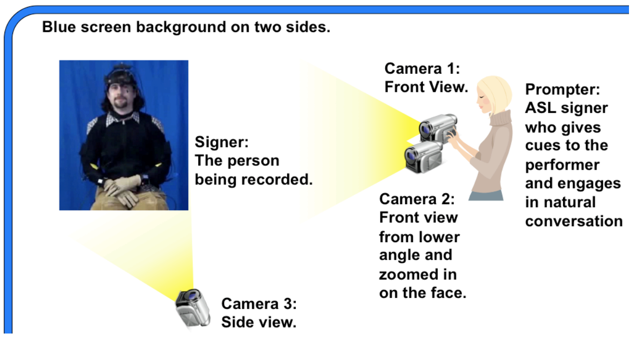



Figure 2: Diagram of an overhead view of our motion-capture studio setup.

Prior to data collection, a prompter who is an ASL signer engaged in natural ASL conversation sitting facing the signer being recorded (Figure 2); during data collection, the prompter gave a prompt to the recorded signer. All of our motion-capture recording sessions are videotaped to facilitate later linguistic analysis and annotation (details in section 5). Three digital high-speed video cameras film front view, facial close-up, and side views of the signer (Figure 2); a similar camera placement has been used in video-based ASL-corpora-building projects (Neidle et al., 2000). The front view is similar to the top left image in Figure 1, but it is wider. The facial close-up view is useful when later identifying specific non-manual facial expressions during ASL performances, which are essential to correctly understanding and annotating the collected data. Videotaping the session may facilitate "clean up" of the motion-capture data in post-processing, during which algorithms can be applied to adjust synchronization of different sensors or remove "jitter" or other noise artifacts from the recording.

To record natural ASL signing (with spontaneous and fluent use of spatial reference points in a multi-sentential single-signer discourse), we are collecting non-scripted passages; so, it has been essential for us to identify appropriate ways to prompt for the type of passages we wish to collect, to support our research. It is important for our research that we collect sign language passages in which signers establish different numbers of points in space to refer to people, places, or things under discussion (SRPs). Further, it is important that the passages do not contain too many classifier predicate (CP) expressions, which are a linguistic construction in ASL that also uses the space around the signer's body. CPs are not our current research focus, and because they lead signers to use space around their bodies in a different way than SRPs, we don't want to record stories that contain a lot of CPs, relative to the number of SRPs.

During our multi-year project, we have experimented with different forms of prompting strategies to elicit ASL signing in which signers establish different numbers of pronominal reference points in space, continuing signing for more time, and in which they do not make frequent use of CPs. Thus, the analysis of the different prompting strategies in one year of our project guided our data collection procedure for the following year. We identified the most effective prompts, and we stopped using some prompts with high CP/SRP ratios. We continued to use those prompts that led to long passage lengths, high number of SRP points established, and low CP/SRP ratios (Lu & Huenerfauth, 2011a; Huenerfauth & Lu, 2010a). The topics of the passages include signers discussing their personal histories, their recollection of news stories/movies, their explanation of encyclopedia articles, their opinion about a hypothetical scenario, their comparison of two persons or things, their description of a page of photos, and their recounting short narratives (Lu & Huenerfauth, 2011a; Huenerfauth & Lu, 2010a). Table 1 lists the prompts we used in the collection of this released corpus data, and brief description of each prompting strategy. Some of our prompting approaches involved showing pictures to a signer. Figure 3 shows an example of what a page of photos looked like for the "photo page" prompts.

Figure 3: Example of what a page of photos looked like for the "photo page" prompts.

| **Type of Prompt** | **Description of This Prompting Strategy** |
|---|---|
| News Story | Please read this brief news article (about a funny or memorable occurrence) and recount the article. |
| Compare (people) | Compare two people you know: your parents, some friends, family members, etc. |
| Compare (not people) | Compare two things: e.g. Mac vs. PC, Democrats vs. Republicans, high school vs. college, Gallaudet University vs. NTID, travelling by plane vs. travelling by car, etc. |
| Photo Page | Look at this page of photos (of people who are in the news recently) and then explain what is going on with them. |
| Personal Narrative | Please tell a story about an experience that you had personally. |
| Personal Intro/Info | Introduce yourself, describe some of your background, hobbies, family, education, etc. |
| Recount Movie Book | Recall a book you've read recently or a movie you saw, and then explain the story as you remember it. |
| Opinion / Explain Topic | Please explain your opinion on this topic (given) or explain the concept as if you were teaching it to someone. |
| Wikipedia Article | Read a brief Wikipedia article on some topic and then explain/recount the information from the article. |

Table 1: Types of prompts used.

Figure 4 lists how many passages we have collected using each of the different prompting strategies in this released corpus. The total number of passages of each prompt-type collected from each signer varies because the recording session was intentionally kept relaxed and conversational to promote more natural signing. Sometimes performers were verbose in their response to a prompt, but other times, they could think of little or nothing to say for a particular prompt. Further, since performers were recorded for only one hour (after the motion-capture equipment was set-up and calibrated), we rarely had sufficient time to try all of the different prompt-types during each performer's recording session.

This release of our corpus contains 9717 glosses in total (signer #1: 3962 glosses, signer #2: 3121 glosses, signer #3: 2634 glosses). The total length of video is 87.7 minutes (signer #1: 2048 seconds, signer #2: 1786 seconds, signer #3: 1426 seconds). The average number of glosses per passage is 54 (signer #1: 82 glosses per passage, signer #2: 53 glosses per passage, signer #3: 37 glosses per passage). The average video length of the passages collected is 99 seconds (signer #1: 158 seconds per passage, signer #2: 92 seconds per passage, signer #3: 68 seconds per passage). Figures 5 and 6 show histograms of passage length for each signer (measured in the number of signs performed or the number of seconds of the video recording). Figure 7 shows a sample of a transcript of a passage, in which the signer was elicited using the "Photo Page" style of prompt (Figure 3). Table 2 explains the notations we used for annotation in the transcript.

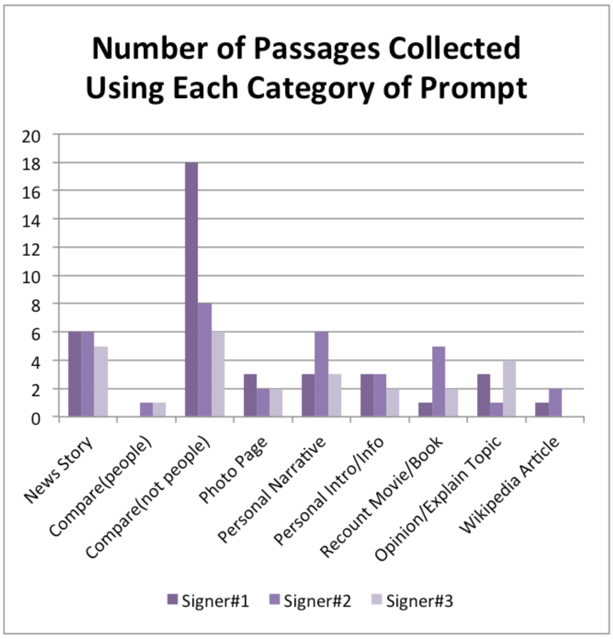



Figure 4: The number of passages in our released corpus that were collected using each category of prompt.

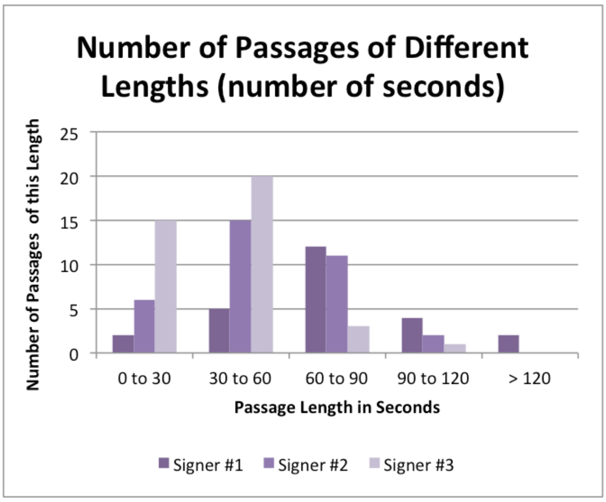



Figure 5: Length of ASL passages collected for each signer.

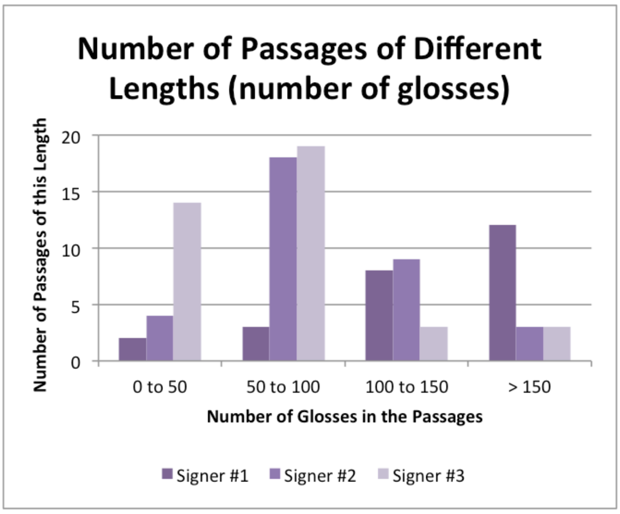



Figure 6: Number of glosses in the collected ASL passages.

PRESIDENT fs-OBAMA NOW IX-1-s:1
OUR PRESIDENT PRESIDENT fs-BUSH
RECENT FINISH THREE YEAR AGO
IX-1-s:2 FOURTY-THREE #TH
fs-OBAMA FOURTY-FOUR BOTH
DIFFERENT #WHAT BLACK IX-1-s:1
FIRST BLACK PRESIDENT WOW
IX-1-s:S LIKE fs-OBAMA BECAUSE
IX-1-s:1 DEMOCRAT …

Figure 7: A sample excerpt of a transcript of a passage.

| Type of notation | Explanation of this notation |
|---|---|
| fs-X | Fingerspelled word |
| IX-1-s:1 | Index sign (pointing), handshape-#1, singular, spatial reference point #1. |
| IX-1-s:2 | Index sign (pointing), handshape-#1, singular, spatial reference point #2. |
| #X | Lexicalized fingerspelled word. |
| IX-1-s:S | "I" or "ME": Index sign (pointing), handshape-#1, singular, signer/self. |

Table 2: The notations in the transcript in Figure 7.

## 5. Motion-Capture Equipment, Recording, and Post-Processing

As shown in Figure 2, three high-definition digital video cameras recorded front view, side view, and facial close-up views of the signer. The video in this corpus release has been separated into individual video clips for each passage; each clip has been trimmed from the full-length video recording of the entire data collection session. The start and end keyframes (marking the beginning and end of each passage) were identified by ASL native signers in our research group who watched the videos at the end of the recording session. To facilitate synchronizing the three video files (front, side, face close-up) during our post-processing, a strobe light was flashed once at the start of the recording session. Thus, as soon as the start- and end-times for each passage were identified in one of the three videos, it was straightforward to calculate the appropriate start- and end-times for the other two videos. All the videos are released as QuickTime MOV format files, of size 640x480, no audio, with a frame rate of 29.97 fps. If there is an interest from researchers in obtaining the original high definition video of the full recording session, this may be available in a future corpus release.

Since our goal in creating this corpus was to learn how to control the movements of an animated signing character, we knew that we would need to identify hand locations and joint angles of the human signer's body throughout the performance. Asking human annotators to write down 3D angles and coordinates from a video recording is time-consuming and inexact. Using computer vision techniques to automatically track the movements of a human's body in a video is also challenging due to the complex shape of the hands/face, rapid speed, and frequent occlusion of parts of the body during sign language. Thus, we chose to employ motion capture technology during the collection of our corpus, as a more reliable and accurate way to record a precise level of movement detail from a human sign language performance.

Full details about our equipment configuration have been previously described in (Huenerfauth & Lu, 2010a), but this information is briefly summarized here. For our corpus, we record handshape; hand location; palm orientation; eye-gaze vector; and joint angles for the wrists, elbows, shoulders, clavicle, neck, and waist. We use a novel combination of commercially available motion-capture equipment for this project, which includes: two Immersion CyberGloves®, an Applied Science Labs H6 head-mounted eye-tracker, an Intersense IS-900 inertial/acoustic tracker (for tracking the location and orientation of the signer's head, which is necessary for calculating an eye-gaze vector in a room), and an Animazoo IGS-190 bodysuit which uses a set of magnetic/inertial sensors.

To facilitate synchronization of the three videos and the data stream from the Animazoo IGS-190 body suit, we asked the signer in the motion-capture equipment to perform a very quick head movement (turning the head to one side) immediately after the strobe light was flashed at the start of the recording; this action was easily identifiable in the videos and the motion-capture data.

Our motion-capture data was recorded using Autodesk MotionBuilder, which is a general-purpose 3D animation program that enables the input of motion-capture data streams during a live recording session. For the convenience of future researchers viewing the motion-capture data recording, we have also inserted a virtual human figure into the MotionBuilder data file with body proportions that are based on the human signer being recorded. The body segments of this virtual human are linked to the data streams of the motion-sensors on the body suit. If we had not inserted such a virtual human figure into the MotionBuilder file, then the data stream recorded would merely consist of the individual location and orientation values of the sensors on the body suit. There would be no easy way for future researchers to quickly visualize of the sensor data. Of course, the raw sensor data is also accessible in the MotionBuilder file if researchers require it.

Since we do not use motion capture techniques to record facial expressions of the signers being recorded, the virtual human figure inserted into the MotionBuilder files does not have any facial details. In addition, we only record the upper body movement (from the hip joint upwards) of the human signers while they sit on a stool, so the position of the legs of the virtual human character in the MotionBuilder file is not meaningful. The eye-tracking data recorded from the signers will require additional post-processing by our research team, and it has not been included in this initial corpus release.

To minimize errors in the motion-capture data we recorded, we carefully calibrated the cybergloves worn by signers; details of the cyberglove calibration protocol we have designed for use in sign language recording projects appears in (Lu & Huenerfauth, 2009). Evaluations of the resulting hand motion-capture accuracy we achieve with the cybergloves is also included in (Lu & Huenerfauth, 2009).

The Animazoo bodysuit system requires information about the lengths of the body segments (the bone lengths) of the human being recorded; this data is needed so that the system can determine how the human is posed during a recording session based on the data from the sensors placed on each segment of the body. Prior to the recording session, we measured the body proportions of signer by photographing each of them while standing in a cube-shaped rig of known size. In this way, we obtained bone-length information for each signer (which can be determined from the resulting photographs using software which accompanies the Animazoo system).

While great care was taken in calibrating the various motion-capture equipment, there are still some errors in the body position that are visually apparent in the motion-capture data. For instance, sometimes when one of the human's hands touches the other, it is apparent that the hands of the virtual human character do not touch precisely. So, there are some retargeting errors in the motion-capture data stream, which future researchers using this data may need to further process, depending on their research goals. We may seek additional methods of cleaning-up and post-processing the collected motion-capture data for our future corpus releases.

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Sign Performed* | fs-OSAMABINLADEN | IX-1-s:1 | AMERICA | NUMBER | ONE | MOST | WANT | FIND | MAN | FINALLY | fs-US | CAPTURE | IX-1-s:1 |
| *SRP#1 Establishment* | | OSAMABINLADEN | | | | | | | | | | | |
| *SRP#1 References* | | e | | | | | | | | | | | r |

Figure 7: Example of a timeline from a passage from our corpus that contains an SRP.

The motion-capture data has been post-processed to adjust the timing synchronization of the motion-capture equipment. We found that it was challenging to perfectly synchronize the body movement data from the body suit and the hand movement data from the cybergloves, due to inconsistencies in the data transfer rate of the equipment and its small drift over time. To fix this timing issue, we asked researchers at our lab who are native ASL signers to watch the virtual human figures in our MotionBuilder files and to carefully edit (delay or advance) the timing of the glove data relative to the body suit data. In this way, we were able to verify that we have an accurate synchronization of the glove and body suit data streams; each of the recorded passages in our corpus was checked in this manner.

We are releasing our motion capture data in two files formats: FBX files and BVH files. FBX format files are the original file format owned by Autodesk and used by Autodesk MotionBuilder; this is the original recording file with the virtual human character (based on the human signer's body proportions) inserted into it. Next, we converted the FBX files into BVH files, which is a popular file format for 3D animation analysis and processing. BVH files are ASCII format files that contain two types of information: (1) a hierarchy of body segments sizes and joints for a figure and (2) rows of numerical data that correspond to information for all of the joints on a frame-by-frame basis. This corpus release also includes time mapping information between the motion-capture files and the videos for each signer.

## 6. Corpora Annotation

A team of native ASL signers (including students from deaf high schools in New York) annotated the data using the SignStream™ annotation tool (Neidle et al., 2000). The linguistic annotations for each passage have been cross-checked by at least two other native ASL signers on our research team. The long-term goal of our project is to annotate: sign glosses (with time alignment to the recorded video); part-of-speech of each gloss; syntactic bracketing (NP, VP, clause, sentence); and non-manual signals (role shift, negation, WH-word questions, yes-no questions, topicalization, conditionals, and rhetorical questions).

In addition, we annotate spatial reference points (SRPs) when they are established during a passage, which discourse entity is associated with each SRP, when referring expressions later refer to an SRP, and when any verbs are spatially inflected to indicate an SRP. These SRP establishments and references are recorded on parallel timeline tracks to the glosses and other linguistic annotations.

Figure 7 shows an example of a timeline from a passage from our corpus that contains an SRP; it is a timeline of an ASL passage discussing when Osama bin Laden was captured. In the example, the first time that the signer points to a location in 3D space around his body (glossed as "IX-1-s:1"), he establishes an SRP at that location to represent "Osama bin Laden." This SRP is referred to again later in the passage when the signer performs another "IX-1-s:1" sign. A loose translation of the passage in Figure 7 would be: "Osama bin Laden was America's No. 1 most wanted man; finally, the US captured him…"

Figure 7 shows the following rows of information:

- Row 1: Sign Performed: This row shows the sequence of glosses. While there is internal consistency in gloss labels used within our project, we have not employed a comprehensive system of "ID-glosses" like those of (Johnston, 2009). However, we may further standardize and edit our gloss notations in a future release of our corpus.
- Row 2: SRP#1 Establishment: This row indicates when the first spatial reference point ("SRP #1") is established by the signer somewhere in the signing space. When an SRP is established, then an annotation is added to this line with start- and end-times that align to the sign or phrase that established the existence of this SRP. The label of the annotation is meant to be a brief gloss of the entity referenced by this SRP. If there is a second SRP established in the signing space, then a new annotation row is added to the file for that additional SRP. Note that the integer after the colon at the end of the gloss "IX-1-s:1" indicates that the pointing sign is referring to SRP #1. A pointing sign directed at SRP #2 (if one were established) would appear as "IX-1-s:2". In this manner, each SRP is assigned an index number, and the gloss of each pronominal or verb-inflection reference to an SRP is marked with this index number (following a colon at the end of a gloss in the transcription).
- Row 3: SRP#1 References: This row indicates whenever a gloss or phrase in the passage references an SRP that has already been established in the signing space. Specifically, this row corresponds to SRP#1. On the first reference to a location for an SRP, this row receives an annotation with a label "e" (for "establishment"), and subsequent references to this SRP during the passage are indicated with an annotation added to this row with a label "r" (for "reference").

Figure 8 and Figure 9 illustrate the average number of SRP establishments (how many unique SRPs are established per passage) and the average number of SRP references per passage for each signer in our released corpus.

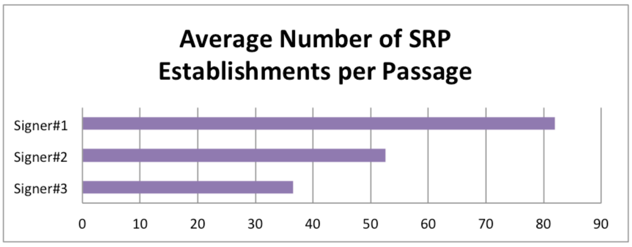



Figure 8: Average number of SRPs established per passage for each signer in our released corpus.

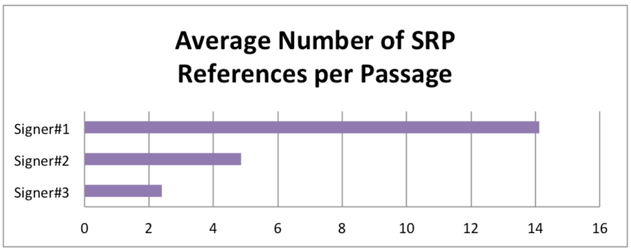



Figure 9: Average number of SRP references per passage for each signer in our released corpus.

In this first release of the corpus, we are distributing the time-aligned glosses, the annotation of the establishment of spatial reference points, and the English translation for each of the passages collected. We anticipate publishing additional layers of annotation in future corpus releases.

Text files of annotation information can be exported from the SignStream software, and we are including these plaintext versions of the annotation in this corpus release. The text files consist of all the annotation information and the file name of the video being annotated; each line in the file contains one type of annotation, such as glosses, SRP establishment, or SRP references. Each item of annotation is followed by its start and end frame numbers, corresponding to the video.

## 7. Sub-corpus of Inflected Verbs

A goal of our research on ASL animation is to design mathematical models of the movements of signers' hands during the production of inflected verbs (whose motion path and orientation is affected by how the SRPs for their subject and object have been set up in the signing space). In prior work (Lu & Huenerfauth. 2010b; Lu & Huenerfauth. 2011b), we needed larger numbers of examples of specific inflected verbs for all possible arrangements of subject and object in the signing space. This would not be possible to extract from our corpus because it is small in size and we would not be able to find all the possible combinations of each verb; so, we had to collect a special corpus of ASL verb movements.

For this corpus, we were only interested in obtaining information about the location and orientation of each of the signers' hands, not the information about head movement, eye gaze movement, or handshape. Thus, we used a motion-capture equipment configuration which was faster to set up, easier for the signer to put on, faster to calibrate, and easier to post-process. The trade-off is that less specific human body movement information was recorded, but this was sufficient for our ASL inflected verb research (Lu & Huenerfauth. 2010b; Lu & Huenerfauth. 2011b). Thus, we used our Intersense IS-900 system alone to record both head and hand data for our verb corpus. Previously, the IS-900 system was used only for head-tracking as part of our more complex equipment set-up for our unscripted multi-sentence corpus (section 5). The acoustical/inertial IS-900 system uses a ceiling-mounted ultrasonic speaker array (Figure 10) and a set of directional microphones on a small sensor to record its location and orientation.

For each verb, the signer was recorded performing it for different arrangements of the subject and object in the surrounding signing space. A set of color-coded squares were placed around the recording studio at various angles in a 180-degree arc in front of the signer; these targets were used as the subject and object SRPs for the various performances of inflected verb signs, for example, a white color target on the left could be the subject, and an orange color target on the right could be the object. We found this use of color targets in the room to be less error-prone than other approaches for collecting many samples of ASL inflecting verbs.

We have made use of these recordings in our prior research (Lu & Huenerfauth, 2011b) to produce models of the motion-path of ASL verbs, and we decided to also release this data to the research community – to facilitate the work of ASL linguistics or animation researchers studying ASL verbs. This "verb" corpus contains a high-resolution video recording of the signer during the collection and the plaintext data files from the IS-900, which consists of a tab-delimited file with columns for: the time code (milliseconds) and for each sensor: the location coordinates (x, y, z) and orientation (yaw, pitch, roll). Sensors were placed on both of the signer's hands, the signer's torso, and the signer's head. We recorded this "verb" corpus from three signers (different people than those recorded in the unscripted multi-sentential discourse discussed in section 5). This small corpus contains several hundred instances of eight ASL inflected verbs: ASK (one-handed version), GIVE, MEET, SCOLD, TELL, EMAIL, COPY, and SEND. The fact that we were able to make use of these recordings to produce models of ASL verb movement, which native ASL signers judged to be of good quality in an experimental study we conducted (Lu & Huenerfauth, 2011b), is good evidence that the quality of the motion-capture data is sufficient for supporting computational linguistic research on these verbs.

Figure 10: Close-up views of the hand-mounted sensor used in the motion capture sub-corpus data collection.

## 8. Conclusion

To address the lack of linguistically annotated ASL corpora with sufficient 3D movement detail for animation research, we began a multi-year project to collect and annotate a motion-capture corpus of ASL. In this paper, we are releasing the first portion of the "CUNY ASL Motion-Capture Corpus," which has been collected and annotated at our laboratory at Queens College of the City University of New York (CUNY). Our goal is for the digital 3D body movement and handshape data we collect from native signers to become a permanent research resource for NLP researchers, ASL linguists, and sign language animation researchers. This corpus will allow researchers to create new ASL generation technologies in a data-driven manner by analyzing the subtleties in the motion data and its relationship to the linguistic structure.

Our initial research focus is to model where signers tend to place spatial reference points in the signing space. Another early goal of our research is to discover patterns in the motion paths of inflecting verbs and model how they relate to layout of SRPs. These models we develop could be used in ASL generation software or could be used to partially automate the work of humans using ASL-scripting systems.

Because we are still collecting, post-processing and annotating this corpus, we plan to provide additional releases of this corpus in future years. This paper has suggested various additional forms of annotation and motion-capture data that we intend to release in the future, and we welcome feedback from the research community about how this resource can be made more useful and accessible.

## 9. Acknowledgements

This material is based upon work supported in part by the US. National Science Foundation under award number 0746556 and award number 1065009, by The City University of New York PSC-CUNY Research Award Program, by Siemens A&D UGS PLM Software through a Go PLM Academic Grant, and by Visage Technologies AB through a free academic license for character animation software. Jonathan Lamberton assisted with the recruitment of participants and the conduct of data-collection sessions. Kenya Bryant, Wesley Clarke, Kelsey Gallagher, Amanda Krieger, Giovanni Moriarty, Aaron Pagan, Jaime Penzellna, Raymond Ramirez, and Meredith Turtletaub have also assisted with data-collection and have contributed their ASL expertise to the project.

## 10. References

Cormier, K. (2002). *Grammaticalization of Indexic Signs: How American Sign Language Expresses Numerosity.* Ph.D. Dissertation, University of Texas at Austin.

Huenerfauth, M., Lu, P. (2010a). Eliciting Spatial Reference for a Motion-Capture Corpus of American Sign Language Discourse. *Proceedings of the Fourth Workshop on the Representation and Processing of Signed Languages: Corpora and Sign Language Technologies, The 7th International Conference on Language Resources and Evaluation (LREC 2010)*, Valetta, Malta.

Huenerfauth, M., Lu, P. (2010b). Modeling and Synthesizing Spatially Inflected Verbs for American Sign Language Animations. In *Proceedings of The 12th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2010),* Orlando, Florida, USA. New York: ACM Press.

Huenerfauth, M., Lu, P. (2012). Effect of Spatial Reference and Verb Inflection on the Usability of American Sign Language Animations. *Universal Access in the Information Society.* Berlin/Heidelberg: Springer.

Johnston, T. (2009). Creating a Corpus of Auslan within an Australian National Corpus. In *Michael Haugh et al. (eds.), Selected Proceedings of the 2008 HCSNet Workshop on Designing the Australian National Corpus,* Somerville, MA: Cascadilla Proceedings Project, pp. 87--95.

Lu, P., Huenerfauth, M. (2009). Accessible motion-capture glove calibration protocol for recording sign language data from deaf subjects. In *Proceedings of The 11th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2009),* Pittsburgh, Pennsylvania, USA. New York: ACM Press, pp. 83--90.

Lu, P., Huenerfauth, M. (2010). Collecting a Motion-Capture Corpus of American Sign Language for Data-Driven Generation Research. *Proceedings of the First Workshop on Speech and Language Processing for Assistive Technologies (SLPAT), Human Language Technologies: The 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics,* Los Angeles, CA.

Lu, P., Huenerfauth, M. (2011a). Collecting an American Sign Language Corpus through the Participation of Native Signers. In *Proceedings of the International Conference on Universal Access in Human-Computer Interaction (UAHCI).* Orlando, Florida, USA.

Lu, P., Huenerfauth, M. (2011b). Synthesizing American Sign Language Spatially Inflected Verbs from Motion-Capture Data. *The Second International Workshop on Sign Language Translation and Avatar Technology (SLTAT), The 13th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2011),* Dundee, Scotland, United Kingdom.

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

Meier, R. (1990). Person deixis in American Sign Language. In *S. Fischer, P. Siple (Eds.) Theoretical issues in sign language research, Vol 1, Linguistics.* Chicago: University of Chicago Press, pp. 175--190.

Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., Lee, R. (2000). *The Syntax of American Sign Language: functional categories and hierarchical structure.* Cambridge, MA: MIT Press.

Padden, C. (1988). *Interaction of morphology and syntax in American Sign Language.* New York: Garland.

Traxler, C. (2000). The Stanford achievement test, ninth edition: national norming and performance standards for deaf and hard-of-hearing students. *Journal of Deaf Studies & Deaf Education* 5(4), pp. 337--348.

# Dicta-Sign – Building a Multilingual Sign Language Corpus

**Silke Matthes[1], Thomas Hanke[1], Anja Regen[1], Jakob Storz[1], Satu Worseck[1], Eleni Efthimiou[2], Athanasia-Lida Dimou[2], Annelies Braffort[3], John Glauert[4], Eva Safar[4]**

[1]Institute of German Sign Language and Communication of the Deaf, University of Hamburg, [2]Institute for Language and Speech Processing / "Athena" R.C., [3]LIMSI/CNRS, [4]University of East Anglia
{silke.matthes,thomas.hanke,anja.regen,jakob.storz,satu.worseck}@sign-lang.uni-hamburg.de,
{eleni_e,ndimou}@ilsp.gr, annelies.braffort@limsi.fr, {J.Glauert,E.Safar}@uea.ac.uk

**Abstract**

This paper presents the multilingual corpus of four European sign languages compiled in the framework of the Dicta-Sign project. Dicta-Sign researched ways to enable communication between Deaf individuals through the development of human-computer interfaces (HCI) for Deaf users, by means of sign language. Sign language resources were compiled to inform progress in the other research areas within the project, especially video recognition of signs, sign-to-sign translation, linguistic modelling, and sign generation. The aim for the corpus data collection was to achieve as high a level of naturalness as possible with semi-spontaneous utterances under lab conditions. At the same time the elicited data were supposed to be semantically close enough to be comparable both across individual informants and for all four sign languages. The sign language data were annotated using iLex and are now made available via a web portal that allows for different access options to the data.

**Keywords:** Sign language technologies, multilingual sign language resources, annotation

## 1. Introduction

Within the framework of the Dicta-Sign project (2009-2012) sign language resources were compiled for four European languages: British, German, Greek, and French Sign Language (BSL, DGS, GSL, and LSF). These resources were used to inform progress in other research areas within the project, especially sign recognition, sign-to-sign translation, linguistic modelling, and sign generation, which then in turn was used to improve sign language technology. At the same time the data are to serve as a self-contained resource for future research.

In a first step, a multilingual lexical database providing a core lexicon of approximately 1000 entries in the four project sign languages was built. The shared list of concepts chosen for the lexicon are of everyday use or specifically related to the field of Dicta-Sign's main topic, European travel. Signs were recorded for each language and annotated assigning gloss labels, form description (HamNoSys) and a rough meaning.

In a second step, a new corpus on the domain "Travel across Europe" was produced by using the same elicitation materials for all four sign languages. Prior to the project, parallel corpus collection for sign languages had only been undertaken in minimal sizes or for spoken language simultaneously interpreted into several sign languages, but not for semi-spontaneous signing by native signers. Because of the "oral" nature of sign language and the risk of influences from written majority languages the collection of parallel sign language data is a difficult task. Corpus planning therefore needs to balance between naturalness of the data to be collected on the one side and the degree of parallelisability of the data across languages on the other side. Within Dicta-Sign, the aim for the data collection was to elicit sign language data as natural as possible with semi-spontaneous utterances under lab conditions. With respect to parallelisability of the sign language data, elicitation tasks had to be designed that result in semantically close answers without predetermining the choice of vocabulary and grammar (Matthes et al. 2010).

Corpus data collection took place in each of the four countries involved in the project and the sign language data were annotated using iLex. A web portal was developed to allow access to the corpus data for research purposes.

## 2. Compilation of the Multilingual Corpus

A multilingual corpus on the domain "Travel across Europe" was compiled for the four sign languages involved in the project (BSL, DGS, GSL and LSF). Elicitation tasks were developed specifically for the project's purposes. After recording had taken place in all four countries, the sign language data were annotated on different levels.

### 2.1 Corpus Data Elicitation

With the objective of gaining sign language data as natural as possible on the one hand and comparable across languages as well as individual informants on the other hand elicitation tasks and materials were designed specifically for the Dicta-Sign corpus collection. One key point in the planning was to film Deaf informants in pairs, interacting with each other. The tasks therefore mostly required the active involvement

of both conversational partners, asking them to discuss and negotiate on certain topics or to describe and explain things to the partner. The elicitation material consists of 10 different tasks, aiming at a session length of approximately two hours, and covers different interaction formats ranging from monologues to sequences of very short turns, also with different levels of predictability. It includes communication for transport by different means and contexts as well as related personal experiences (Matthes et al. 2010).
The complex studio setup that was decided to be used for Dicta-Sign's data collection consisted of seven cameras, two of them stereo cameras (Hanke et al. 2010a). The different camera perspectives (front, side and bird's eye view) were to help annotators interpret the signing. The additional stereo cameras provide footage that allows image analysis to reconstruct 3D information and help automatic processing. In each country, 16 to 18 informants were filmed in sessions lasting about two hours each. Not counting task explanations or material that needed to be excluded for certain reasons, the corpus now consists of 8 to 10 hours of signed data from 14 to 16 different signers per language.
A variety of post-processing steps were needed before annotation work could start, most importantly providing backup data, compression of the video files as well as precise frame-by-frame synchronisation.

## 2.2 Annotation

Corpus annotation work for all four sign languages within the Dicta-Sign project was carried out using iLex, an annotation environment that is linked to a lexical database (Hanke/Storz 2008). The video data were integrated into iLex and transcripts were produced for all tasks. At UHH, where Session Director had been used to run the elicitation sessions (Hanke et al. 2010a), it was possible to provide automatic tagging specifying start and end of the individual tasks and subtasks using time information from the Session Director log files. The annotation consists of a basic annotation on sign level for subsets of the sign language data as well as content tags that allow detecting comparable content across different signers and languages.

### 2.2.1 Sign Level Annotation

Sign level annotation of the corpus data is now available for about 40 minutes up to 5.5 hours per language, including segmentation of signs, lemmatisation, form description, as well as further details depending on the individual language. With regard to segmentation of the continuous signing it was decided to treat transitional movements between individual signs not as part of either sign (i.e. there are gaps between two signs during which the articulators move from the end of one sign to the beginning of the next). This approach, though more time consuming than segmenting once in the middle of a transition, offers advantages for subsequent processing: Firstly it results in the fact that a token tag only represents that part of the signal that is described by HamNoSys. Secondly, variation between tokens is much lower than if the transition would be part of the sign.
After segmentation the individual signs were lemmatised, i.e. unique glosses were assigned by means of type-token matching. In iLex this is done by linking tags to type entries in the database, which results in filling the transcript and a growth of the sign language database at the same time. A form description of the sign types was added using HamNoSys (Hanke 2004).
As a further step to enrich the corpus data, individual project partners conducted extra annotation work on data from their respective sign language: For the DGS data mouth patterns were annotated by assigning either written German words to represent mouthings or "MG" as a preliminary tag for mouth gestures. Furthermore, in addition to the annotation of HamNoSys on type level form deviations between types and the respective tokens were tagged as such in order to provide more reliable training data for image processing. The GSL data include a tagging of clause boundaries, and for LSF pointing, buoys and depicting signs were annotated using categories close to those proposed in the Auslan annotation guideline provided by Trevor Johnston (2011), and for pointing and depicting signs some indication on the use of the signing space were added. Furthermore, English translation is provided for parts of the German, Greek, and French subcorpora as well as a French translation for the majority of the LSF data.

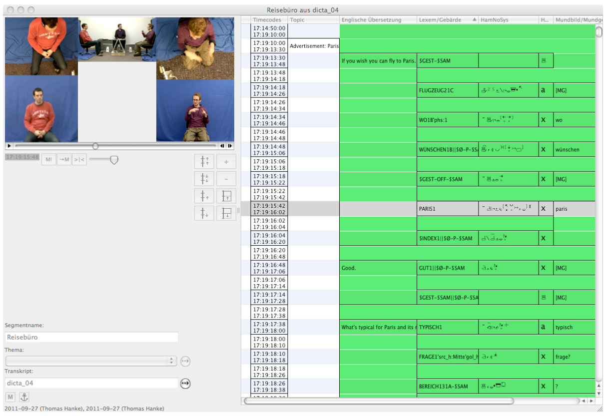


Figure 1: Transcript of Task "Travel Agency" (DGS informant)

### 2.2.2 Content Tagging

The Dicta-Sign multilingual corpus is not a parallel corpus in the classical sense as the "oral" nature of sign language as well as the risk of influences from written majority languages do not allow for such an approach. Instead, the aim within Dicta-Sign was to elicit semantically close answers without predetermining the choice of vocabulary and grammar. In order to allow identifying video sections in the corpus with comparable content across individual informants and languages, especially for those parts without sign level annotation, content tags were assigned that reflect the topics the informants signed about. The detailedness of the content tags varies across the different tasks, ranging from very broad content descriptions that mainly reflect the given structure of the subtasks to a more detailed specification of the topics covered (see examples below).

*Example 1:*
For the task "Public transportation" (task 1, category "Route descriptions") the informants are asked to explain how to get from a certain place to another using public transportation. A map is provided to both of them displaying different means of public transport and stations. In five subtasks different stations are given as departure and destination points and each informant is asked to suggest one possible route per subtask.
For each subtask between nine and 12 different routes were described. While many of the 60 informants described similar routes, several routes occurred only once or twice. Mapping information was needed to compare information from the different sign languages: Route codes were agreed on and pictures were produced for each of the routes in order to ease the mapping (see pictures below). Discussion about the chosen routes was included in the route tags, but in cases were further discussion evolved (e.g. advantages of taking the bus) this was tagged separately.

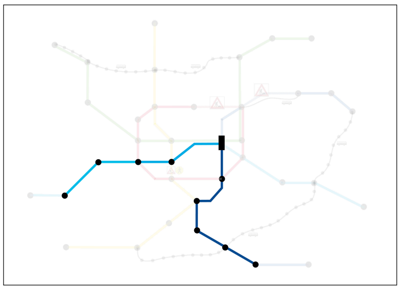


Figure 2: Route R2.2 (by 23 signers)

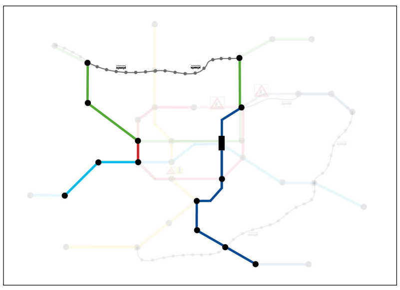


Figure 3: Route R2.6 (by 1 signer)

*Example 2:*
For the task "At the airport" (task 4, category "Description of Places and Activities") one informant is asked to explain the procedures taking place at the airport as if the other has never travelled by plane before. Pictures displaying different aspects as checking in, boarding, and baggage claim are shown in chronological order.

Figure 4: "At the airport" (German version)

The content tags distinguish the different steps that were described by the individual informants. Most of the topics are directly related to the elicitation material – as e.g. *check-in*, *information board*, *security check*, *food and drinks on board*, and *baggage claim* – and were covered by almost all 23 informants across three sign languages.[1] However, additional topics occurred: e.g. *Preparation of the trip* was mentioned by four informants (DGS, GSL, and LSF), *airplane fuelling* by three informants (GSL and LSF) and *Amusement activities on board* by 10 informants altogether (DGS, GSL, and LSF).

*Example 3:*
For the task "Expectation & Reality" (task 6, category "Narration") the informants were asked to tell short stories based on picture cards showing somebody's expectations of a certain situation and the actual situation. Topics of the stories were: small hotel room, cancelled flight, crowded museum, posh restaurant, rained off BBQ, and missed sunset.

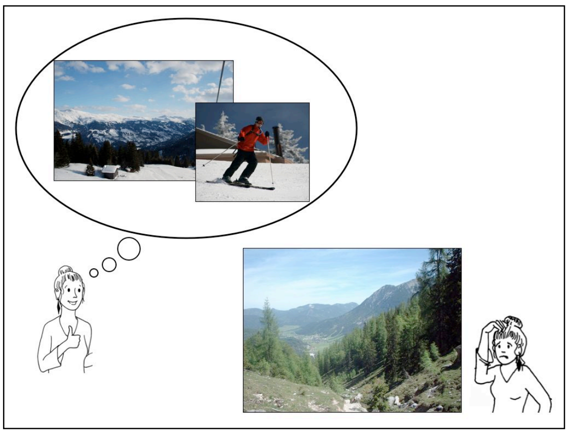


Figure 5: Example from the task explanation

[1] Only one informant per pair performed this task. For BSL tagging of this task is not available.

With respect to content tagging, the signing was first segmented into individual stories (i.e. subtasks), using Session Director log file information where available, and further divided into the 'expectation' and the 'reality' part of the story. For this task, applying only two content tags per subtask seems appropriate to mark the content, as the individual stories told by the informants are comparably short. For example, in the DGS data the content tags are – depending on the subtask – of a length ranging from 15sec up to 1min:19sec, with the 'expectation' part always being slightly longer than the 'reality part'. This results in stories with an average length of about 1min:20sec. Both parts of the six stories could be detected in the data of almost all informants of the four sign languages.

### 2.3 Metadata

Personal metadata was collected from the informants by means of questionnaires based on the IMDI standard with sign language-specific extensions as defined in Crasborn/Hanke 2003. As that set covers a variety of purposes for metadata (e.g. to support language acquisition studies), but does not explicitly define subparts, Dicta-Sign defined a subset that seemed suitable for the kind of study conducted here and also minimised the questionnaire filling effort for the informants.

Metadata was collected in a finer granularity than appropriate for publication, however standards are not yet available that specify suitable coarsenings for such data. Therefore, two levels of coarsening were defined within Dicta-Sign for different publicity levels of informant data (see below on portal structure). For example, the informant's date of birth is converted to the age in years for restricted access and age range (e.g. 41-50) for public access.

For the time being, data are made available in IMDI session file format. We plan, however, to convert these data into the CMDI component structure.

## 3. Exploitation

### 3.1 The Dicta-Sign Web Portal

A web portal was developed to allow access to the Dicta-Sign language resources for public use as well as research purposes. It can be accessed from the Dicta-Sign website: http://www.dictasign.eu/Main/Portal. Besides Dicta-Sign's basic lexicon and further training data for sign recognition the portal presents the multilingual corpus, allowing for different access options to the data.

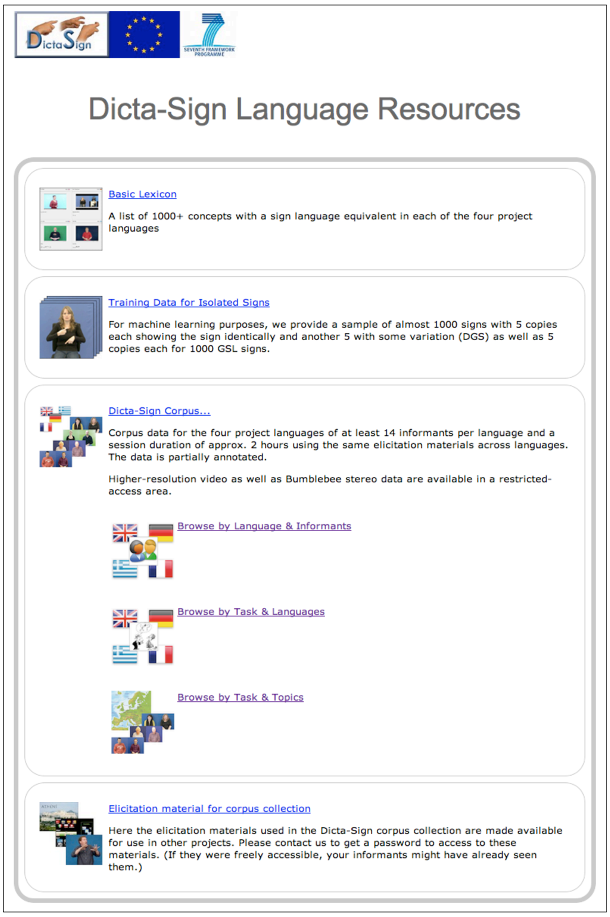



Figure 6: The Dicta-Sign web portal

#### 3.1.1 Layout

The Dicta-Sign web portal offers different approaches to access the corpus data:

- *By Language & Informants:* For every sign language the available recording sessions (i.e. pairs of informants) are listed. Via the sessions all tasks performed by the respective informants as well as informant metadata information can be accessed.
- *By Task & Languages:* For each task that the informants were asked to perform a short description as well as the elicitation material is provided. Grouped by languages, all data-by-task items for a certain task are listed and can be accessed. In addition, the content tags defined for each task are presented.
- *By Task & Topics:* This approach makes use of the topics identified as part of the annotation process. By listing all content tags of the individual tasks it allows access to comparable data across individual informants and languages.

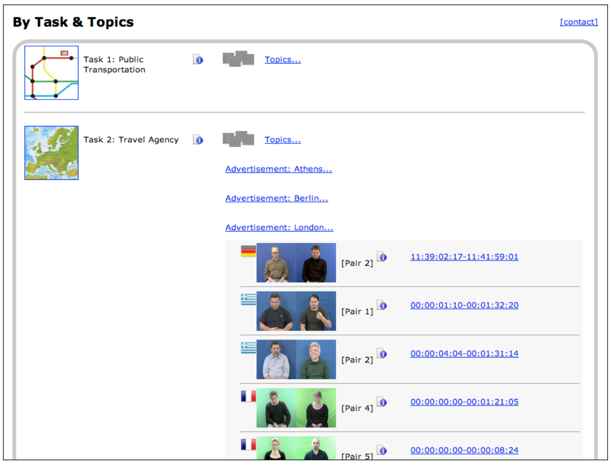



Figure 7: Content tags in the web portal

### 3.1.2 Access Levels

There are three accessibility levels to the corpus data:

- *Publicly available on the project site:* Metadata and "appetisers" to the video and transcript data. Information on elicitation materials is at the same level of detail as previously published (Matthes et al. 2010).
- *Restricted access for researchers:* Elicitation materials available for researchers' own purposes as well as to fully understand the data collection. Video and transcript data as available in a standard format (H.264 for video, ELAN and iLex export format for transcripts), more detailed metadata.
- The third level is not available online, but requires arrangements between the researcher interested and the individual partner owning the data. This includes higher-resolution less compressed video and stereo data and even more refined metadata.

The corpus video and annotation data linked to the portal via the different access options are made available as videos with and without subtitles, in iLex export format as well as in ELAN format. The elicitation material including task explanations in the respective sign language is provided as Keynote or PowerPoint documents.

Additionally the web portal includes contact forms for researchers who request higher-resolution and stereo data from individual partners or ask to contact informants to be given access to more detailed metadata or to suggest additional data collection.

### 3.2 Finding the most Parallel Content Tags

The content tagging, as described in chapter 2.2.2, facilitates a rough comparison of the corpus data on the semantic level. Via the "Task & Topics" approach of the portal access to individual topics is provided and allows for direct comparison across languages and individual informants. The problem remains how, for a given topic tag, to find the closest match in another language, from the set of identically tagged stretches of signing offered by the portal.

Here we report on the experiments undertaken to gain a better understanding of what can be done for sign language corpora. For written language texts, a variety of similarity measures have been suggested in the literature, often relying on probabilistic models. As the needed statistical data are not yet available for sign language lexicons, we started with a very simple measure, namely lexical overlap count relative to sample size within one language (DGS) in the "At the airport" task. Not surprisingly, this measure highly depends on lexical variation. In fact, it becomes useless if signers with different sign dialects are involved. However, computing overlap in the semantic domain (concept entries assigned as meanings to the types) and thereby eliminating the influence of lexical variation provided results coming close to the annotators' intuition.

In order to apply this approach to content tags from different languages, a common semantic basis such as compatible SignNets in the sense of WordNets would be needed. Dicta-Sign has provided a list of 1000 concepts and signs in each of the four project languages for each of these. In many cases, WordNet sense keys could be assigned to the concepts whereas in other cases the sign languages require a granularity not provided by a WordNet for English.

Now we used the same measure as before, but only the instances of types with meanings in the 1000 concepts list could be taken into account. In our test case – DGS-BSL – this meant a reduction of the counts to one third, to sample sizes of 10-80 concepts. Overlap measures were no longer comparable, but seemed to provide tendencies nevertheless.

In order to provide more reliable measures, larger cross-language resources would be needed, ideally a "EuroSignNet".

## 4. Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 231135.

## 5. References

Crasborn, O., Hanke, T. (2003). Metadata for sign language corpora. Online available at http://www.let.ru.nl/sign-lang/echo/docs/ECHO_Metadata_SL.pdf (last seen 30/03/2012).

Hanke, T. (2004). HamNoSys – Representing Sign Language Data in Language Resources and Language Processing Contexts. In: Streiter, O., Vettori, C. (Eds). LREC 2004, Workshop proceedings: Representation and processing of sign languages. Paris: ELRA, pp. 1-6.

Hanke, T., Storz, J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. In: Crasborn, O., Hanke, T., Efthimiou, E., Zwitserlood, I., Thoutenhoofd, E. (Eds.): Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages. Paris: ELRA, pp. 64-67.

Hanke, T., König, L., Wagner, S., Matthes, S. (2010a). DGS Corpus & Dicta-Sign: The Hamburg Studio Setup. In: Dreuw, P. et al. (Eds.): LREC 2010. 7th International Conference on Language Resources and Evaluation. Workshop Proceedings. W13. 4th Workshop on Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. May 22/23 2010. Valetta – Malta. Paris: ELRA, pp. 106-109.

Hanke, T., Storz, J., Wagner, S. (2010b). iLex: Handling Multi-Camera Recordings. In: Dreuw, P. et al. (Eds.): LREC 2010. 7th International Conference on Language Resources and Evaluation. Workshop Proceedings. W13. 4th Workshop on Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. May 22/23 2010. Valetta – Malta. Paris: ELRA, pp. 110-111.

Johnston, T. (2011). Auslan Corpus Annotation Guidelines. Online available at http://www.auslan.org.au/video/upload/attachments/AuslanCorpusAnnotationGuidelines30November2011.pdf (last seen 30/03/2012).

Matthes, S., Hanke, T., Storz, J., Efthimiou, E., Dimou, N., Karioris, P., Braffort, A., Choisier, A., Pelhate, J., Safar, E. (2010). Elicitation Tasks and Materials designed for Dicta-Sign's Multi-lingual Corpus. In: Dreuw, P. et al. (Eds.): LREC 2010. 7th International Conference on Language Resources and Evaluation. Workshop Proceedings. W13. 4th Workshop on Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. May 22/23 2010. Valetta – Malta. Paris: ELRA, pp. 158-163.

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

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

**Abstract**

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

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

## 1. Introduction

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

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

## 2. Swedish Sign Language Corpus

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

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

## 3. Lexicography

A lexicographic work of Swedish Sign Language was initiated in 1988 in the Sign Language section of the Department of Linguistics at Stockholm University. In 2001, the first dictionary resulting from the project went online, and it was entitled the *Digital version of Swedish Sign Language Dictionary*. A follow-up dictionary called the *Swedish Sign Language Dictionary* was created in 2008 and has been in development since (see Mesch, Wallin & Björkstrand, in this volume). Today, it has approximately 8,000 sign entries. There are also nine lexical databases with specialized lexicons in Swedish Sign Language, totaling 4,300 sign entries, of which some are also found in the Swedish Sign Language Dictionary, henceforth the SSLD. In the SSLD, a sign is described in terms of its handshape, orientation, location, movement, Swedish translations, and sign transcription. The lexicographic work group associated with this project documents the vocabulary of Swedish Sign Language and continuously updates the web-based dictionaries.

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

## 4. Annotation work with glosses

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

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

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

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

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

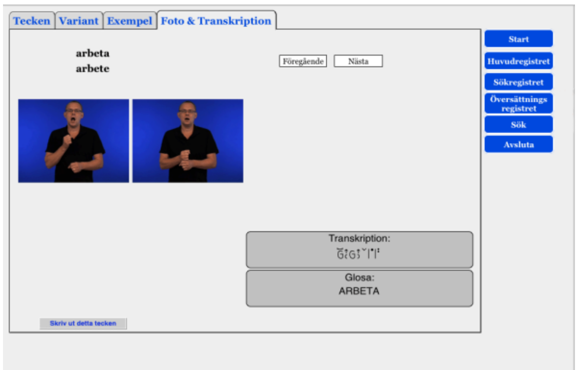



Figure 1: Reference dictionary with suitable gloss for annotators

## 5. Different point of views for glossing

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

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

As shown, there are some challenges to finding a suitable gloss in the SSLC annotation. For example, for the sign EREKTION *'erection'* in figure 3, the corpus annotator would write a more detailed gloss with the purpose of indicating a correct meaning, whereas in the SSLD a more general meaning would be chosen. That is, it does not pertain only to the stiffening of a man's penis, but also to a woman's clitoris. In the SSLC, this would be glossed as PENIS-STYV (*'penis-stiff'*). This means that this particular sign entry in the SSLD needs to be changed.

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

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

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

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

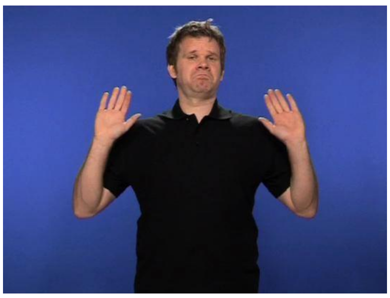


Figure 4: Dictionary entry: inte mitt problem (*not my problem*).

GET-1 TWO HUNDRED

ONE-MORE MORE MEMBER

(g-)PF REALLY MEMBER

CLEAR

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

## 6. Back to the dictionary

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

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

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

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

## 7. Conclusion

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

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

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

## 9. Acknowledgements

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

## 8. References

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

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

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

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

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

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

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

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

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

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

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

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

# Sign Language Resources in Sweden: Dictionary and Corpus

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

**Abstract**

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

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

## 1. Introduction

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

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

## 2. Online Swedish Sign Language dictionaries

The first version of the online dictionary was created in 2001; it was called the *Digital version of the Swedish Sign Language Dictionary* and included 3,132 sign entries. This was the result of lexicology work initiated in 1988 by the Sign Language section of the Department of Linguistics at Stockholm University. This dictionary includes video files for signs taken from a printed dictionary of the Swedish National Association of the Deaf (1997). Then, nine databases with specialized lexicons in Swedish Sign Language, with a total of 4,300 sign entries, were created during the years 2004–2011. The *Swedish Sign Language Dictionary*, the SSLD, was created in 2008 and has been in development since. Today, it has approximately 8,000 dictionary sign entries, including parts of the specialized databases mentioned above (see figure 1). The lexicographic work group documents Swedish Sign Language vocabulary and continuously updates the web-based dictionaries. All these databases are created as FileMaker Pro database files and are available online at http://www.ling.su.se/teckensprak

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

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

Figure 1. Lexicographic SSL databases

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

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

Figure 2. Sign entry in the SSLD

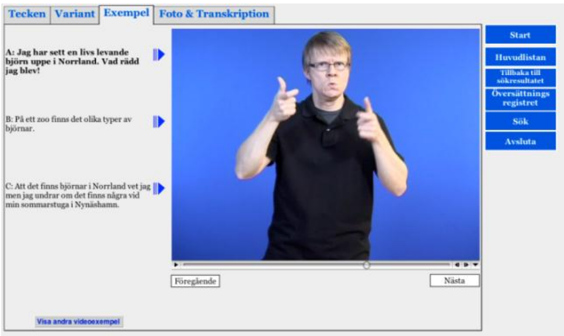


Figure 3. Example usage in sentences in the SSLD

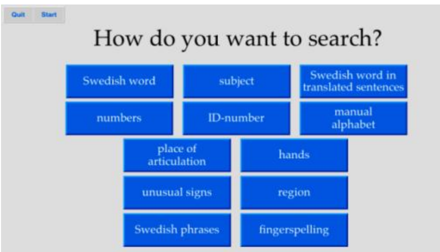


Figure 4. Searching fields in the SSLD

As shown in Figure 4, there is a variety of search paths one can take. These paths are a) Swedish word; b) subject, which is a list of signs associated with a specific domain, such as technology, family and clothing; c) Swedish words in translated sentences, which can be selected to show the dictionary user how specific signs are used in sentences; d) numeral, which consists of a list of numeral sign or sign with incorporated number; e) ID-number, for those cases in which the user knows the number of the sign entry; f) manual alphabet, which displays the Swedish manual alphabet as a list; g) place of articulation; i) handshape for one or two hands; j) unusual/old signs that exist but are not in common use today (often signs used by older generations, but not younger ones); k) regions, in which signs are divided by region of location and former location of deaf schools; l) sign language phrases such as 'good morning', 'how are you?' or 'cannot afford'; and m) fingerspelling, which consists of fingerspelling signs, fingerspelling affixes and fingerspelling parts in sign compounds.

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

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

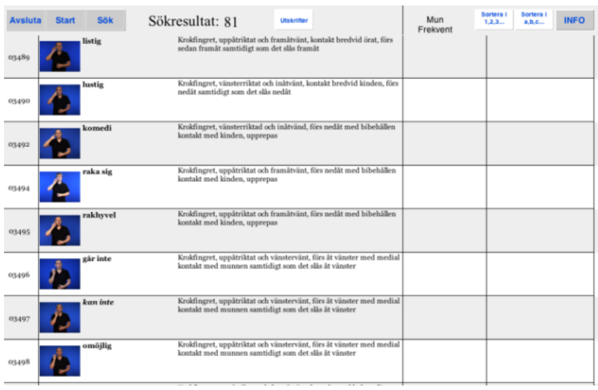


Figure 5. Sign list in the SSLD

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

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

Table 1. The most common search paths for SSLD users

## 3. Upcoming: Swedish Sign Language Corpus

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

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

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

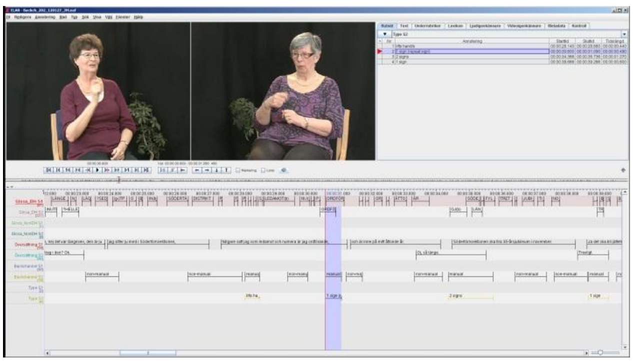


Figure 6. An example illustrating conversation materials in the SSLC with ELAN

## 4. Two resources for different uses

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

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

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

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


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

## 5. Combined resources

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

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

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

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

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

The next task for the project will be to create study material of sign language structure, using both of these resources. The material should include in sign description of the sign structure (how one or two hands perform a movement in a particular position), sign formation processes, parts of speech, sentence formation, spatial organization and variation in language use. The material is supposed to consist of short texts with hyperlinks to the video examples. It will be available on the web portal of the Department of Linguistics probably in the beginning of 2013, with links to the SSLD and the discourses of SSLC.

## 6. Conclusion

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

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

## 7. Acknowledgements

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

## 8. References

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

Mesch, J., Wallin, L., Nilsson, A-L. & Björkstrand, T. (2010). Corpus and Dictionary – separate or together? Poster presented at Sign Linguistics Corpora Network Workshop 4: Exploitation, December 3, 2010, Berlin.

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

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

Skollink. http://www.skollink.se

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

Svenskt teckenspråkslexikon. (1997). Sveriges dövas riksförbund.

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

# A conceptual approach in sign language classification for concepts network

**Cédric Moreau**
INS HEA
Institut National Supérieur de Formation et de Recherche pour l'Education des Jeunes Handicapés et les Enseignements Adaptés.
58, 60 av des Landes 92150 Suresnes France
UMR 7023, Université Paris 8
2 rue de la Liberté, 93200 Saint-Denis France
cedric.moreau@inshea.fr

**Abstract**

Most websites presupposes a conceptual equivalence between a written word and a sign. In such tools, signs, which don't have strict written equivalent lexicons, can't be found. The collaborative website OCELLES project LSF/French tries to give the opportunity to obtain several signs for a unique concept, with the possibility of uploading a sign without being constrained by written language. Although word checking in a written text is quite easy, it is not the case for sign checking in a video.
Today studies are carried out in the field of gesture recognition, but all the sign language linguistic parameters cannot be considered as such. Indeed, they have to be used simultaneously during communication interactions.
Our approach Based upon the semiological Cuxac model (Cuxac, 2000) and Thom morphogenesis theory (Thom, 1973), could help to find a sign in a sign dictionary without using written language.

**Keywords:** sign language, LSF, morphogenesis, catastrophe theory, OCELLES

## 1. Representation of lexicalized and iconic sign on internet websites.

According to Cuxac (2000), two discursive enunciation strategies co-exist in sign language, using visual-gestual channel, you can choose to communicate either by showing or not. It means you can "let people see" your experience with a visually sequence of signs, or you can use lexicalized signs which don't bear any resemblance with the experience you describe.
Today, most websites propose only lexicalized signs and overlook all iconic signs which are the most used ones, depending on speech type studies (Sallandre, 2001). This approach comes from methodology choices. The conception of these websites presupposes a conceptual equivalence between a written word and a sign. Users are invited to upload a lexicalized sign from a chosen written word. In such similarity based tools, signs which don't have strict written equivalent lexicons or iconic signs can't be found.
The collaborative website OCELLES project LSF/French tries to give the opportunity for sign users to obtain several signs for a unique concept, so that they can use them as signifiers, without being limited in their choice (they could chose lexicalized or iconic signs). (Moreau & Mascret, 2010).

## 2. Signs access on website

It seems that, with the possibility of uploading a sign without being constrained by written language, problems could be solved. But which sign access are deaf people provided with when looking for a sign, in a conceptual network like OCELLES, when they have no idea what the equivalent written word is? Although word checking in a written text is quite easy, it is not the case for sign checking in a video.
How can a specific sign in a video be found? Today studies are carried out in the field of gesture recognition (Dreuw & Ney, 2008; Lefebvre-Albaret & Dalle, 2010) but all the linguistic parameters cannot be taken into consideration, for instance specific parameters of sign languages (handshape, movement, place (Stokoe, Casterline, & C –Cronenberg, 1965), orientation (Friedman, 1977; Liddell, 1980; Moody, 1980; Yau, 1992), but also symmetry (Filhol, Braffort, & Bolot, 2007), …). These linguistic parameters cannot be considered as such. Indeed, even if a human mind can discern one from the other, as isolated significant elements, they have to be used simultaneously during communication interactions. Contrary to vocal languages, realizing a signifying form in a sign language cannot be made through a succession of distinct realizations of isolated and non-signifying elements. Minimal realization structures in sign language may be ranged on a growing complexity scale, starting from the formal transfer (infra-conceptual level) and going up to the double transfer (level where several actors, location parameters and utterances can be combined) (Cuxac, 2000). These various structures use the same linguistic parameters during the same realization laps of time. (Moreau & Mascret, 2010)
If we consider these elements, we can observe that few websites propose thematic approaches making it possible to find a sign through labels including animated signs. In most of existing tools, deaf users have to master written language which often isn't their natural language. Deaf people can't find a sign directly in a document the same way vocal speakers can find a word in a text or in a dictionary.

## 3. Theoretical and conceptual framework

### 3.1 Hypothesis

The approach which has currently been chosen, is both theoretical and general, it could help find a sign in a sign dictionary or sign ontology without using written language. Based upon the semiological Cuxac model (Cuxac, 2000) and Thom morphogenesis theory (Thom, 1973), we consider a sign as a constellation of pregnant (stable and perpetual) parameters. A sign is a dynamic form i.e. a set of space discontinuities which changes in time.

### 3.2 Space and internal dynamics

According to Petitot[1], in the "catastrophe theory", a substratum has a spatial extension, in which each point has a local physic. This local process was called internal dynamic by Thom (1973). Therefore, each point has an internal dynamic. Spatial extension of substratum works as a coupling mode between internal dynamics, what Thom calls space control. Position in control space creates interactions between local dynamics and others which are nearby. These interactions propagate spatially and the coupling exists thanks to characteristic substratum mechanisms. Space becomes mainly a coupling factor, which connects internal dynamics. Space isn't a container, but an interaction principle between internal dynamics.

When we move spatially, internal dynamics, which result from couplings, are transformed and deformed. But, for one point, these internal dynamics define the local state of substratum. So, when some critical values are crossed while moving, the internal dynamic modifies internal states of the system.

Some domains are logically found within some internal states which predominate each time. Each domain is delimited by boundaries. So, domains with boundaries define the concept of form. Each form means that substratum space is broken.

Dynamic can be defined as a process which minimizes energy level. At a given point, internal dynamic is described by a function of potential. The internal states are the minima of this function. This principle is an optimization principle. Thom calls this first category: elementary catastrophe (Thom, 1973).

The second category is called generalized catastrophe. This approach corresponds to complex situations including many sorts of dynamics. The theorem shows that in each dynamic, there has to be some dissipation or gradient decrease, in the shape of a depression, the minima of which is called system attractor.

### 3.3 Isomorphism

According to the Gestalt theory, we postulate an isomorphism between the world and the way the person perceives it.

This dynamic and topological representation must obviously be understood in a broad sense: abstract and complex. If we perceive a handshape it doesn't mean that this handshape will physically take shape in our brain. It is not a strict coding of our sensations, particularly concerning our perception of space and time.

"Thom claims that the principle organizing the combination of meaning-carrying units in language corresponds to the principle underpinning the configuration of phenomenal parts into intelligible wholes in perception. The rationale of this claim is biological: it seems sensible to suggest, as Thom says (Thom, 1980 b p. 180), that language has evolved from the necessity of (or the advantage inherent in) conveying to others the significant changes (i.e. the catastrophes) in the environment. This entails—as Thom with no further argument asserts—that the syntactic structure "naturally" reflects the dynamic structure of the external catastrophe." (Bundgaard & Stjernfelt, 2010)

## 4. Application in the sign language

### 4.1 Process of a sign formulation

The sign achievement process is considered as an optimization process. A sign looks acceptable to a sign language speaker, if it complies with signing constraints. A sign is considered as acceptable when meeting with meaningful linguistic units interactions.

### 4.2 Space of the sign and conceptual space

We consider two isomorphic spaces: a sign space and a conceptual space.

During sign procedure, its form changes into a potential gradient, under the influence of internal variables, resulting both from internal constraints and the period of achievement. Every minima of this space corresponds to a system attractor.

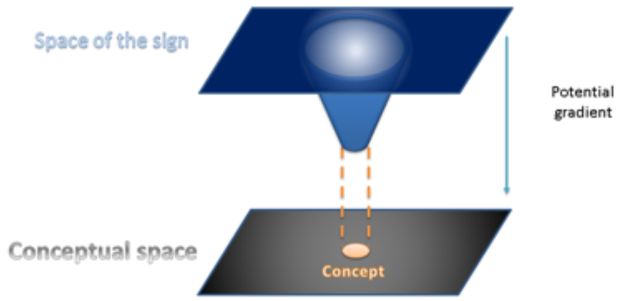



Figure 1: Space of the sign and conceptual space

The sign space is the sum of morphemic subspaces. Each subspace is evidence of a morpheme and characteristic of its internal states.

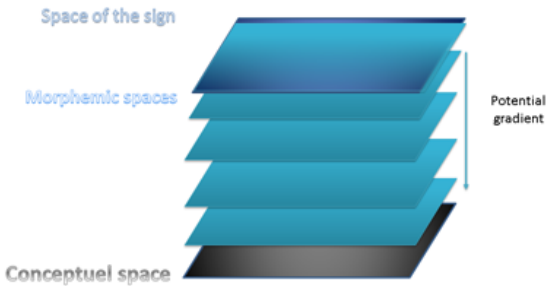



Figure 2: Morphemic subspaces

[1] http://www.archivesaudiovisuelles.fr/FR/_video.asp?format=68&id=117&ress=477&video=81606

## 4.3 General Principle

In these conditions:

- A sign is defined by the same generative potential, which found the mutual determination of agents, by extension the parameters and the morphemes. The potential is the source of the structure,
- The number of linguistic parameters characterizing a sign isn't a priori defined,
- There is an initial equiprobability of linguistic parameters,
- The possible perceptual stability of one or several linguistic parameters can evolve during the realization,
- An attractor results from a morpheme, which could use several linguistic parameters,
- The spaces of the signs and the conceptual spaces are countless and can overlap.

## 4.3 Illustration

This approach based upon the perception-conception character of sign language helps consider lexicalized signs and also high iconicity structures. It helps make the distinction between each lexicalized sign together with a couple of high iconicity structures, which are close to one other.

From the structure of high iconicity in which the form [TABLE LOUIS XV] for example, is reinvested during a transfer, we can emphasize:

- some morphemic subspaces in which the lexicalized sign [TABLE] displays appears in the first instance,

Figure 3: [TABLE] (extract of sign [TABLE LOUIS XV])[2]

- some morphemic subspaces in which the proforms (Cuxac, 2003), specify the distinctive form of the table appropriate to the style [LOUIS XV].

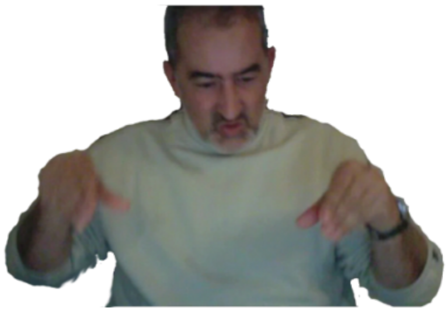


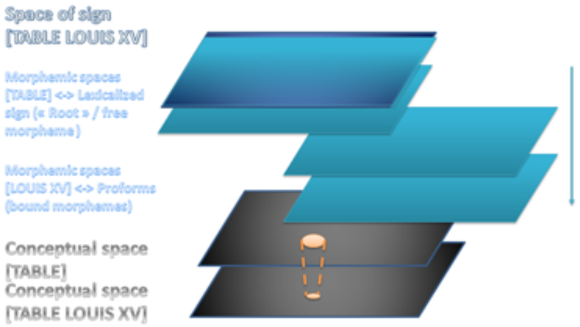
Figure 4: [LOUIS XV] (extract of sign [TABLE LOUIS XV])

[2] On the picture : Moez a French Sign Langage native speaker

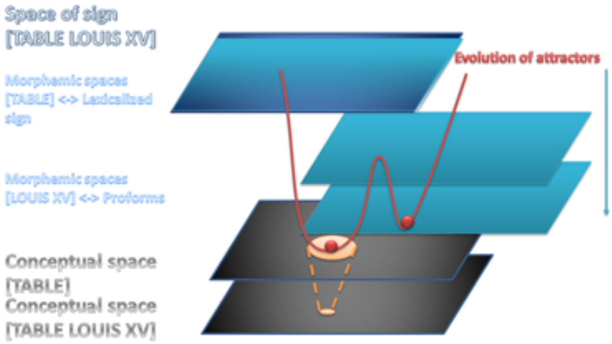
The conceptual space [TABLE LOUIS XV] is a conceptual subspace of [TABLE].

Figure 5: Table of style Louis XV[3]

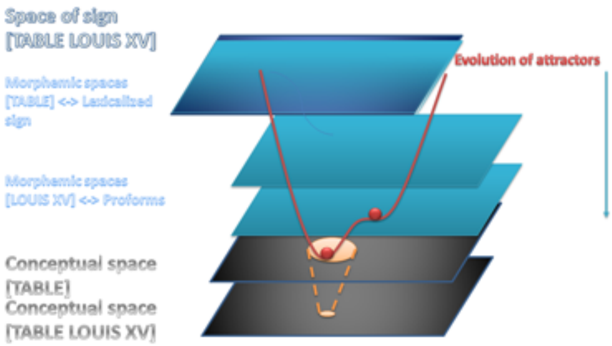
This process can be illustrated by the plans below. Attractors change during period of sign realization. Agents are symbolized by red balls.



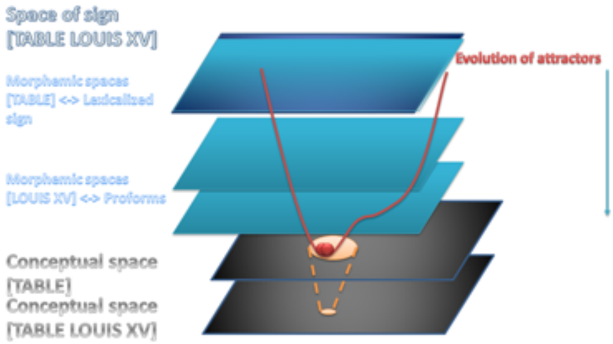

- 1 -




- 2 -




- 3 -




- 4 -

Figure 6: Example of evolution of realization of a structure of high iconicity [TABLE LOUIS XV]

When we approach the bottom of the basin of attraction (the minima of the attractor or "chreode" (Thom, 1973)), which corresponds to the exact meaning of the concept, parameters which are not the most pregnant one contribute to the exact determination of the concept, and they can modify the surface of the basin.

The iconic signs send back to conceptual subspace of a greater granularity than the conceptual space of the lexicalized signs.

The use of the perceptual stability of the morphemes of “secondary” morphemic spaces allow the distinction between two structures of high iconicity, close to each other.

For example, [TABLE LOUIS XV] and [TABLE HENRI II] have the same first morphemic space, which comes from lexicalized sign [TABLE]. Their “secondary” morphemic spaces relative to the legs of the table, for example, allows to distinguish them.

Figure 7: Table style Henri II[4]

The differences between these two signs particularly concern the use of given proforms which specify the distinctive shape of the leg of each table.


Figure 8: Spaces of signs [TABLE LOUIS XV] & [TABLE HENRI II] and conceptual spaces

3 http://www.mariealbertfurniture.com/images/items/Table/03019th1.jpg

4 http://www.french-warehouse.com/Images/Table.jpg

According to the idea introduced by (Blum, 1973; Psotka, 1978; Koenderink, 1984; Koenderink & VAN DOORN, 1986) spreading boundaries on which the process of genesis of a relational shape bases on, is transmitted as a front of wave (Petitot, 1991).
The transition between two signs is characterized in the abstract space by specific type of pass: transition between two lines level tangent. These characteristic points could be used to identify the position (Petitot, 1991) and the determination of the relative distance between close concepts.

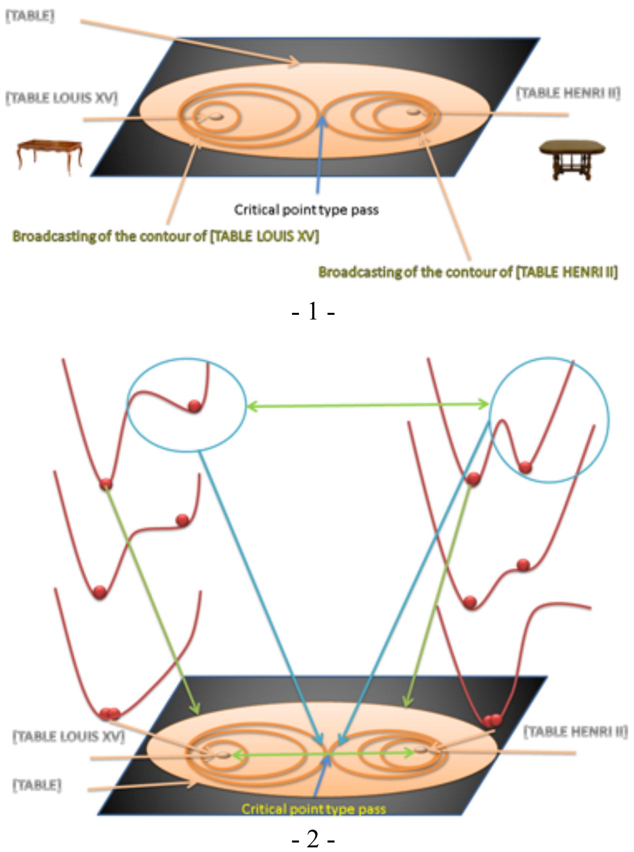



Figure 9: Proximity of concepts [TABLE LOUIS XV] and [TABLE HENRI II]

## 5. Perspectives for sign access on website

Perspectives of this theoretical and general work could be used in the future as a way of accessing a sign in bilingual or monolingual (sign language) dictionaries or ontology, like OCELLES project.
From perception of sign speakers, perspectives of this theoretical and general work could be used in the future as a way of accessing a sign in bilingual or monolingual (sign language) dictionaries or ontology, like OCELLES project. Every user will be able, for example, to give his perceptive point of view about every sign by proposing a morphemic cutting and weighting way based on every linguistic parameter (by proposing eventually new ones). If the sign access is unsuccessful, users will always be able to use isomorphism between sign and conceptual spaces.

## 6. References

Blum, H. (1973, February). Biological Shape and Visual Science. *Journal of Theoretical Biology, 2*(38), 205-287.

Bundgaard, P. F., & Stjernfelt, F. (2010). René Thom's Semiotics and its Sources. In W. Wildgen, & W. W. Per Aage Brandt (Ed.), *Semiosis and Catastrophes René Thom's Semiotic Heritage* (Vol. 10). Wildgen, Wolfgang / Brandt, Per Aage (eds.). Retrieved from Center for Semiotik Nordisk Institut Humaniora Aarhus Universitet: http://www.hum.au.dk/semiotics/docs2/pdf/bundgaard_peer/rene_thoms_semiotics.pdf

Cuxac, C. (2000). Les voies de l'iconicité.

Cuxac, C. (2003). Une langue moins marquée comme analyseur langagier : l'exemple de la LSF. *Nouvelle Revue de l'AIS (Adaptation et Intégration Scolaires)*(23), 19-30.

Dreuw, P., & Ney, H. (2008). Towards Automatic Sign Language Annotation for the ELAN Tool. *LREC Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora.* Marrakech, Morocco.

Filhol, M., Braffort, A., & Bolot, L. (2007). Signing avatar : Say hello to Elsi ! *International Workshop on Gesture in Human-Computer Interaction and Simulation (GW).* Lisbonne.

Friedman, L. A. (1977). *On the other hand: new perspectives on American sign language.* New York: Academic Press.

Koenderink, J. J. (1984). The Stucture of Images. *Biological Cybernetics*(50), 363-370.

Koenderink, J. J., & VAN DOORN, A. (1986). Dynamic Shape. *Biological Cybernetics*(53), 383-396.

Lefebvre-Albaret, F., & Dalle, P. (2010). Video Retrieval in Sign Language Videos : How to Model and Compare Signs? In N. C. Tapias (Ed.), *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10).* Valletta, Malta: European Language Resources Association (ELRA).

Liddell, S. K. (1980). *American Sign Language Syntax.* The Hague: Mouton De Gruyter.

Moody, B. (1980). *Histoire et Grammaire, tome I.* Paris: Ellipses.

Moreau, C., & Mascret, B. (2010). Organizing data in a multilingual observatory with written and signed languages. In E. L. (ELRA) (Ed.), *Language Resources and Evaluation Conference (LREC 2010), 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies* (pp. 168-171). Valletta, Malta: Nicoletta Calzolari (Conference Chair) and Khalid Choukri and Bente Maegaard and Joseph Mariani and Jan Odijk and Stelios Piperidis and Mike Rosner and Daniel Tapias.

Petitot, J. (1991, Septembre). Syntaxe topologique et grammaire cognitive. *Langages*(103), pp. 97-128.

Petitot, J. (n.d.). *Les modèles morpho-dynamiques en* *sciences cognitives et sciences humaines.* Retrieved Février 4, 2012, from Archives Audiovisuelles de la Recherche: http://www.archivesaudiovisuelles.fr/FR/_video.asp?format=68&id=117&ress=477&video=81606

Psotka, J. (1978). Perceptual Process that May Create Stick Figures and Balance. *Journal of Experimental Psychology, (Human Perception and Performance)*, 101-111.

Sallandre, M.-A. (2001, Novembre 9). Va-et-vient de l'iconicité en langue des signes française. *15*.

Stokoe, W. C., Casterline, D., & C –Cronenberg, C. G. (1965). *A dictionary of American Sign Language on linguistic principles.* Washington D.C: Gallaudet College Press.

Thom, R. (1973). *Modèle mathématiques de la morphogénèse.* Paris: Christian Bourgois Éditeur.

Thom, R. (1980 b). *Paraboles et Catastrophes* (éd française Flammarion (1983) ed.). Milan: G. Giorello.

Yau, S.-C. (1992). *Création Gestuelle et début du Langage Création de langues gestuelles chez les sourds isolés.* Hong Kong: Langages Croisés.

# A New Web Interface to Facilitate Access to Corpora: Development of the ASLLRP Data Access Interface (DAI)

**Carol Neidle*, Christian Vogler****
*Boston University, **Gallaudet University
*Boston University Linguistics Program, 621 Commonwealth Avenue, Boston, MA 02215
**Gallaudet University, Technology Access Program, 800 Florida Ave NE, Washington DC 20002
E-mail: carol@bu.edu, christian.vogler@gallaudet.edu

**Abstract**

A significant obstacle to broad utilization of corpora is the difficulty in gaining access to the specific subsets of data and annotations that may be relevant for particular types of research. With that in mind, we have developed a web-based Data Access Interface (DAI), to provide access to the expanding datasets of the American Sign Language Linguistic Research Project (ASLLRP). The DAI facilitates browsing the corpora, viewing videos and annotations, searching for phenomena of interest, and downloading selected materials from the website. The web interface, compared to providing videos and annotation files off-line, also greatly increases access by people that have no prior experience in working with linguistic annotation tools, and it opens the door to integrating the data with third-party applications on the desktop and in the mobile space. In this paper we give an overview of the available videos, annotations, and search functionality of the DAI, as well as plans for future enhancements. We also summarize best practices and key lessons learned that are crucial to the success of similar projects.

**Keywords**: web interfaces, access to corpora, corpus management

## 1. Introduction

Linguistically annotated video corpora for signed languages can be enormously valuable for research in linguistics and computer-based sign language recognition, with many other potential applications, including education. Construction of such corpora is time-consuming, and linguistically controlled data collection yielding high-quality video files requires resources and interdisciplinary collaboration. The substantial investment in corpus development will have greatest benefit if corpora can be shared widely.

A significant obstacle, however, to broad utilization is the difficulty in gaining access to the specific subsets of data and annotations that may be relevant for particular types of research. With that in mind, we have developed a web-based Data Access Interface (DAI), to provide access to the expanding datasets of the American Sign Language Linguistic Research Project (ASLLRP), available at http://secrets.rutgers.edu/dai/queryPages/. The DAI facilitates browsing the corpora, viewing videos and annotations, searching for phenomena of interest, and downloading selected materials from the website. We have also found these same tools invaluable for verifying the consistency of our annotations.

Here we give an overview of available video files and linguistic annotations, summarize current functionalities of the DAI, and discuss directions for ongoing development. We also offer some lessons learned that might be of interest to others engaged in corpus management.

## 2. Available data sets

The DAI now allows access to the National Center for Sign Language and Gesture Resources (NCSLGR) Corpus, ASL videos collected and linguistically annotated at Boston University. Synchronized video files, available in compressed and uncompressed formats, show the signing from the front and side and include a close-up view of the face. Linguistic annotations of manual and nonmanual components of the signing have been carried out using SignStream® (Neidle 2002b) and are available in XML format. Manual signs are represented by unique gloss labels. Annotation conventions are documented (Neidle 2002a, 2007).

Annotations are available for 19 short narratives (1002 utterances) plus 885 additional elicited utterances, all from Deaf native signers of ASL (with most of these data coming from four signers). This constitutes a total of 1,888 linguistically annotated utterances, including 1,920 distinct canonical signs (grouping together close variants) and 11,861 total sign tokens.

Linguistic annotations include the start and endpoints of each sign, identified by a unique gloss label, part of speech, and start and end points of a range of non-manual behaviors (e.g., raised/lowered eyebrows, head position and periodic head movements, expressions of the nose and mouth) also labeled with respect to the linguistic information that they convey (serving to mark, e.g., different sentence types, topics, negation, etc.). The annotations are available via an XML format. For the DTD and documentation of the XML format, see http://www.bu.edu/asllrp/ncslgr-for-download/download-info.html.

## 3. Functionalities of the interface

As shown in Figure 1, the DAI user can search for specific text in gloss fields and can narrow the search to specific classes of signs or search for particular types of classifiers or parts of speech. Figure 2 displays a small section of the alphabetical listing of all signs (based on the selection in Figure 1), with sign variants grouped together, enabling the user to select a particular sign or variant (and potentially a specific signer). Selecting FINISH from this

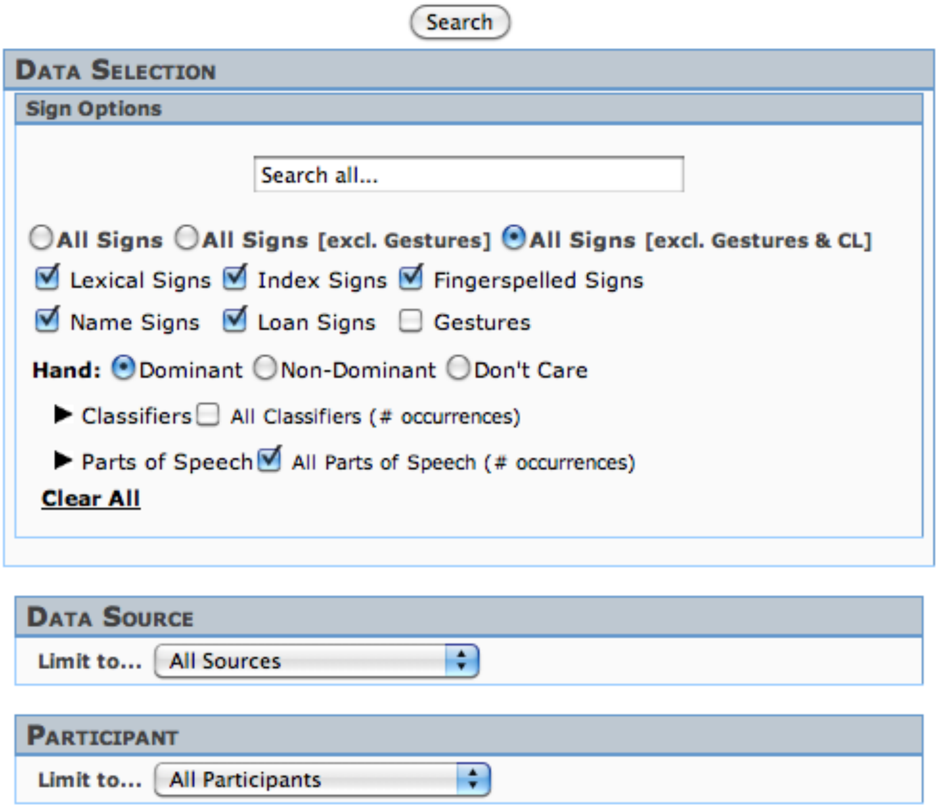



Figure 1: DAI Screenshot showing sample search query

| Sign | Occurrences | Benjamin Bahan | Freda Norman | Lana Cook | Marlon Kuntze | Michael Schlang | Norma Bowers Tourangeau |
|---|---|---|---|---|---|---|---|
| ▼ **FINE** (6) | 44 | 4 | 1 | 0 | 1 | 38 | 0 |
| (2h)FINE | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| FINE | 35 | 3 | 0 | 0 | 0 | 32 | 0 |
| FINE+ | 4 | 0 | 0 | 0 | 0 | 4 | 0 |
| FINE++ | 2 | 1 | 0 | 0 | 0 | 1 | 0 |
| FINE+++++ | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| FINEwg | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| **FINGERSPELL** | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| ▼ **FINISH** (3) | 110 | 27 | 2 | 0 | 4 | 43 | 34 |
| (1h)FINISH | 2 | 1 | 0 | 0 | 0 | 1 | 0 |
| FINISH | 107 | 26 | 2 | 0 | 4 | 41 | 34 |
| FINISH-shake | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| ▼ **FIRE** (3) | 5 | 5 | 0 | 0 | 0 | 0 | 0 |
| (1h)FIRE | 3 | 3 | 0 | 0 | 0 | 0 | 0 |
| (2h)alt.FIRE+++ | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| FIRE | 1 | 1 | 0 | 0 | 0 | 0 | 0 |

Figure 2: DAI Screenshots showing subset of results

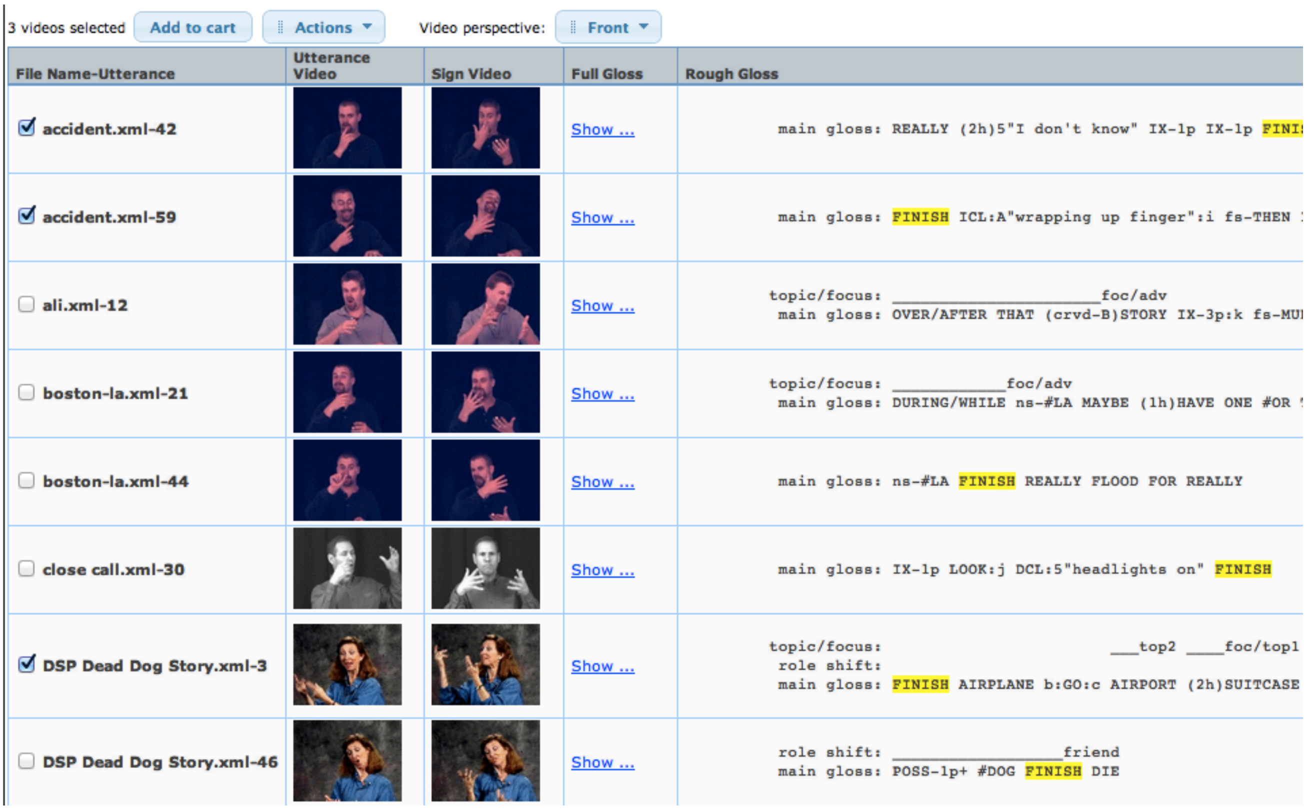



Figure 3: Display of sentences with FINISH, with annotations selected for later download.

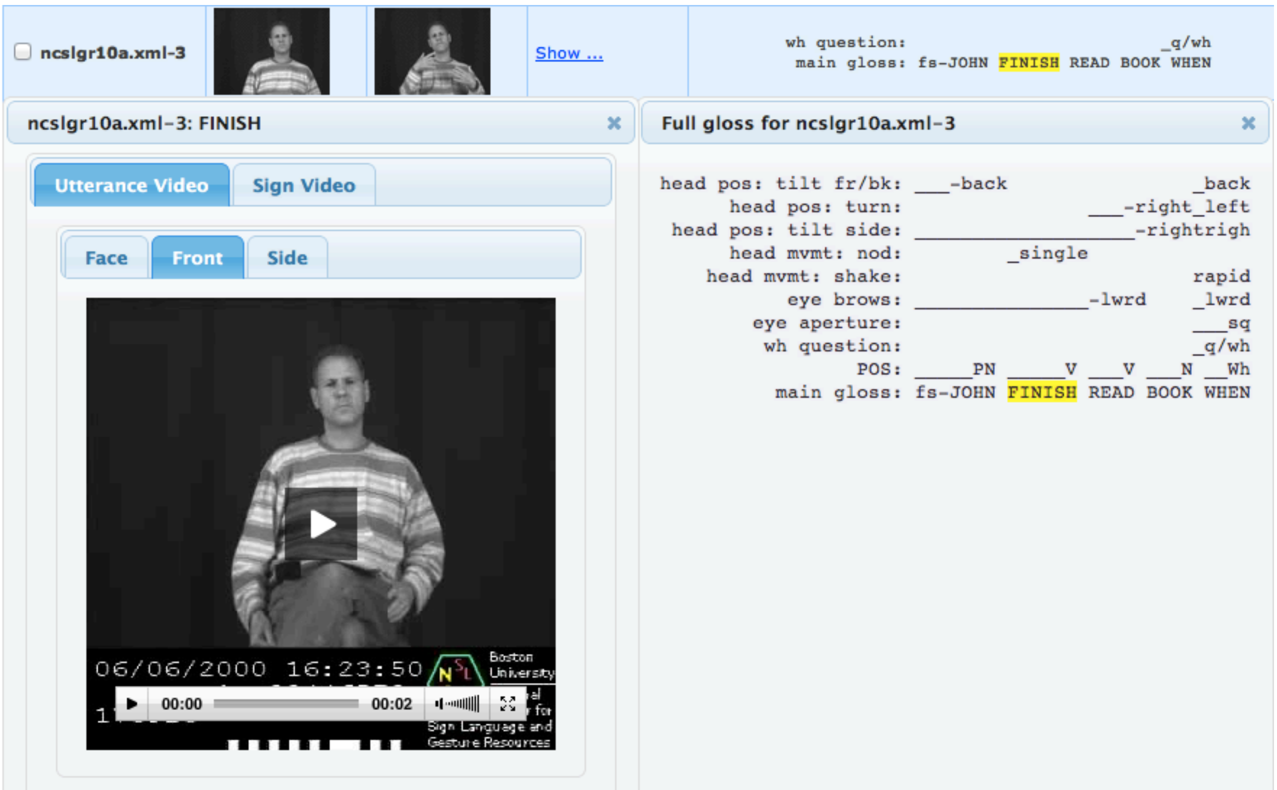



Figure 4: User can view the detailed gloss, and play movies for the sign and the utterance in which it occurs from multiple viewpoints

chart would bring up the display in Figure 3, which includes still images of the relevant material; here the user can switch among the available camera perspectives for each annotation (frontal and face, in some cases also side and stereo camera view), and mark annotations for later download. It is also possible to play back online the videos corresponding to just the sign FINISH or to the utterance containing it, and to display a more complete transcription (Figure 4).

After annotations have been marked for later download, users can call up the download tool. This tool allows them to select which specific video files to download for the selected annotations, where the available choices are sign only, the utterance containing the sign, or the entire story video in which the sign occurs, or any combination of these (along with the linguistic annotations in the XML format). This greatly increases the utility of the DAI, as it is possible to focus on specific signs or linguistic phenomena, and easily obtain a collection of all available videos that exhibit them.

## 4. Best practices and lessons learned

Managing large corpora and making them available to the community entails a unique set of challenges. We present some key lessons that we believe are essential to the success of any similar project.

### 4.1. Presentation of data

**Presenting signs as still images saves users and annotators time and effort.** If the start and end frames of each annotation are presented as thumbnail images, users may be able to detect at a glance whether an annotation is of interest. As compared with having access only to videos (which are time-consuming to watch), availability of still images also greatly speeds up validation and consistency checks – if an annotation is inconsistent with the other ones in the same category, it is likely to manifest in a difference in still frames.

### 4.2. Resource Management

**Keep metadata separate from file names and assets.** Enforcing a consistent coding scheme across thousands of file names and file headers is nearly impossible. It is much easier to keep metadata consistent and up-to-date if it is encoded in a centralized spreadsheet or database. Note that although it may seem to be useful to have some indication of the file's contents in the file naming convention, the downside is that if any of the metadata changes or is corrected later, the file name also would have to be updated to reflect the change, which can break existing external links to the asset.

**Designate only one asset as the authoritative source on metadata, and auto-generate other assets from there.** Having metadata available in multiple formats is often unavoidable; for example, it may need to be present in the database tables, in a spreadsheet for easy manipulation by the team maintaining the corpus, as a web page, and in a textual format for easy distribution to third parties. Unfortunately, there is a high risk of ending up with conflicting metadata for assets, which would result in having to sort out the conflict manually in a laborious process. It follows that only one of these formats can be updated with new and corrected information, and it must be very clear throughout the lifecycle of the project which one it is to be. Moreover, all other metadata assets need to be automatically (i.e. programmatically) generated from the authoritative source, so as to avoid introducing inconsistencies due to human error. Automating this process also makes it more likely that the information is always kept up-to-date across all formats.

**Separate file location and names.** Files can move, as systems are upgraded, or redundancy is built in. If the location is encoded separately, only this part needs to be updated, rather than every link to a file. The DAI uses a two-part schema of the form:

*<url prefix> <path to file>*

where URL prefix points to a location on the server that hosts a collection of related content, such as all XML annotation files, or all videos from a specific camera. Moving the collection to a different location entails updating only a single row in the table that contains the affected URL prefix.

**Be mindful of cross-platform issues.** Different operating systems have different restrictions on file names; for instance, colons are not allowed on Windows. This can cause problems both for users who want to download the data sets, and for copying the data across hard drives with different file systems (such as copying from HFS+ to NTFS and vice versa). In a large corpus that has thousands or even tens of thousands of assets, running into these problems can result in significant delays and expenses. Choosing the intersection of all the restrictions on Windows, Mac OS X and Unix variants – or even restricting file names to alphanumeric characters – is the safest way to proceed, and should be planned and done before any of the data are collected.

### 4.3. Development Processes

**Plan for continuity.** In an academic environment, the design and development must be managed by a project lead, who can commit long-term, and who has the skills to review other contributors' designs and code. Leaving students, who can drop out at any moment or graduate, in charge of the project will induce significant expenses and delays. The project lead, in particular, must understand the overall design of the project, so as to hold hands with new members while they get up to speed on the design and code.

**Use version control on all source files *and* third-party libraries.** The time *will* come when a bug is introduced that can be triaged only by investigating an earlier project revision. Any third-party dependencies must be included in those revisions to guard against the possibility of newer versions of a library being incompatible with the older version of the source code, or the case where a third-party library introduces a bug. Having version control also allows for easy separation of development and release branches, and makes it easy to fix bugs on the release branch, without having to wait for the development branch to get into a releasable state. Modern distributed version control systems, such as Git and Mercurial, make this mode of development especially simple and painless for the developers.

### 4.4. Database Design

**Think queries, not data format.** The types of queries that need to be supported drive the design of the database and the tables. They inform every decision that pertains to the tables, the relationships between tables, database views, and choice of indices, and can result in a representation of the annotations that is markedly different from the one chosen for the annotation file format. Doing the design off the annotation file format is a sure way to run into data management and performance problems down the road. For a concrete example, consider the organization of information in tiers in the annotation file formats and in the program chosen to carry out the annotations. If two tiers are tightly linked – such as in the case of tagging a gloss with the part of speech and the canonical form of the sign – queries are much more efficient if this linkage is made clear in an explicit relationship in the database tables, rather than using a generic tier model.

**Use a collection of tags.** Standardized tagging of annotations (e.g., is a sign fingerspelled? plural? does it use a non-standard location? etc.) provides a powerful and efficient way to search for specific linguistic phenomena. In fact, in the DAI the distinctions among lexical signs, loan signs, classifiers, name signs, fingerspelled signs, indexed signs, and gestures are implemented in this manner (see also Figure 1). In the annotation file formats some of this information may come from separate tiers or be implicit in the naming conventions of glosses. In the DAI database population process, however, this information is extracted and put in an explicit relationship with the annotations, as explained in the previous point on queries.

## 5. Plans for future development and integration of additional data types

Planned enhancements to the DAI include:

1) Integration of other types of corpora;
2) Functionalities to enable additional types of searches;
3) Providing annotations in additional formats;
4) Display of various kinds of statistical information;
5) Integration of new technologies, as they become available.

### 5.1. Integration of additional types of corpora

The interface will be modified to allow integration of other types of corpora, including the American Sign Language Lexicon Video Dataset (ASLLVD), a corpus containing over 3,000 citation forms of lexical signs, each produced by between 1 and 6 native signers, resulting in a total of about 9,000 tokens, which have been annotated for start and end handshapes, among other things (Neidle et al. 2012). Design decisions will have to be made about how best to allow users to move easily among the different types of data sets, e.g., to look up a sign to see variations in production of citation forms by different signers, and to see the sign in context in examples from our corpus of continuous signing.

We would also like to allow access, through this interface, to portions of the *Deaf Studies Digital Journal* (DSDJ) http://dsdj.gallaudet.edu/, edited by Ben Bahan and Dirksen Bauman. (See also http://www.gallaudet.edu/News/Pioneering_digital_journal_to_launch_November_4.html.)

### 5.2. Additional search functionalities

It will, before long, be possible to search for

- Grammatical constructions, such as questions (of various types), negations, conditionals, relative clauses (correlatives), topics;
- Nonmanual signals, such as eye aperture, head tilt, raised/lowered eyebrows, body lean;
- Words in the English translation field.

Searches for text based on Sign ID (represented by a unique English-based gloss label) corresponding to a specific ASL sign will include the ability to restrict text searches to whole word (by default) or to search for text strings. We will incorporate searches based on:

- Video properties: e.g.: types of available viewpoints (frontal, side, stereo, face); availability of color video
- Availability of calibration data
- Subject wearing long/short sleeves
- Subject wearing glasses.

We will also explore the possibilities of allowing searches to be based on:

- Frequency (making it possible to search for items that have a minimum number of tokens);
- Sign duration;
- Number of subjects (making it possible to search for productions of a minimum number of signers);
- Specific signers (making it possible to view the set of productions of one or several individuals);
- Characteristics of signers (particularly as the corpus grows), such as gender or age range.

Furthermore, since the new corpora will include other types of annotations, we will also need to extend the search functionalities to enable appropriate searches of those data sets. This will include the ability to limit searches to specific sign types (lexical signs, index signs, classifier constructions of different types, fingerspelled signs, loan signs, name signs). We will also provide a way to search for the initial and/or final hand shape for the sign, or other phonological properties of the sign (e.g., signs containing a particular hand shape or movement type, or signs articulated with one or two hands).

We welcome suggestions about features that might be useful for different communities of potential users. This web-based interface could be especially useful for those who use ASL as a primary language and for those learning the language. We will be working with prospective users from these groups to design tools to facilitate the kinds of access that might be anticipated.

### 5.3. Annotations in other formats

Although the annotations currently are available in an easy-to-parse SignStream-specific XML format, we realize that researchers have their own preferences with respect to what annotation software they use. We plan to make the annotations, at a minimum, available in the ELAN EAF format and welcome suggestions as to what other formats should be supported.

### 5.4. Display of statistical information about the corpora

We plan to add functionality to view statistics about common metrics for measuring the size of the available corpus, including number of utterances, signs, length of the videos, size in MB, and so on. These numbers will make it possible to compare the key characteristics of the corpus to related work at a glance.

### 5.5. Integration of new technologies for display of, and access to, data, as they become available

Our future plans include the display of information on annotations in new ways. One of these ways consists of integrating data that can be measured by a computer as opposed to humans, such as graphs and numbers showing changes in eyebrow height and head movement for large samples of the corpus.

We also plan to facilitate the integration of the data with third-party and mobile applications. The biggest promise of having the DAI available on the web, as opposed to distributing it off-line, lies in making it available as an online service, such that cloud-based applications can take advantage of it. For example, a sign language dictionary available on mobile devices would be able to search for and retrieve concrete usage examples for the sign in question, which can be an invaluable tool for second language learners. Taking this approach will also enable other creative ways to use a corpus that we have not yet even envisioned.

## 6. Resources

- Database access interface: http://secrets.rutgers.edu/dai/queryPages/
- XML file format and DTD: http://www.bu.edu/asllrp/ncslgr-for-download/download-info.html.

## 7. Acknowledgments

The production of the NCSLGR corpus has involved significant contributions by many people. The Computer Science personnel responsible for the technical aspects of the video capture and management of video files included Stan Sclaroff and Vassilis Athitsos. The linguistic aspects of the data elicitation and annotation were carried out by Carol Neidle, Dawn MacLaughlin, Robert G. Lee, Benjamin Bahan, Michael Schlang, and many other ASLLRP participants. We are indebted to David Greenfield, the programmer for SignStream® through version 2, and to Iryna Zhuravlova who has been developing the Java reimplementation with additional features. Preliminary versions of the interface were developed by Ben Waber, Eugene Khafizov, and others. Thanks also to Richard Goodrow for his assistance with current development efforts. We are grateful for support from the National Science Foundation (NSF) that made this research possible (CNS-1059218 and CNS-1059221). The work reported here on development of the interface has also been supported in part by grants from NSF (CNS-0958442 and CNS-0964385).

## 7. References

Neidle, Carol. 2002a. *SignStream™ Annotation: Conventions used for the American Sign Language Linguistic Research Project.* American Sign Language Linguistic Research Project Report No. 11, Boston University.

—. 2002b. SignStream™: A Database Tool for Research on Visual-Gestural Language. *Journal of Sign Language and Linguistics 4*.203-14.

—. 2007. *SignStream™ Annotation: Addendum to Conventions used for the American Sign Language Linguistic Research Project.* American Sign Language Linguistic Research Project Report No. 13, Boston University.

Neidle, Carol, Ashwin Thangali & Stan Sclaroff. 2012. Challenges in Development of the American Sign Language Lexicon Video Dataset (ASLLVD) Corpus. Paper presented to the 5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC 2012, Istanbul, Turkey, 2012.

# Challenges in Development of the American Sign Language Lexicon Video Dataset (ASLLVD) Corpus

**Carol Neidle*, Ashwin Thangali**, Stan Sclaroff****
Boston University
*Linguistics Program, 621 Commonwealth Ave.; **Computer Science Department, 111 Cummington St.
Boston University, Boston, MA 02215
E-mail: carol@bu.edu, tvashwin@cs.bu.edu, sclaroff@bu.edu

**Abstract**

The American Sign Language Lexicon Video Dataset (ASLLVD) consists of videos of >3,300 ASL signs in citation form, each produced by 1-6 native ASL signers, for a total of almost 9,800 tokens. This dataset, including multiple synchronized videos showing the signing from different angles, will be shared publicly once the linguistic annotations and verifications are complete. Linguistic annotations include gloss labels, sign start and end time codes, start and end handshape labels for both hands, morphological and articulatory classifications of sign type. For compound signs, the dataset includes annotations for each morpheme. To facilitate computer vision-based sign language recognition, the dataset also includes numeric ID labels for sign variants, video sequences in uncompressed-raw format, camera calibration sequences, and software for skin region extraction. We discuss here some of the challenges involved in the linguistic annotations and categorizations. We also report an example computer vision application that leverages the ASLLVD: the formulation employs a HandShapes Bayesian Network (HSBN), which models the transition probabilities between start and end handshapes in monomorphemic lexical signs. Further details and statistics for the ASLLVD dataset, as well as information about annotation conventions, are available from http://www.bu.edu/asllrp/lexicon.

**Keywords**: American Sign Language, lexicon, computer-based handshape recognition

## 1. Introduction

The American Sign Language Lexicon Video Dataset (ASLLVD) arose from collaboration among computer scientists and linguists to develop sign lookup technology (Athitsos et al., 2010). Several multimedia resources for ASL are under development, but available interfaces for sign lookup remain less than optimal. The ideal interface would enable users to search the dataset simply by video-recording a sign and relying on computer-based sign recognition for lookup.

To train computer algorithms to distinguish and recognize ASL signs, we created a corpus with ~3,000 signs from up to six native signers. Our sign recognition and retrieval algorithms rely in part on linguistic models. Initial research has focused on the benefits for robust sign recognition of exploiting constraints on the relationship, in monomorphemic lexical signs, between start and end handshapes (and between the two hands, in two-handed signs) (Thangali et al., 2011).

Linguistic annotations have been carried out to facilitate this research. Specifically, we assigned each sign a unique gloss label; identified variants of specific lexical items; and labeled start/end handshapes. This corpus will be shared publicly once verifications are complete. It will also be integrated with another corpus that we already make available for online browsing and download: our National Center for Sign Language and Gesture Resources (NCSLGR) corpus, which can be accessed from http://www.bu.edu/asllrp/ (see Neidle & Vogler (2012)). Through extensions to our web interface, it will be possible to search our lexical and continuous signing data in various ways, and to go back and forth between different data types, e.g., between viewing a sign in citation form or produced in a natural context.

For verifying these large data samples—to enforce consistency in labeling and in groupings of sign variants—we have developed a powerful tool: the Lexicon Viewer and Verification Tool (LVVT). We will (a) describe the data collection, (b) discuss challenges for elicitation, consistent annotation, and classification of data, (c) present a brief overview of the data that we have amassed and statistics thereof, (d) describe a computer science research project that leverages the detailed annotations of the ASLLVD dataset, and (e) outline directions for future research.

## 2. Data collection

Videos were captured using four synchronized cameras. Thus for each sign production, we have a side view of the signer, a close-up of the head region, a half-speed high resolution front view, and a full resolution front view.

The consultants, ASL native signers, were shown video prompts (from the *Gallaudet Dictionary of American Sign Language* (Valli, 2002)) and asked to reproduce the signs as they naturally would (or not, if they do not use that sign). Signers did not always produce the same sign shown in the prompt. In cases where a signer recognized and understood that sign but used a different sign or a different version of the same sign, divergences showed up in the data set. So, in reality, a given stimulus resulted in productions that may have varied in any of several different ways: production of a totally different but synonymous sign; production of a lexical variant of the same sign; production of essentially the same sign but differing in subtle ways with respect to the articulation.

As displayed in Figure 1, we collected a total of 3,314 distinct signs, including variants (for a total of 9,794 tokens). Among those were 2,793 monomorphemic lexical signs (8,585 tokens) and 749 tokens of compounds, which

| Class of signs | Number of signs | Number of sign variants | # sign variants with { 1, 2, 3, 4… } consultants | | # tokens (examples) per sign { 1,2,…,6, >6 } | | Number of sign tokens | |
|---|---|---|---|---|---|---|---|---|
| Monomorphemic lexical signs | 2,284 | 2,793 | x1 | 621 | 587 | x1 | 8,585 | Two-handed |
| | | | x2 | 989 | 858 | x2 | | 5,713 |
| | | | x3 | 394 | 386 | x3 | | 67% |
| | | | x4 | 563 | 491 | x4 | | |
| | | | x5 | 85 | 142 | x5 | | One -handed |
| | | | x6 | 141 | 154 | x6 | | 2,873 |
| | | | | | 175 | >6 | | 33% |
| Compound signs | 289 | 329 | x1 | 129 | 117 | x1 | 749 | |
| | | | x2 | 106 | 107 | x2 | | |
| | | | x3 | 48 | 46 | x3 | | |
| | | | x4 | 33 | 33 | x4 | | |
| | | | x5 | 4 | 11 | x5 | | |
| | | | x6 | 9 | 13 | x6 | | |
| | | | | | 2 | >6 | | |
| Number signs | 76 | 88 | | | | | 260 | |
| Loan signs | 46 | 52 | | | | | 136 | |
| Classifier constructions | 27 | 31 | | | | | 38 | |
| Fingerspelled signs | 21 | 21 | | | | | 25 | |
| ALL | 2,742 | 3,314 | -- | -- | -- | -- | 9,794 | |

Figure 1. Overview of statistics from the dataset

provide fertile ground for studying assimilation effects. Column 4 shows the total number of sign variants we have as produced by 1 signer, 2 signers, etc. Since in some cases we had more than one example per signer, the total number of tokens per sign was, in some cases, greater than 6.

## 3. Resources to be made available

Linguistic annotations are in the final stages. Once this has been completed, the video files and associated annotations will be made publicly available. Details about this will be provided on our website when the materials are ready for release (http://www.bu.edu/asllrp/lexicon).

### 3.1. Video data

Video sequences will be made available in uncompressed-raw format, along with camera calibration sequences and software for skin region extraction. Hand location bounding box coordinates (either in each video frame or only for the start and end frames of a sign) will be accessible for a subset of signs in the dataset.

### 3.2. Linguistic annotations

Linguistic annotations, carried out using SignStream®3 (beta), will also be made available in XML format. These include gloss labels and start/end time codes for each sign, labels for start and end handshapes of both hands, morphological classifications of sign type (lexical, number, fingerspelled, loan, classifier, compound), and articulatory classifications (1- vs. 2-handed, same/different handshapes on the 2 hands, same/different handshapes for sign start and end on each hand, etc.). For compound signs, the dataset includes annotations as above for each morpheme. To facilitate computer vision based sign language recognition, the dataset also includes numeric ID labels for variants of a sign.

## 4. Challenges faced for linguistic annotation and categorization of signs

This data set will serve as the basis for development of sign lookup technology. That is, we ultimately want to be able to identify automatically, from a video, the identity of the sign that was produced, so that this can serve as an entryway for lookup in an ASL dictionary. Some of the decisions with respect to annotation were made with this kind of application in mind. For example, for such research, it is essential that there be a 1-1 correspondence between sign and label. The American Sign Language Linguistic Research Project (ASLLRP) based at Boston University has been using unique gloss-based ID labels throughout the development of all of our corpora — including our NCSLGR corpus — since the early 1990's.[1] Although our annotation conventions (Neidle, 2002, 2007) have evolved slightly to deal with issues that have

[1] For further discussion of ID-glosses, in particular, and the types of issues that arise in the annotation of signed language corpora, see Johnston (2010).

arisen as the corpus has expanded (a 2012 version documenting recent modifications is currently in preparation), the essential goal with respect to the gloss-based labeling of signs has remained constant:

> To facilitate both linguistics and computer science research, we have tried our best to settle on conventions to ensure that every time a particular English gloss is used, it corresponds to a unique ASL sign, and conversely, that the same ASL sign will have a predictable English gloss. (Neidle, 2007: p. 3)

There were challenges in ensuring consistency across annotators, and in assigning unique gloss labels while also enforcing consistency with glossing conventions for our other corpus. There were also challenges involved in assigning consistent handshape labels to hand configurations that sometimes did not exactly match any of our 86 canonical handshapes (we included an 87th that we labeled as a "relaxed handshape" and an "other" option when the handshape used failed to correspond with any of the other handshapes):

http://www.bu.edu/asllrp/cslgr/pages/handshape-palette.html

For handshapes that fell in between two of our existing handshapes, the danger is that what might appear, from the annotations, to be variations in production might, in fact, turn out to be merely inconsistencies in how the same handshape had been annotated. To some extent, this is unavoidable given the gradient nature of some of the handshape productions, but the ability to view exemplars of a given sign together, and to search for handshape annotations across the dataset, makes it considerably easier to do side-by-side comparisons and to increase the degree of consistency in the annotations.

We encountered various thorny issues in assessing variation: When should two productions be considered distinct signs, variants of the same sign, or the same variant of a single sign? In principle, we did not separate out as variants productions differing solely with respect to general ASL linguistic processes (not specific to the particular lexical item). For example, productions that differed in an alternation between a flat-B and B-L handshape (e.g., for the dominant hand of OVER/AFTER) were considered to be instantiations of the same sign variant, since there is, in general, widespread variation between these two handshapes, not restricted to this particular sign. In fact, there are 157 forms where variation between these two handshapes was attested. There are other cases, however, where the manifestation of two different handshapes is tightly linked to the particular sign, an example being the alternation between the A and 5 start handshape in MAN (or WOMAN). This kind of alternation is not widespread and is restricted to a small set of lexical items. Thus, MAN and (5)MAN have been distinguished in glossing and classified as two variants of the same sign. These examples are illustrated in Figure 2.

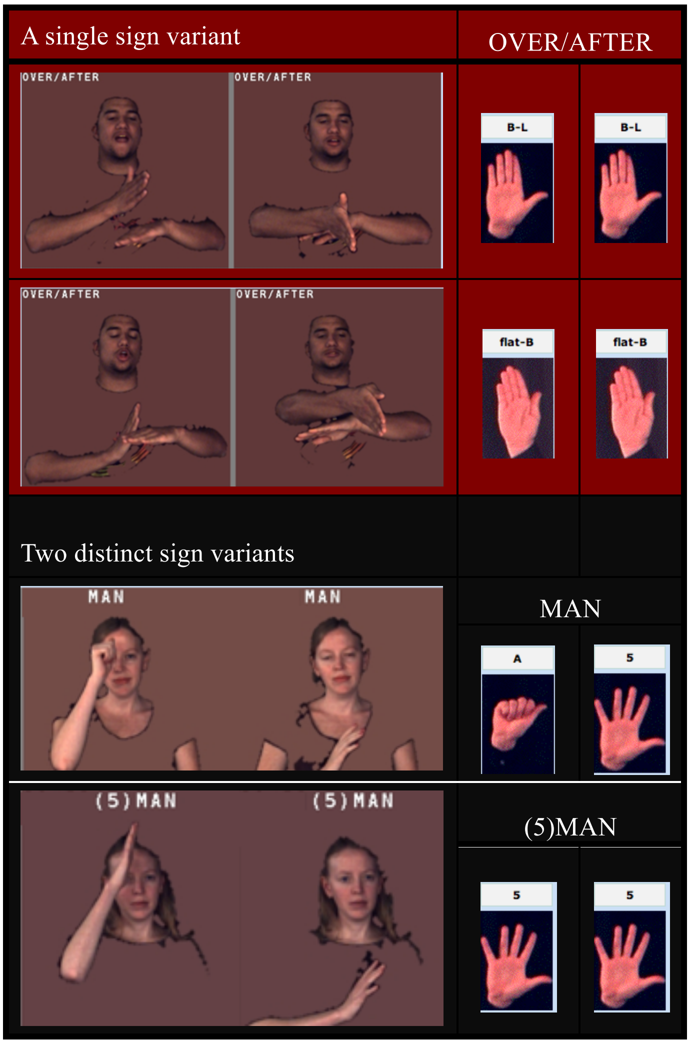



Figure 2. Predictable variation in handshapes [B-L/flat-B] vs. lexically dependent variation [A/5]

However, the status of handshape variations was not always clear, particularly because we often had only one or few tokens of each sign per signer, so issues of inter- vs. intra- signer variations were sometimes difficult to tease apart at the time annotations were initially conducted. Such issues are quite interesting, though, and become more tractable when we can examine patterns across the entire dataset and probe further with signers about the equivalency or non-equivalency in their own signing of specific handshape variations for a given sign. Given the intended application (computer-based sign lookup), we focused on the way the signs were produced. In the case of homonyms, we used the same gloss for all, despite the fact that it was often impossible in the labeling to account for the full range of meanings. We expect that the eventual dictionary lookup will provide access to the various distinct meanings that can be associated with a given production. However, here again, there were some difficult cases, where some but not all realizations of two given signs were distinguishable from one another. For example, we classified CHEW and WASH in Figure 3 as distinct signs, even though in many cases, it would be hard to distinguish productions of the two.[2]

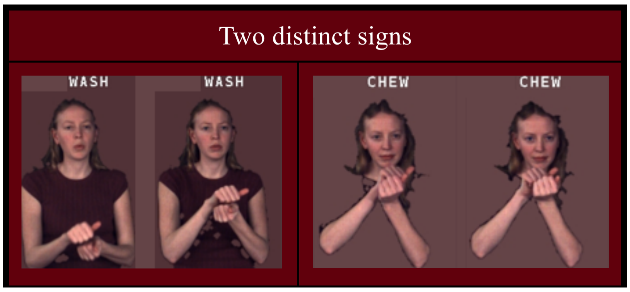



Figure 3. WASH vs. CHEW

In such cases where there is some degree of similarity that may be relevant for an eventual lexical lookup process, we have grouped the lexical items together, but we consider them to be linguistically distinct.

We generally did not separate out signs for which production differed only in the number of repetitions or reduplications of the base form, even though this frequently (but not always) results in a difference in meaning. We indicated the number of repetitions through the use of the + symbol, but considered the productions that differed in this way to be instantiations of the same sign variant. In cases where the productions with and without reduplication differ in meaning, disambiguation would need to occur at the dictionary lookup stage.

The challenges that we have faced with annotation and categorization of signs—which are far too numerous and varied to list exhaustively in the present context—are not unique to this project. The same kinds of issues necessarily face other sign lexicon projects. For that reason, we believe that the kind of tool discussed in the next section has the potential to facilitate such efforts and increase the accuracy of annotations and classifications.

## 5. Tool for browsing and verification

The Lexicon Viewer and Verification Tool (LVVT) was conceived and developed to aid in viewing, comparing, verifying, and modifying SignStream® annotations. The LVVT is designed to assist the annotator in the daunting task of ensuring consistency of the labeling of glosses and articulatory attributes across several thousand tokens.

In developing the LVVT we drew inspiration from the search and browsing functionality implemented in the ASLLRP Data Access Interface (DAI) (Neidle & Vogler, 2012). The LVVT extends the DAI's feature set by enabling users not only to browse the data, but also to modify displayed attributes for signs. Presently, the attributes supported are the gloss labels, start/end handshapes, start/end timecodes in video, and the morphological and articulatory classifications of signs. In addition to presenting an interface for the annotator to search, browse, compare and modify annotations for signs, we believe an important contribution of the LVVT is in facilitating groupings of signs to be constructed.

We define a two-level grouping layout for signs in the lexicon dataset so as to clearly distinguish cases where we have distinct *signs* from those in which we are dealing with *sign variants*.

(1) Occurrences of a given sign may be subdivided into several distinct variants. Occurrences classified as belonging to a single *sign variant* are deemed to differ from one another only as a result of general language processes that are not sign-specific. As mentioned in Section 4, we do group together signs that differ in the presence or absence of reduplication (indicated by '+'); thus all examples considered to be instantiations of a single variant may not be identical in meaning. A sign with four variants is illustrated in Figure 4 (bottom): ABORTION_2 differs from ABORTION in the orientation of the non-dominant hand, (1h)ABORTION is a one-handed sign and, (S)ABORTION uses a different start handshape on the dominant hand.

(2) Loosely related (but distinct) signs can be further organized by means of higher-level groupings. This is intended solely to aid in navigating the dataset. These groupings are for our convenience in working with the data and have no linguistic significance.

Each grouping of signs and sign variants is annotated with a unique gloss label, and with a pair of numeric IDs to denote its location in the upper and lower levels of the two-level grouping layout.

The general listing of signs in the sign index is shown in the left column of Figure 4 (top). The higher-level groupings are visible from the presence (or absence) of a ♦ (diamond) prefix, which indicates that a contiguous sequence of gloss labels belong to the same sign collection, e.g., HOW-MANY and (1h)HOW-MANY are in one high-level grouping; HUMBLE, (H)HUMBLE, and (1)HUMBLE are in another. In both of those cases, those groupings contain a single sign with more than one variant. Note also, however, that HUSBAND (a monomorphemic sign) and the closely related compound BOY+MARRY (from which HUSBAND evolved), are also grouped together, albeit as distinct signs.

The LVVT has proven to be very useful for comparing similar forms, for ensuring consistency of annotations, and for determining how they should best be categorized in relation to one another. Of particular benefit is the ability to view still images of the start and end frames together across the range of sign tokens, and to play the video files from two different camera views of multiple signers producing the same sign simultaneously. Figure 5 depicts a snapshot of a video sequence presented to the annotator for the purpose of verifying consistency in the grouping. By viewing the data in these ways, we can discover sign variants that had not previously been noticed as distinct by the annotators, and conversely can discern similarities in production of signs that previously may have been categorized as distinct.

[2] According to Vicars (2012), "the movement of 'wash' is two steady circles. The movement of 'chew' is [very slightly] more elliptical and uses a bit (but not much) more shoulder/elbow movement as the hand circles toward the body."

Through use of the LVVT, corrections to glosses, start/end frames, handshapes, and/or classifications and groupings of signs can also be carried out directly, in a simple, intuitive way, e.g., by clicking on a handshape icon associated with a sign to bring up the handshape palette, then clicking to select a replacement for an erroneous handshape. The user interface elements for annotating the gloss and other attributes for each sign are displayed in the last column in Figure 4 (top).

Various corpus properties can also be displayed, and many different types of searches can be performed. For example, Figure 6 shows part of a chart illustrating, for monomorphemic signs, the most likely end handshape given a particular start handshape. The particular start and end handshape combinations can (with a single mouse click) be entered into a search box in the LVVT, and all relevant examples will be listed. Search queries can be carried out for particular handshapes (start and/or end of dominant and/or non-dominant hands), potentially in combination with a variety of morpho-phonological properties and categorizations.

The LVVT also includes an interface for working with compound forms. The LVVT presents the annotator with the same set of features for annotating morphemes in

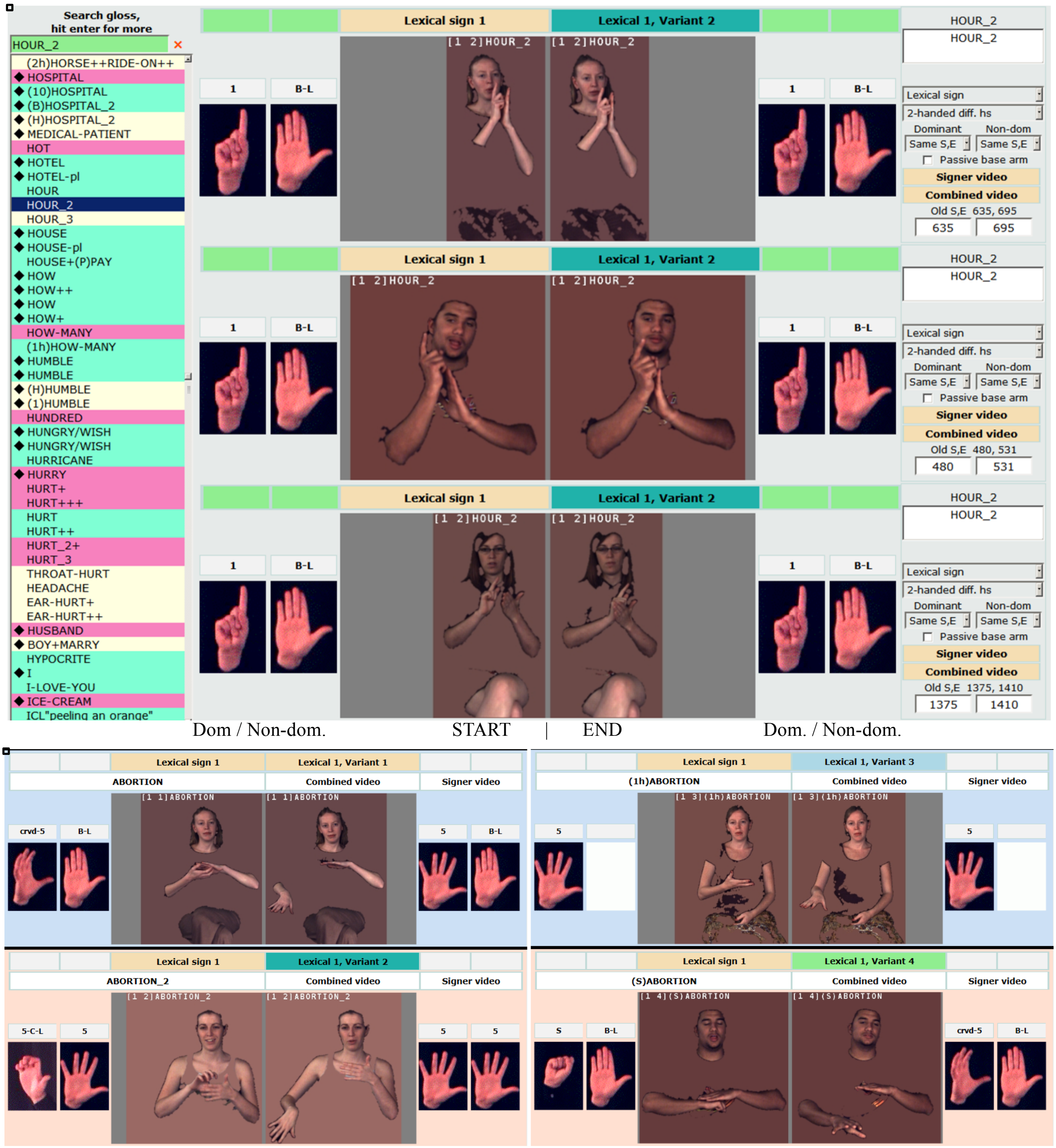



Figure 4. Lexicon Viewer & Verification Tool (LVVT): main page with listing of signs (top) and display of sign variants (bottom)

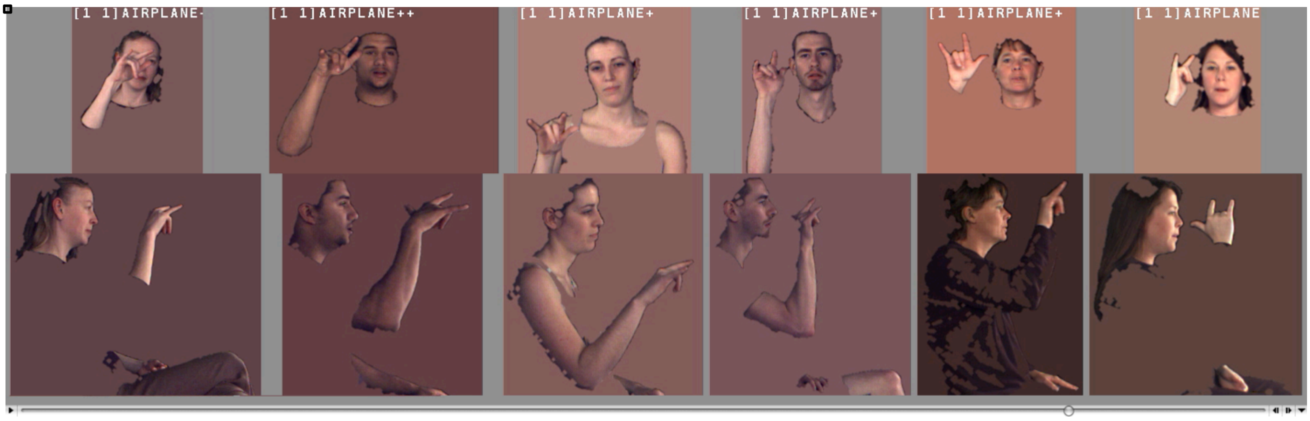



Figure 5. The LVVT enables combined videos (front and side views) to play simultaneously

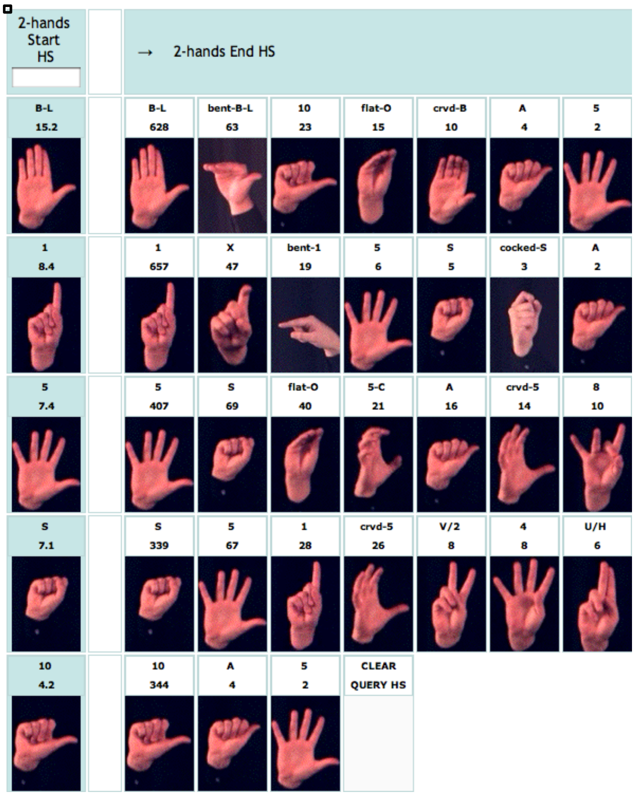



Figure 6. Excerpt of chart showing likely end handshape given the start handshape on the left; shown in order of decreasing frequency

compound signs as are available for monomorphemic signs. This interface makes it simple to view all compounds that share a particular morpheme, for example. The morphemes in compound signs can also be compared to their non-compound versions to ascertain consistencies in glossing and other annotated attributes.

## 6. Computer science research

One important goal of the ASLLVD is to support development and evaluation of algorithms that can distinguish and recognize ASL signs. As an example application, we have developed a computer vision approach for handshape inference that utilizes a HandShapes Bayesian Network (HSBN) (Thangali, et al. 2011), which models the transition probabilities between start and end handshapes in monomorphemic lexical signs (i.e., simple signs).

A challenging aspect of handshape identification by computer from video is the fact that 3D hand configurations are visible only as 2D images. We demonstrate that the HSBN is able to help in the handshape recognition problem by exploiting general properties for how handshapes are sequenced and how their variations are realized in simple signs. While many previous approaches (e.g., Bowden et al., 2004; Liwicki & Everingham, 2009; Vogler & Metaxas, 2004) have trained Hidden Markov Models (HMMs) that are specific for each sign/utterance to be recognized, the HSBN represents phonological properties that are applicable to all simple signs. The HSBN parameters are automatically learned from the linguistic annotations of signs in the ASLLVD dataset.

The annotations for each sign in the ASLLVD that are used in training the HSBN include: the handshape numeric ID for the start and end handshapes on each hand, the bounding box coordinates of each hand in the start and end frames, and a classification denoting each sign as either one-handed, two-handed:different handshapes, or two-handed:same handshapes. The HSBN training algorithm also exploits the property that the signs in the dataset are grouped into variants (as in Figure 8). Since the variations in handshape within each group are produced as a result of general language processes that are not specific to a particular sign, the HSBN representation is able to model such variations.

Figure 7 illustrates the HSBN graphical models for the three main articulatory classes. Each node in the graphical model represents a variable in the HSBN. Each HSBN comprises three layers. The lowest layer represents the actual image observations provided to the model; these are the cropped images of each hand at the start and end of the sign. The nodes in this layer are shaded to indicate that they are observed (given) during training and inference. The middle layer in the HSBN represents the IDs of the realized handshapes on each hand. Nodes in the middle layer are partially shaded to denote that annotations for handshape IDs are available in the training set, but the IDs must be inferred (i.e., they are not given) when the trained HSBN is used for recognizing

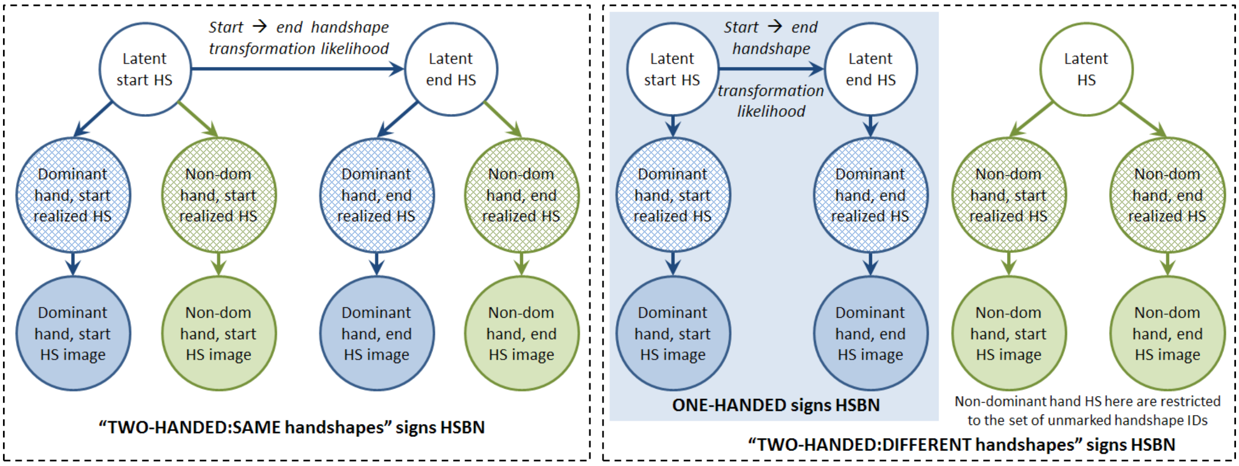



Figure 7. The model on the left represents the HSBN for two-handed:same handshapes signs. The model shown inset on the right represents the HSBN for one-handed signs. In two-handed:different handshapes signs, the HSBN for handshapes on the dominant hand is the same as that of one-handed signs; handshapes on the non-dominant hand are, however, limited to a small number of unmarked handshapes and hence are represented using a separate HSBN.

handshapes in signs. The top layer of the HSBN is a "latent variable" layer that represents the unknown, underlying start and end handshapes (since only realized handshapes are known in training, we model the underlying handshape as a hidden layer in the HSBN whose representation is learned during model training).

The arrows in the graphical models (Figure 7) denote different conditional probability distributions in the HSBN. The horizontal arrow in the top layer represents dependency of the end handshapes on the start handshape, i.e., the likelihood that certain hand configurations appear as end handshapes among simple signs that use a specific start handshape. The arrows connecting the top layer and the middle layer serve two purposes: (a) they represent handshape variability, wherein closely related handshapes may be variant realizations of a hypothetical underlying handshape, and (b) the two pairs of arrows in two-handed-same signs represent bilateral symmetry in the start/end handshapes. The arrows between the middle layer and the lowest HSBN layer represent the relationship between the handshapes that are produced by the signer and their observed images.

Handshapes of the dominant hand in all three sign classes, and handshapes of the non-dominant hand in two-handed:same handshape signs, share the same phonological properties with regard to start/end handshape transition and handshape variation. The HSBN is thus learned using handshape annotations for signs from all classes excluding handshapes on the non-dominant hand in two-handed:different handshape signs. An auxiliary HSBN to model the latter category is much simpler because the handshapes on the non-dominant hand are restricted to a small set of unmarked handshapes without change in handshape between the start and end points of the sign.

The formulation has been evaluated in the task of handshape classification using training and test data taken from the ASLLVD. Handshape recognition accuracy is evaluated on a sequestered test set consisting of 1962 {start, end} handshape image pairs obtained from 657 signs (333 one-handed / two-handed:different handshape and 324 two-handed:same handshapes signs). The remaining 6862 simple signs in the ASLLVD are used in the training. As a baseline handshape recognition method, we use an algorithm to assess similarity in appearance among pairs of handshape images (Thangali et al., 2011). Handshape images of the test signer are excluded from the database used for handshape retrieval. Its rank-1 recognition accuracy is 30.4% (597 of 1962). The proposed HSBN exploits information about handshape candidates retrieved for all {start, end} handshape pairs in the query and thus returns a more coherent collection of inferred handshapes. Performing this inference improves rank-1 recognition accuracy to 44.4% (871 of 1692). We believe that this demonstrates the promise of incorporating linguistic constraints in our recognition system, and the training data from the annotated corpus makes learning such models possible.

## 7. Future aspirations

Once verifications are complete, this set of >3,000 signs, annotated within SignStream®, will be turned into a "sign bank," so that annotators can take advantage of the stored phonological information (which can be further modified) to make the annotation process considerably more accurate and efficient. The annotator will be able to select from available signs and sign variants, and add additional signs or sign variants to the repertoire.

The lexicon corpus data will be released in various forms, including a spreadsheet showing the range of handshape variations for each of the signs in the dataset. For illustration, see Figure 8. This display makes it easy to scan visually for variations in handshapes, for example. As shown in this small sample, the A, S, and 10 handshapes frequently occur in alternation within a single sign variant (despite the fact that they are contrastive for certain signs).

Future plans include integration of the lexicon data with our other datasets, through the Data Access Interface (DAI) that we have been developing, initially to provide

| Main Gloss | Consultant | Main Gloss | Variant | D Start HS | N-D Start HS | D End HS | N-D End HS |
|---|---|---|---|---|---|---|---|
| ACCIDENT | ============= | ============= | ============= | ============= | ============= | ============= | ============= |
| | 1 | ACCIDENT | ACCIDENT | S | S | S | S |
| | 2 | ACCIDENT | ACCIDENT | S | S | S | S |
| | ------------ | ------------ | ------------ | ------------ | ------------ | ------------ | ------------ |
| | 1 | ACCIDENT | (5)ACCIDENT | 5 | 5 | A | A |
| | 2 | ACCIDENT | (5)ACCIDENT | 5 | 5 | 10 | 10 |
| | 3 | ACCIDENT | (5)ACCIDENT | 5 | 5 | S | S |
| | ------------ | ------------ | ------------ | ------------ | ------------ | ------------ | ------------ |
| | 4 | ACCIDENT | (3)ACCIDENT | 3 | 3 | A | A |

Figure 8. Excerpt from the summary spreadsheet showing the variants produced by each of the consultants

access to our NCSLGR corpus (of continuous signing: sentences and short narratives), as described by Neidle & Vogler (2012). The interface will be designed to enable searching through the corpora separately, using appropriate tools for each, as well as going back and forth between display of lexical citation forms and of signs in context. Thus, this will require enhancement of our web interface to facilitate searching, browsing and downloading the kind of data and annotations that are contained in the ASLLVD. Ultimately, the plan is to incorporate many of the search functionalities of the LVVT into our main web interface, the DAI.

The LVVT in its current implementation employs signs in citation form. However, we envision that future versions of this system might also collate signs from continuous signing corpora (such as our NCSLGR corpus) where start/end annotations for individual signs are available. This extension could provide a seamless interface for viewing and synchronizing linguistic annotations across what are presently disparate datasets.

Finally, we are pursuing development of a lookup tool to facilitate access to multimedia materials such as ASL dictionaries. Modifications of interfaces we have developed for working with this kind of data (e.g., within the Java reimplementation of SignStream®, where tools are provided to facilitate intuitive data entry of phonological and morphological information) could also allow users to specify partial information about articulatory properties in order to improve upon results of computer-based search and retrieval.

## 8. Acknowledgments

We are grateful to Vassilis Athitsos, Joan Nash, Donna Riggle, Jessica Scott, Ben Bahan, Rachel Benedict, Naomi Berlove, Elizabeth Cassidy, Lana Cook, Jaimee DiMarco, Danny Ferro, Alix Kraminitz, Indya Oliver, Caelan Pacelli, Braden Painter, Chrisann Papera, Tyler Richard, Tory Sampson, Dana Schlang, Jon Suen, and Iryna Zhuravlova for invaluable assistance with this project. The work reported here has been supported in part by grants from the National Science Foundation (IIS-0705749, CNS-0855065, IIS-0964385).

## 9. References

Athitsos, Vassilis, Carol Neidle, Stan Sclaroff, Joan Nash, Alexandra Stefan, Ashwin Thangali, Haijing Wang & Quan Yuan. 2010. Large Lexicon Project: American Sign Language Video Corpus and Sign Language Indexing/Retrieval Algorithms. Paper presented to the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. LREC, Malta, May 22-23, 2010, 2010.

Bowden, Richard, David Windridge, Timor Kadir, Andrew Zisserman & Michael Brady. 2004. A Linguistic Feature Vector for the Visual Interpretation of Sign Language. *ECCV*.390-401.

Johnston, Trevor. 2010. From archive to corpus: Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics 14*.106-31.

Liwicki, S. & M. Everingham. 2009. Automatic recognition of fingerspelled words in British Sign Language. Paper presented to the Computer Vision and Pattern Recognition Workshops, 2009.

Neidle, Carol. 2002. *SignStream™ Annotation: Conventions used for the American Sign Language Linguistic Research Project.* American Sign Language Linguistic Research Project Report No. 11, Boston University.

—. 2007. *SignStream™ Annotation: Addendum to Conventions used for the American Sign Language Linguistic Research Project*. American Sign Language Linguistic Research Project Report No. 13, Boston University.

Neidle, Carol & Christian Vogler. 2012. A New Web Interface to Facilitate Access to Corpora: Development of the ASLLRP Data Access Interface (DAI) In *5th Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon, LREC 2012*. Istanbul, Turkey.

Thangali, Ashwin, Joan P. Nash, Stan Sclaroff & Carol Neidle. 2011. Exploiting Phonological Constraints for Handshape Inference in ASL Video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Valli, Clayton (ed.) 2002. *The Gallaudet Dictionary of American Sign Language*. Washington, DC: Gallaudet University Press.

Vicars, William G. 2012. ASLU: English to ASL Dictionary: http://www.lifeprint.com/dictionary.htm.

Vogler, Christian & Dimitris Metaxas. 2004. Handshapes and movements: Multiple-channel ASL recognition. *Springer Lecture Notes in Artificial Intelligence (Proceedings of the Gesture Workshop '03, Genova, Italy) 2915*.247-58.

# English-ASL Gloss Parallel Corpus 2012: ASLG-PC12

**Achraf Othman, Mohamed Jemni**
Research Laboratory LaTICE, University of Tunis
5, Av. Taha Hussein, B.P. 56, Bab Mnara, 1008 Tunis, Tunisia
E-mail: achraf.othman@ieee.org, mohamed.jemni@fst.rnu.tn

**Abstract**

A serious problem facing the community of researchers in the field of sign language is the absence of a large parallel corpus for signs language. The ASLG-PC12 project proposes a rule-based approach for building a big parallel corpus of English written texts and American Sign Language glosses. We present a novel algorithm that transforms an English part-of-speech sentence to an ASL gloss. This project was started in the beginning of 2011 as a part of the project WebSign, and it offers today a corpus containing more than one hundred million pairs of sentences between English and ASL glosses. It is available online for free to promote development and design of new algorithms and theories for American Sign Language processing, for example statistical machine translation and related fields. In this paper, we present tasks for generating ASL sentences from the Gutenberg Project corpus that contains only English written texts.

**Keywords:** American Sign Language, Parallel Corpora, Sign Language

## 1. Introduction

To develop an automatic translator or any other tool that requires a learning task for Sign Languages, the major problem is the collection of parallel data between text and Sign Language. A parallel corpus contains large and structured texts aligned between source and target languages. They are used to do statistical analysis and hypothesis testing, checking occurrences or validating linguistic rules on a specific universe. Since there is no standard and sufficient corpus for Sign Language (Morrissey & Way, 2007; Morrissey S. , 2008), to develop statistical machine translation that requires pre-treatment prior to the execution of the process of learning which needs an important volume of data.

For these reasons, we started to collect pairs of sentences between English and American Sign Language Gloss. And due to absence of data, especially in ASL and in other side there exists a huge data of English written text; we have developed a corpus based on a collaborative approach where experts can contribute in the collection and in correction of bilingual corpus and also in validation of the automatic translation. Experts are people that are authorized to validate translations and correct suggestions of translations. ASLG-PC12 project (Othman & Jemni, 2011) was started in 2010, as a part of the project WebSign (Jemni & El Ghoul, 2007) that carries on developing tools able to make information over the web accessible for deaf. The main goal of our project WebSign is to develop a Web-based interpreter of Sign Language (SL). This tool would enable people who do not know Sign Language to communicate with deaf individuals. Therefore, contribute in reducing the language barrier between deaf and hearing people. Our secondary objective is to distribute this tool on a non-profit basis to educators, students, users, and researchers, and to disseminate a call for contribution to support this project mainly in its exploitation step and to encourage its wide use by different communities.

In this paper, we review our experiences with constructing one such large annotated parallel corpus between English written text and American Sign Language Gloss –the ASLG-PC12 (Othman & Jemni, 2011), a corpus consisting of over one hundred million pairs of sentences.

The paper is organized as follow. Section 2 presents a brief description about American Sign Language Gloss. Section 3 presents methods and pre-processing tasks for collecting data from the Gutenberg Project (Lebert, 2008). We present two stages of pre-processing, in which each sentences had been extracted and tokenized. After, we present our method and algorithms for constructing the second part of the corpus in American Sign Language Gloss. Constructed texts were generated automatically by transformation rules and then corrected by human experts in ASL. We describe also the composition and the size of the corpus. Discussions and conclusion are drawn in section 5.

## 2. Background

Several projects, concerned with Sign Language, recorded or annotated their own corpora, but only few of them are suitable for automatic Sign Language translation due to the number of available data for learning and processing. The European Cultural Heritage Online organization (ECHO) published corpora for British Sign Language (Woll, Sutton-Spence, & Waters, 2004), Swedish Sign Language (Bergman & Mesch, 2004) and the Sign Language of the Netherlands (Crasborn, Kooij, Nonhebel, & Emmerik, 2004). All of the corpora include several stories signed by a single signer. The American Sign Language Linguistic Research group at Boston University published a corpus in American Sign Language (Athitsos, et al., 2010). TV broadcast news for the hearing impaired are another source of sign language recordings. Aachen University published a German Sign Language Corpus of the Domain Weather Report (Bungeroth, Stein, Dreuw, Zahedi, & Ney, 2006). In 2010, Sara et al., (Morrissey, Somers, Smith, Gilchrist, & Dandapat, 2010) published a multimedia corpus in Sign Language for machine Translation. In literature, we found many related projects aiming to build corpus for Sign Language. Most of them are based on video recording and we cannot find textual data toward building translation memory. Textual data for Sign Language is not a simple written form, because signs can contain others information line eye gaze or facial expressions. So, for our corpus, we will use glosses to represent Sign Language. In the next section, we will present a brief description about glosses.

## 3. Glossing signs

Stokoe (Stokoe, 1960) proposed the first annotation system for describing Sign Language. Before, signs were thought of as unanalyzed wholes, with no internal structure. The Stokoe notation system is used for writing American Sign Language using graphical symbols. After, others notation systems appeared like HamNoSys (Prillwitz & Zienert, 1990) and SignWriting (Sutton & Gleaves, 1995). Furthermore, Glosses are used to write signs in textual form. Glossing means choosing an appropriate English word for signs in order to write them down. It is not a translating, but, it is similar to translating. A gloss of a signed story can be a series of English words, written in small capital letters that correspond to the signs in ASL story. Some basic conventions used for glossing are as follows:

- Signs are represented with small capital letters in English.
- Lexicalized finger-spelled words are written in small capital letters and preceded by the '#' symbol.
- Full finger-spelling is represented by dashes between small capital letters (for example, A-C-H-R-A-F).
- Non-manual signals and eye-gaze are represented on a line above the sign glosses.

In this work, we use glosses to represent Sign Language. In the next section, we will describe steps for building our corpus.

## 4. English-ASL Parallel Corpus

### 3.1 Problematic issues

As we say in the beginning, the main problem to process American Sign Language for statistical analysis like statistical machine translation is the absence of data (corpora or corpus), especially in Gloss format. By convention, the meaning of a sign is written correspondence to the language talking to avoid the complexity of understanding. For example, the phrase "Do you like learning sign language?" is glossed as "LEARN SIGN YOU LIKE?". Here, the word "you" is replaced by the gloss "YOU" and the word "learn-ing" is rated "LEARN". Our machine translate must generate, after learning step, the sentence in gloss of an English input.

### 3.2 Ascertainment and approach

Generally, in research on statistical analysis of sign language, the corpus is annotated video sequences. In our case, we only need a bilingual corpus, the source language is English and the language is American Sign Language glosses transcribed. In this study, we started from 880 words (English and ASL glosses) coupled with transformation rules. From these rules, we generated a bilingual corpus containing 800 million words. In this corpus, it is not interested in semantics or types of verbs used in sign language verbs such as "agreement" or "non-agreement". Figure 1 shows an example of transformation between written English text and its generated sentence in ASL. The input is "What did Bobby buy yesterday?" and the target sentence is "BOBBY BUY WHAT YESTERDAY?". In this example, we save the word "YESTERDAY" and we can found in some reference "PAST" which indicates the past tense and the action was made in the past. Also, for the symbol "?" it can be replaced by a facial animation with "WHAT". For us, we are based on lemmatization of words. We keep the maximum of information in the sentence toward developing more approaches in these corpora. Statistics of corpora are shown in Table 1. The number of sentences and tokens is huge and building ASL corpus takes more than one week.

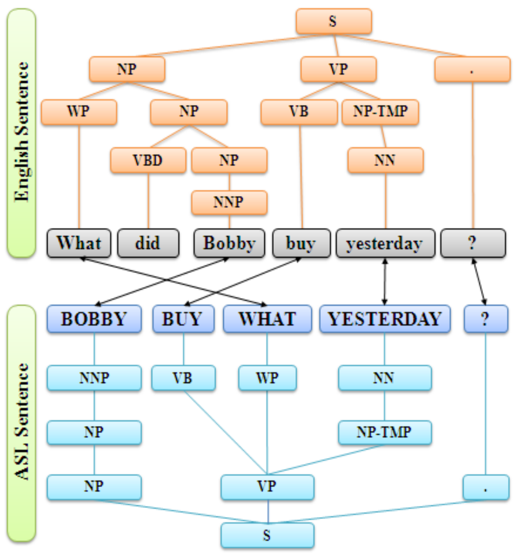



Figure 1: An example of transformation: English input 'What did Bobby buy yesterday?'

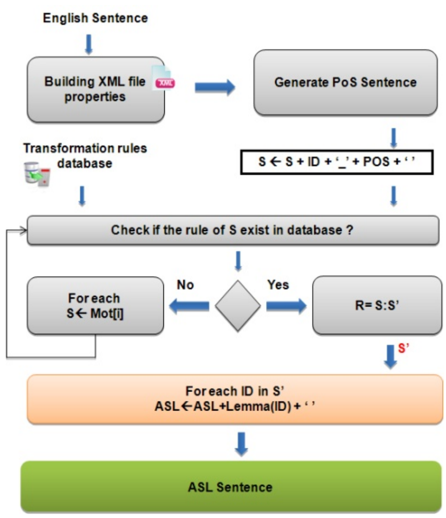



Figure 2: Steps for building ASL corpora

The input of the system is English sentences and the output is the ASL transcription in gloss. In table 2, only simple rules are shown, we can define complex rule starting from these simple rules. We can define a part-of-speech sentence for the two languages. According to figure 3, when we check if the rule of S exists in database, the algorithm will return true, in this case, we apply directly the transformation. Of course, all complex rules must be created by experts in ASL. Table 2 shows some transformation from English sentence to American Sign Language. We present the transformation rule made by an expert in linguistics.

| | Corpus size English | | Corpus size ASL Gloss | |
|---|---|---|---|---|
| | tokens | sentences | tokens | sentences |
| **PART 1** | 280 M | 13 M | 280 M | 13 M |
| **PART 2** | 323 M | 16 M | 323 M | 16 M |
| **PART 3** | 549 M | 27 M | 549 M | 27 M |
| **PART 4** | 292 M | 14 M | 292 M | 14 M |
| **PART 5** | 150 M | 7 M | 150 M | 7 M |

Table 1. Size of the American Sign Language Gloss Parallel Corpus 2012 (ASLG-PC12)

| |
|---|
| **English sentence:** what is your name?<br>**ASL sentence:** IX-PRO2 NAME, WHAT?<br>**Transformation rule**:<br>1_VBP 2_PRP 3_JJ 4_. → 2_PRP 0_DESC- 3_JJ 4_. |
| **English sentence:** Are you deaf?<br>**ASL sentence:** IX-PRO2 DESC-DEAF?<br>**Transformation rule**:<br>1_VBP 2_PRP 3_DT 4_NN 5_. → 4_NN 2_PRP 5_. |
| **English sentence:** are you a student?<br>**ASL sentence:** STUDENT IX-PRO2?<br>**Transformation rule**:<br>1_VBP 2_PRP 3_DT 4_NN 5_. → 4_NN 2_PRP 5_. |
| **English sentence:** do you understand him?<br>**ASL sentence:** IX-PRO2 UNDERSTAND IX-PRO3?<br>**Transformation rule**:<br>1_VB 2_PRP 3_VB 4_PRP → 2_PRP 3_VB 4_PRP |

Table 2. Example of full sentences transformation rules

In figure 2, we describe steps to transform an English sentence into American Sign Language gloss. The input of the system is the English sentence. Using CoreNLP tool, we generate an XML file containing morphological information about the sentence after tokenization task. Then, we build the part-of-speech sentence and thanks to the transformation rules database, we try to transform the input for each lemma. In some case, we can found that the part-of-speech sentence doesn't exist in the data-base, so, we transform each lemma. Transformation rule for lemma is presented in table 3. In the last step, we add an uppercase script to transform the output. The transformation rule is not a direct transformation for each lemma, it can an alignment of words and can ignore some English words like (the, in, a, an, etc.).

### 3.3 Transformations rules

Not all transformation rules used to transform English data were verified by experts in linguistics. We validate only 800 rules and transformation rules for lemma. We cannot validate all rules because there exist an infinite number of rules. For this reason, we developed an application that offer to experts to enter their rules from an English sentence, without coding. The application is just a simple user interface that contains lemma transformation rule, and the expert will compose lemma. After that, he save the result and rebuild the corpora. The built corpus is a made by a collaborative approach and validated by experts.

### 3.4 Collecting data from Gutenberg

Acquisition of a parallel corpus for the use in a statistical analysis typically takes several pre-processing steps. In our case, there isn't enough data between English texts and American Sign Language. We start collecting only English data from Gutenberg Project toward transform it to ASL gloss. Gutenberg Project (Lebert, 2008) offers over 38K free ebooks and more than 100K ebook through their partners. Collecting task is made in five steps:

Obtain the raw data (by crawling all files in the FTP directory).

- Extract only English texts, because there exist ebook in others languages than English like German, Spanish. We found also files containing ADN sequences.
- Break the text into sentences (sentence splitting task).
- Prepare the corpora (normalization, tokenization).

In the following, we will describe in detail the pre-processing steps to clean collected data.

### 3.5 Sentence splitting, tokenization, chunking and parsing

Sentence splitting and tokenization require specialized tools for English texts. One problem of sentence splitting is the ambiguity of the period "." as either an end of sentence marker, or as a marker for an abbreviation. For English, we semi-automatically created a list of known abbreviations that are typically followed by a period. Issues with tokenization include the English merging of words such as in "can't" (which we transform to "can not"), or the separation of possessive markers ("the man's" becomes "the man 's"). We use also an available tool for splitting called Splitta (Gillick, 2009). The models are trained from Wall Street Journal news combined with the Brown Corpus which is intended to be widely representative of written English. Error rates on test news data are near 0.25%. Also, we use CoreNLP tool (Toutanova & Manning, 2000; Klein & Manning, 2003). It is a set of natural language analysis tools which can take raw English language text input and give the base forms of words, their parts of speech.

### 3.6 Releases of the English-ASL Corpus

The initial release of this corpus consisted of data up to September 2011. The second release added data up to January 2012, increasing the size from just over 800 sentences to up to 800 million words in English. A forthcoming third release will include data up to early 2013 and will have better tokenization and more words in American Sign Language. For more details, please check the website (Othman & Jemni, 2011).

## 5. Discussions and conclusion

We described the construction of the English-American Sign Language corpus. We illustrate a novel method for transforming an English written text to American Sign Language gloss. This corpus will be useful for statistical analysis for ASL. We present the first corpus for ASL gloss that exceeds one hundred million of sentences available for all researches and linguistics. During the next phase of the ASLG-PC12 project, we expect to provide both a richer analysis of the existing corpus and others parallel corpus (like French Sign Language, Arabic Sign Language, etc.). This will be done by first enriching the rules through experts. Enrichment will be achieved by automatically transforming the current transformation rules database, and then validating the results by hand.

## 6. References


Athitsos, V., Neidle, C., Sclaroff, S., Nash, J., Stefan, A., Thangali, A., et al. (2010, May 22-23). Large Lexicon Project: American Sign Language Video Corpus and Sign Language Indexing/Retrieval Algorithms. " *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, LREC* .

Bergman, B., & Mesch, J. (2004). *ECHO data set for Swedish Sign Language (SSL).* Department of Linguistics, University of Stockholm.

Bungeroth, J., Stein, D., Dreuw, P., Zahedi, M., & Ney, H. (2006). A German Sign Language Corpus of the Domain. *Fifth International Conference on Language Resources and Evaluation*, (pp. 2000-2003). Genoa, Italy.

Crasborn, O., Kooij, E. v., Nonhebel, A., & Emmerik, W. (2004). *ECHO data set for Sign Language of the Netherlands (NGT).* Department of Linguistics, Radboud University Nijmegen.

Gillick, D. (2009). Sentence Boundary Detection and the Problem with the U.S. *Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* , pp. 241-244.

Jemni, M., & El Ghoul, O. (2007). An avatar based approach for automatic interpretation of text to Sign language. *9th European Conference for the Advancement of the Assistive Technologies in Europe.* San Sebastian.

Klein, D., & Manning, C. D. (2003). Accurate Unlexicalized Parsing. *Proceedings of the 41st Meeting of the Association for Computational Linguistics* , pp. 423-430.

Lebert, M. (2008). *Project Gutenberg (1971-2008).* University of Toronto & Project Gutenberg.

Morrissey, S. (2008). Assistive translation technology for deaf people: translating into and animating Irish sign language. *12th International Conference on Computers Helping People with Special Needs.* Linz.

Morrissey, S., & Way, A. (2007). Joining hands: developing a sign language machine translation system with and for the deaf community. *Conference and Workshop on Assistive Technologies for People with Vision and Hearing Impairments: Assistive Technology for All Ages.* Granada.

Morrissey, S., Somers, H., Smith, R., Gilchrist, S., & Dandapat, S. (2010, May). Building a Sign Language corpus for use in Machine Translation. *Proceedings of the 4th Workshop on Representation and Processing of Sign Languages: Corpora for Sign Language Technologies* , pp. 172-177.

Othman, A., & Jemni, M. (2011). *American Sign Language Gloss Parallel Corpus 2012 (ASLG-PC12).* Retrieved from http://www.achrafothman.net/aslsmt

Prillwitz, S., & Zienert, H. (1990). Hamburg notation system for sign language: Development of a sign writing with computer application. *nternational Studies on Sign Language and Communication of the Deaf* (pp. 355–379). Hamburg, Germany: Signum Press.

Stokoe, W. (1960). *Sign Language Structure: An Outline of the VisualCommunication Systems of the American Deaf.* Linstok Press, SilverSpring.

Sutton, V., & Gleaves, R. (1995). SignWriter— the world's first sign language processor. *Deaf Action Committee for SignWriting* .

Toutanova, K., & Manning, C. D. (2000). Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger. *the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora* , pp. 63-70.

Woll, B., Sutton-Spence, R., & Waters, D. (2004). *ECHO data set for British Sign Language (BSL).* Department of Language and Communication Science, City University (London).

# Sign Language Documentation in the Asia-Pacific Region: A Deaf-Centered Approach

**Felix Sze, James Woodward, Gladys Tang, Jafi Lee, Ka-yiu Cheng, Joe Mak**
Centre for Sign Linguistics and Deaf Studies, Department of Linguistics and Modern Languages,
The Chinese University of Hong Kong
Shatin, Hong Kong
E-mail: felix_cslds@cuhk.edu.hk, woodyvn@yahoo.com, gtang@cuhk.edu.hk, jafi_cslds@cuhk.edu.hk, yiu_cslds@cuhk.edu.hk, joemak_cslds@cuhk.edu.hk

**Abstract**

In this paper, we would like to share our experience in training up Deaf individuals from the Asian-Pacific countries to compile sign language dictionaries and conduct sign language research through the 'Asia-Pacific Sign Linguistics Research and Training Program' at the Chinese University of Hong Kong. The program, fully funded by the Nippon Foundation, is a multi-country, multi-phase project which aims at nurturing Deaf people to become sign language researchers through a series of credit-bearing training programs at the diploma and higher diploma levels. The training covers three major areas: Sign Linguistics, Sign Language Teaching and English Literacy. One important part of the training involves the production of sample dictionaries of the Deaf trainees' own sign languages. To confirm the dictionary entries, the Deaf trainees conduct surveys in the Deaf communities in their home countries from time to time and as a result a substantial amount of lexical variants have been collected. An online database, called the Asian SignBank, is now being developed to house these lexical data and facilitate further research. Apart from basic search functions, the SignBank also incorporates detailed phonetic features of individual signs and a materials-generating function which allows quicker production of dictionaries in the future.

**Keywords:** sign language documentation, sign linguistic training, sign language lexicography, the Asia-Pacific region

## 1. Introduction

In this paper, we would like to share our experience in training up Deaf individuals from countries in the Asian-Pacific region to compile sign language dictionaries and conduct sign language research through 'the Asia-Pacific Sign Linguistics Research and Training Program' (hereafter APSL Program) at the Chinese University of Hong Kong. Specifically, we will discuss the types of training given to these Deaf people, the field work methods that they use to collect lexical variation data during the compilation of sign language dictionaries, and the design of the Asian SignBank, an online platform being developed to house these lexical data.

## 2. The APSL Program

### 2.1 Founding Philosophy

Sign language research has flourished in an increasing number of countries over the past few decades, particularly in the States and Europe. Apart from its academic contributions, the accumulated research findings have brought forth remarkable breakthroughs in the recognition of the value of sign languages to Deaf communities and in the education systems of Deaf children. In the Asia-Pacific Region, however, misconceptions and prejudice against sign languages and Deaf people still abound due to a lack of research. To help improve the situation of the Deaf, there is a pressing need to document sign languages in the region and develop sign language resources that can support future developments. Upon seeing this need, the Nippon Foundation offered a major donation to the Chinese University of Hong Kong to set up the APSL Program in 2003. This program is a multi-country, multi-phase project that aims at documenting sign languages and empowering the Deaf communities in the Asia-Pacific region by providing sign linguistic training to Deaf individuals.

### 2.2 Training for the Deaf researchers

During the first phase of the program (2003-2007), a total of 270 hours of on-site linguistic training was provided to deaf fluent signers in Hong Kong (4 trainees), the Philippines (6 trainees), Cambodia (6 trainees) and Vietnam (8 trainees). The training scheme included introductory courses on the formational structure of sign languages, grammatical structure of sign languages, lexical structure of sign languages, sociolinguistics, lexicographical study of sign languages, and applied sign language linguistics. One important component of this phase was the production of practical dictionaries and sign language teaching materials, which were deemed indispensable for the promotion of sign languages in the local communities.

In the second phase (2006 – 2012), the APSL Program was further expanded to support training to both Deaf and hearing individuals with the long-term goal of establishing sign linguistics research units at the university level in the participating countries. [1] Collaboration was sought with the Regional Secretariat for Asia and the Pacific of the World Federation of the Deaf (WFD RSA/P) in identifying the participating countries and contacting local Deaf organizations for

[1] Negotiation with the Nippon Foundation about further extension of the APSL Program is now underway.

recruiting potential Deaf trainees. Selected Deaf adults, who are fluent signers of their own sign languages, were brought to Hong Kong to receive centralized research training at the Chinese University of Hong Kong. The first cohort, consisting of five deaf adults from Sri Lanka, 4 from Indonesia and 2 from Hong Kong, commenced their training in November 2007. Recently, in November 2010, 5 Japanese, 2 Fijian and 2 Hong Kong Deaf adults were admitted as the second cohort. Most of these Deaf trainees had only completed 10th/11th grade without opportunities of further education in their home countries.

The training in Phase II covers three major areas: sign linguistics, sign language teaching and English literacy. Five diploma programs and one higher diploma program (1,350 hours of training in total) are jointly offered by the Centre for Sign Linguistics and Deaf Studies and the School of Continuing and Professional Studies of the Chinese University of Hong Kong. They are:

- Diploma Program in Basic Sign Language Lexicography for the Deaf;
- Diploma Program in English Literacy and IT Application for the Deaf;
- Diploma Program in Sign Language Studies for the Deaf;
- Diploma Program in General Studies for the Deaf;
- Diploma Program in English Literacy Skills for the Deaf; and
- Higher Diploma in Sign Linguistics and Sign Language Teaching,

The sign linguistic component covers areas such as sign language lexicography, phonology, morphology, syntax, sociolinguistics, sign language research projects, language acquisition, etc., which equip the Deaf trainees with basic knowledge and skills for compiling dictionaries and documenting their own sign languages. The purpose of the English component is to ensure that the trainees develop sufficient reading and writing skills for accessing information and conducting research independently. The sign language teaching modules include teaching methodology, materials development, syllabus design, language assessment and practicum. It is hoped that these trainees will become competent sign language teachers and take the lead of promoting sign language in their home countries in the future.

### 2.3 Training for the Hearing researchers and inter-university collaboration

While grooming deaf sign language researchers is essential to Deaf empowerment in the long run, we also see the importance of nurturing hearing researchers to work as collaborators with Deaf people in promoting the study of sign languages. Hence, efforts have been made to set up inter-university links with the participating countries for recruiting committed hearing students to come to CUHK to receive sign linguistic training at the master level. So far, the University of Indonesia (Indonesia), the University of Kelaniya (Sri Lanka), the University of the South Pacific (Fiji) and the University of Tokyo (Japan) have agreed to be our collaborators. For each country, a maximum of two hearing students with good signing skills will be recruited. At present, two hearing Indonesian students are receiving their training in Hong Kong. More are expected to come in the near future.

## 3. Documentation of Sign Languages

### 3.1 Compiling dictionaries and teaching materials

Besides studying, the Deaf trainees of the APSL Program (Phase II) are involved in the production of sample dictionaries and teaching materials of their own sign languages. The dictionaries basically adopt parameters in sign language phonology as the principles for entry ordering. Signs are first broadly classified in terms of handshapes and ordered accordingly. Within each handshape, the signs are further ordered according to the number of hands, palm orientation, etc. For the entries, both line drawings/still photos and video clips are produced. The line drawings/still photos are used for producing printed dictionaries, whereas the video clips will be used for producing electronic dictionaries or be placed online for public access in the future. We decided to provide training on different ways to produce dictionaries because the format of the dictionaries Deaf trainees will make in the future depends on the socio-economic situation of their own communities. For developed countries where computers are commonplace, electronic dictionary is the way to go; printed dictionaries with line-drawings will be suitable for developing countries where computers are still rare and printing cost is high for full images. So far, the Sri Lankan team has finished the drafts of Book 1 and Book 2 of the sample dictionaries, each consisting of around 250 sign entries. The Indonesian team is working on the dictionaries of two signing varieties, one in Jakarta and one in Yogyakarta. For both varieties, the drafts of Book 1 and 2 are also completed. Both the Sri Lankan and Indonesian students are preparing the manuscripts of the teaching materials. As for the Japanese and Fijian trainees, they just started the compilation recently, and hopefully the draft of the first dictionary booklet will be ready by the end of 2012.

### 3.2 Field surveys, lexical variation and other sign language data

Note that in the process of dictionary compilation, the Deaf trainees are required to conduct surveys in their home countries once or twice a year to verify the sign entries and look for possible lexical variants. What they usually do is prepare photos/movie clips for eliciting the target signs, and look for fluent deaf signers in their countries as informants. The lexical variants, when found, are videotaped. These surveys provide useful information on the lexical variations across signers in the Deaf community. So far, over 3000 lexical variants have been collected from Sri Lankan Sign Language and the two signing varieties in Indonesia respectively.

In the fieldwork survey, the trainees also collected some other signing data such as picture descriptions, narratives and conversations from the signing informants. These data can be used for further linguistic research or as references for making teaching materials. In fact, these lexical variations and discourse data are now being analyzed in the following small-scale research projects by the Deaf/hearing trainees or staff members of the APSL Program:

- Comparing Jakarta Sign Language and Yogyakarta Sign Language
- Sign language use and lexical variations in Jakarta Sign Language
- A comparison of word order in Hong Kong Sign Language, Sri Lankan Sign Language, and Jakarta Sign Language

## 4. Asian SignBank

### 4.1 Design of the infra-structure

The Asian SignBank is an online database developed to facilitate componential analysis and storage of sign entries collected through the APSL Program. It is capable of storing a wide range of linguistic information of a sign, including:

- glosses of signs in their native spoken language and English, to facilitate both local and international access of information;
- individual glosses for each component in compound signs;
- examples which make use of that sign;
- related signs which show variation of the signs in the Deaf community, and
- detailed phonetic information of the sign.

The phonetic notional system of the SignBank is based on the feature analysis proposed by Brentari's Prosodic Model (1998), as shown in Figure 1:

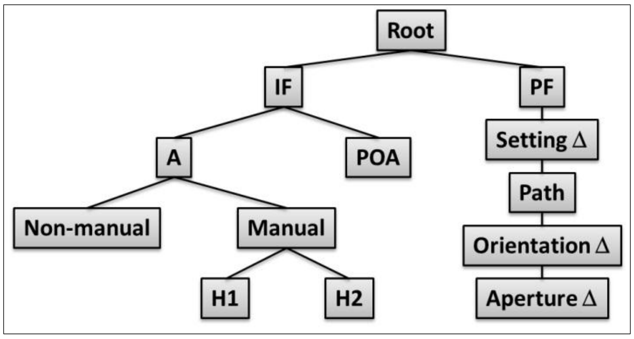


Figure 1. Brentari's Prosodic Model (1998)

For each sign, the phonetic information that can be listed in the system include: inherent features (i.e. static features of a sign, like the handshape, place of articulation, body part, palm orientation, handparts, etc.) and prosodic features (i.e. dynamic features of a sign, like path and shape, setting, orientation, aperture and handshape changes, trilled movements, etc.) (See Figure 2).

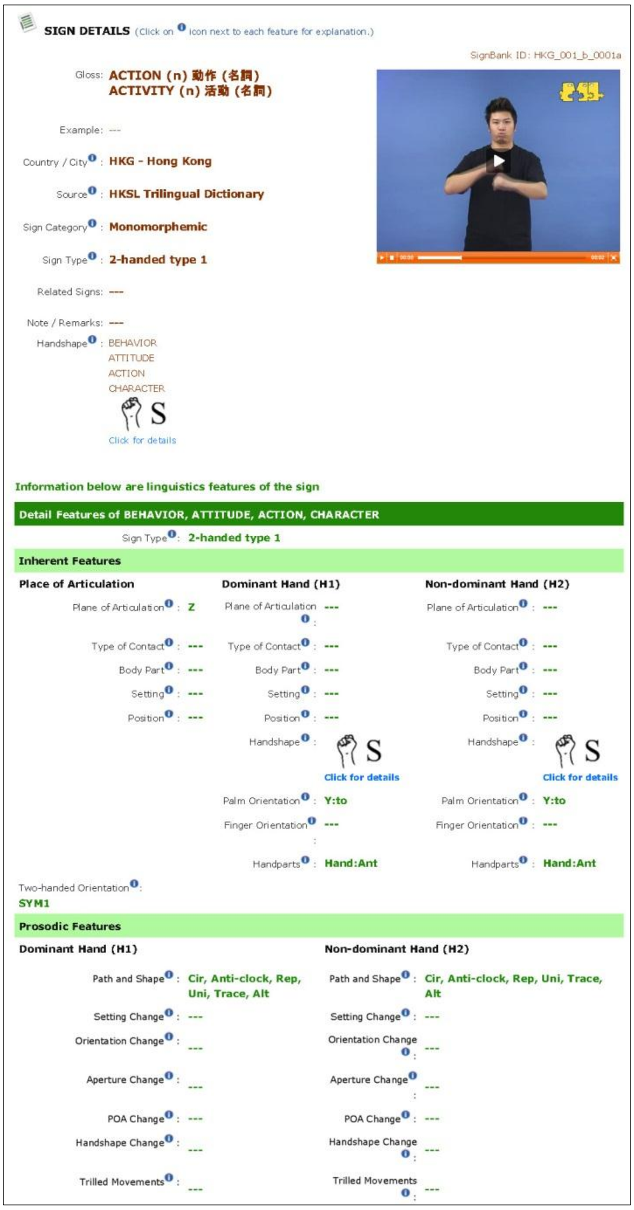


Figure 2: A result page from Asian SignBank showing detailed linguistic features of a sign.

### 4.2 Benefits to sign language research

In addition to providing a storage and viewing platform for the linguistic properties of signs, the Asian SignBank also allows a searching strategy of the signs according to any piece of information one may know about them, from the glosses to the most detailed phonetic information. Searching can be done by a single feature or by a conditioned combination of features. For instance, in Figure 3, a user can search for a sign with a handshape plus location feature (e.g. 5-handshape + articulated on the shoulder). Also, users can look for the definition and pre-defined values of each linguistic feature (Figure 4).

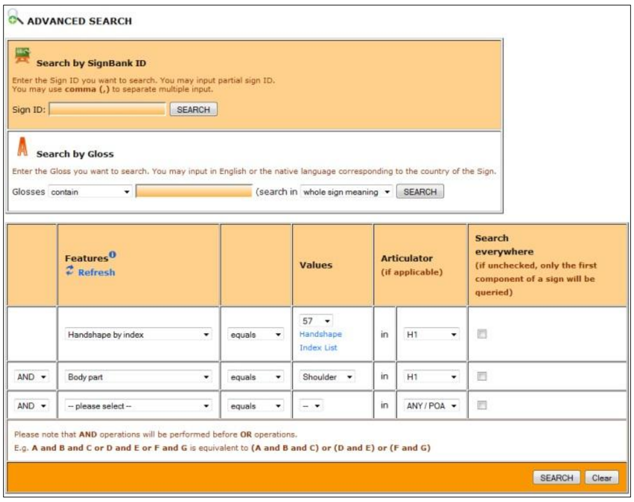



Figure 3. A search page of the Asian SignBank allowing users to query for a sign with different combinations of linguistic features.

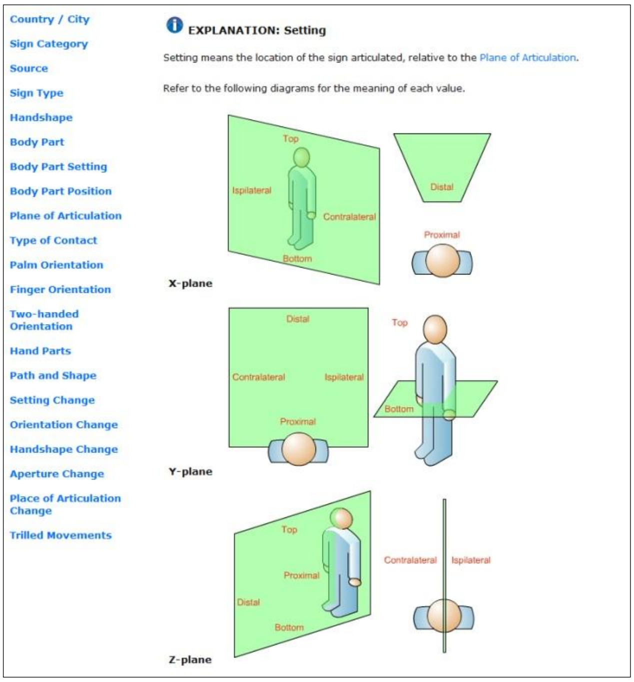



Figure 4. An explanation page describing the definition of each linguistics feature and its corresponding pre-defined values.

On top of serving a sign retrieval purpose, the search functions can serve a descriptive statistical purpose such as frequency of occurrence or co-occurrence of particular features of signs. When the system extends to include other sign languages in Asia, the system will allow a cross-linguistic comparison of signs and their linguistic information. So far the data in the Asian SignBank has already supported the completion of an M.Phil thesis on the movement properties of Hong Kong Sign Language (Mak 2011).

### 4.3 Sign language material-generating function

Recently, a materials-generating function was also added to the system to allow researchers to generate search outputs in formats that (i) facilitate qualitative and quantitative analysis; (ii) allow sign language dictionary and glossary production in customized formats; and (iii) allow sign language materials production that suits different teaching purposes.

For instance, a sign language teacher may want to get a list of all S-handshaped signs for his class, or a sign linguist may want to retrieve from the Asian SignBank with all the signs from different countries that use ‘head’ as the body part. Without the need to install any specific software, users can get the resulting documents in form of Excel spreadsheet for feature analysis, or in PDF format with customizable layout as teaching materials or dictionary.

### 4.4 Current stage of development

At present, the infrastructure of the Asian SignBank has been established and tested with around 2000 Hong Kong signs, 780 Vietnamese (Ho Chi Ming City) signs, 150 Sri Lankan signs, 340 Jakarta signs and 270 Yogyakarta signs. When the system is mature, more signs from these sign languages will be input and analyzed. The next target sign languages for the Asian SignBank will be from the sign language varieties of Japan and Fiji.

## 5. Conclusion

In this paper we’ve introduced how Deaf adults from the Asia-Pacific region are trained to compile sign language dictionaries, produce sign language teaching materials and become sign language researchers/teachers through the APSL Program at the Chinese University of Hong Kong. We’ve also discussed the design of the Asian SignBank, an online database developed to house the lexical data of different sign languages in a way that can facilitate linguistic analysis as well as production of sign language materials for research or pedagogical purposes. Central to the whole APSL Program is our conviction that Deaf people need to in the centre of sign language documentation and sign language research, which is deemed essential for the empowerment and betterment of Deaf people in the region in the long run.

## 6. Acknowledgements

We are heavily indebted to the Nippon Foundation for their trust in us and their generous financial support for the APSL Program. In addition, we would like to thank WFD/RSAP, the local deaf organizations, and the four partner universities for their collaboration and support.

## 7. References

Brentari, D. (1998). *A Prosodic Model of Sign Language Phonology.* Cambridge, MA: MIT Press.

Mak, K.L. (2011). Movement classes and feature organization in Hong Kong Sign Language. Unpublished M.Phil thesis, The Chinese University of Hong Kong.

# Compiling the Slovene Sign Language Corpus

**Špela Vintar, Boštjan Jerko, Marjetka Kulovec**
University of Ljubljana, Dept. of Translation Studies
Aškerčeva 2, SI - 1000 Ljubljana
E-mail: {spela.vintar, bostjan.jerko, marjetka.kulovec}@ff.uni-lj.si

**Abstract**

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

**Keywords:** Slovene Sign Languge, corpus compilation, sign language transcription

## 1. Introduction

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

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

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

## 2. Slovene Sign Language - background

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

Just like elsewhere in the world, being deaf does not equal sign language user and vice versa; many deaf people have learned to communicate primarily through lip reading and speaking, and many sign language users are hearing children of deaf parents or simply hearing users of SSL, for whatever reason.

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

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

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

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

[1] http://www.zveza-gns.si/

## 3. Data collection

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

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

The personal information we are collecting from the signers include:

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

### 2.1 Recording sessions

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

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

### 2.2 Field observations

The organisaton of recording sessions is performed with the help of local deaf clubs. At the beginning of the project we organised a presentation event on the premises of the association of deaf clubs, where the goals of the project and the plans for data collection were presented to local presidents. The responses of the deaf community were cautious. It became clear from the questions and comments that some fears were related to the fact that this was the first time an academic institution launched a project on the topic of SSL, and that the deaf community felt this as an unwelcome intervention or an attempt to "prescribe" or "forbid" certain SSL usage. Having made clear that the aim of the project was primarily to describe the language as it is currently used, there was again some disappointment due to the fact that certain SSL users effectively wish for a certain level of standardisation to occur, for purely practical reasons.

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

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

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

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

## 4. Corpus annotation and processing

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

The first annotation stages include segmentation

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

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

Similar to other sign languages, Slovene sign language uses modifications of signs to express syntactic and semantic relations between items in a sentence. Thus, the base form of the sign TEACH can be modified in various ways to signify *teacher* (male or female), *I teach* (1st person, active voice), *I am being taught* (1st person, passive voice) and so on. Transcribing such tokens involves deciding whether *teach* and *teacher* are to be considered forms of a single lexeme or two separate lexemes, whether the sign for *female teacher* should be tokenized as one sign or two and other similar dilemmas.

SSL has a relatively poor specialized vocabulary, therefore many specific topics need to be signed "creatively", using general signs in new contexts and with new meanings. The mouthing accompanying these creative uses is necessary to infer the intended meaning, but in transcribing such signs we again need to decide whether this creative use constitutes a new lexeme or not.

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

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

## 5. Conclusions and future work

Since we are describing a relatively young and small-scale project, there are currently few conclusions and substantial future work. First of all we plan to finish collecting the materials and in particular proceed with the transcription, as this activity alone generates fundamental theoretical questions. Transcribing will be performed with iLex primarily by the project members, and in certain stages the deaf students will also be involved.

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

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

.

## 6. References


Dotter, F. (2011) Sign Languages and Their Communities Now and in the Future. "Multilingualism in Europe. Prospects and Practices in East-Central Europe", Budapest, March, 25-26, 2011

Hanke, T. and Stolz, J. (2008). iLex – A Database Tool For Integrating Sign Language Corpus Linguistics and Sign Language Lexicography, LREC 2008, May 28 – May 30 2008, poster

Johnston, T. and A. Schembri (2006). “Issues in the creation of a digital archive of a signed language”. In L. Barwick and N. Thiesberger (eds.) *Sustainable data from digital fieldwork*. Sydney: Sydney University Press, 7-16.

König, S., R. Konrad and G. Langer. (2008). “What’s in a sign? Theoretical lessons from practical sign language lexicography”. In J. Quer (ed.) *Signs of the time. Selected papers from TISLR 2004*. Hamburg: Signum, 379-404.

Konrad, R., A. Schwarz, S. König, G. Langer, T. Hanke and S. Prillwitz (2003). *Fachgebärdenlexikon Sozialarbeit/Sozialpädagogik*. Hamburg: Signum. Available at: http://www.sign-lang.uni-hamburg.de/slex/ (accessed: 30 September 2009).

Krammer, K., Bergmeister, E., Dotter, F., Hilzensauer, M., Okorn, I., Orter, R. and Skant, A. (2001). The Klagenfurt database for sign language lexicons. Sign Language and Linguistics, vol. 4, issue 1, pp. 191-201

Prinetto, P. et al. (2011). The Italian Sign Language Sign Bank: Using WordNet for Sign Language corpus creation, ICCIT, 2011 Internation Conference , 29-31 March 2011, pp. 134-137

Schmaling, C. and T. Hanke (2001). *HamNoSys 4.0*. Available at: http://www.sign-lang.uni-hamburg.de/Projekte/HamNoSys/HNS4.0/englisch/HNS4.pdf (accessed: 30 September 2011).

Tarczay, S. (2010.), Pretpostavke profesionalizacije prevoditelja znakovnoga jezika za gluhe i gluhoslijepe osobe. Magistarski rad. Zagreb: Edukacijsko-rehabilitacijski fakultet Sveučilišta u Zagrebu

# A Proposal for Making Corpora More Accessible for Synthesis: A Case Study Involving Pointing and Agreement Verbs

**Rosalee Wolfe[1], John C. McDonald[1], Jorge Toro[2], Jerry Schnepp[1]**
[1]DePaul University Chicago, IL
[2]Worchester Polytechnic Institute Worcester, MA
E-mail: {wolfe,jmcdonald,jschnepp}@cs.depaul.edu, jatoro@wpi.edu

**Abstract**

Sign language corpora serve many purposes, including linguistic analysis, curation of endangered languages, and evaluation of linguistic theories. They also have the potential to serve as an invaluable resource for improving sign language synthesis. Making corpora more accessible for synthesis requires geometric as well as linguistic data. We explore alternate approaches and analyze the tradeoffs for the case of synthesizing indexing and agreement verbs. We conclude with a series of questions exploring the feasibility of utilizing corpora for synthesis.

**Keywords:** annotation standards, corpora, sign language synthesis

## 1. Introduction

Sign language corpora provide a means for developing new insights into sign language (Crasborn, 2008), for supporting the documentation and curation of endangered languages (Johnston & Schembri, 2006), and for enabling alternative methods for evaluating theories, such as those describing patterns in language acquisition (Lillo-Martin & Pichler, 2008). By design, corpora are meant to support future as well as current research.

Sign language corpora could also be a valuable resource for the development of better sign language synthesizers. A synthesizer is an essential component of an automatic translation system between spoken and signed language. It can also serve as a verification tool for transcribing lexical items and can serve as a powerful basis for building flexible educational tools.

Current sign synthesizers excel at recalling items from a lexicon and concatenating them to create sentences. However, much work still needs to be done to expand the flexibility of synthesizers if they are to fulfill their promise, and corpora have the potential of serving a key role in this development.

## 2. Using Corpora for Improving Synthesis

The output of a synthesizer is only as good as the data used to create it. Without access to corpora, researchers miss important cases that synthesis algorithms need to model. New models must be rigorously tested with as many examples as possible. Access to corpora opens the door for thorough testing.

Corpora gathered for analysis provide large, rich collections of exemplars which are useful for algorithm development. They have three advantages over those gathered by synthesis researchers. The first is the level of quality of the recorded data, the second is the general purpose of the recorded data, and the third is the annotations accompanying the recorded data.

### 2.1 High-Quality Recording

Through years of experience, linguists have developed consistent methodologies for elicitation, and have established state-of-the-art recording facilities, designed specifically for capturing sign language. The results are high-quality recordings that preserve as much information as possible.

### 2.2 Generality

The second advantage of corpora gathered for analysis is the general nature of the data. We have found that our own elicitation techniques can become too specific when we are interested in representing a particular language construct for synthesis. As with movie directors, there the overwhelming desire to give such directions such as "now point to the red square." For example, when informants knew we were interested in the placement of indices, it overly influenced how the informants signed the story.

### 2.3 Annotations

If sign language corpora were simply a collection of recordings, their usefulness for synthesis would be limited due to the time investment required to manually search the videos for the desired exemplars. The addition of annotations facilitates time-effective machine searching. Searchable annotations also provide the potential to identify exceptional cases that do not fit standard models. Synthesis algorithms need to incorporate these in order to exhibit the full range of expressiveness of natural signing. While annotation data desired by synthesis researchers and linguistic scholars share many similarities, they differ somewhat in several key areas. To better understand these similarities and differences, the following section describes the organization of our sign synthesis system and lays the groundwork for a possible approach to utilize corpora originally intended for analysis.

## 3. Motivation for treatment of numeric and linguistic data

Ultimately any sign synthesis system must have access to numeric data for creating the postures and timing of animation. Our sign synthesis system combines rules, linguistic labels and numeric data. It has four major components – a handshape editor, a sign transcriber, an expression builder and a sentence generator. The first three provide user interfaces to record and store numeric, phonemic and lexical data, as shown in Figure 1. The fourth combines these data to form complete sentences.

Our earliest component was the handshape editor. It stores the information not in terms of joint rotations but as the handshape features of bend, spread and hook. A mathematical model (McDonald, et al., 2001) converts linguistic features to joint rotations, generating the numeric data required by the underlying animation engine.

The sign transcriber uses handshapes as a basis for creating signs. It then allows for the designation of an articulator, place of articulation and palm orientation. These correspond to the phonemic parameters of handshape, location and palm orientation.

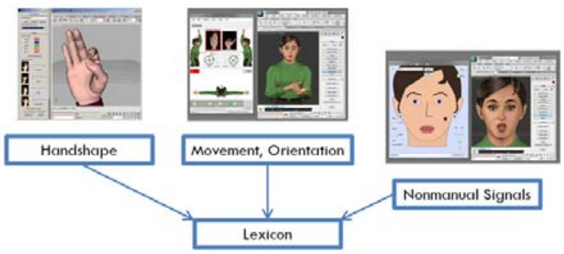



Figure 1: Handshape editor, sign transcriber, expression builder

Initial testing with members of the Deaf community indicated that more flexibility should be incorporated into this approach. Reviewers indicated that signs were awkward, and would demonstrate that sometimes a different location may be preferable to that used in the synthesized animations. As seen in Figure 2, we added a method to control positioning at a very fine level of detail. Although it still carries the linguistic tag "Left Temple", the actual geometric position of the location has changed.

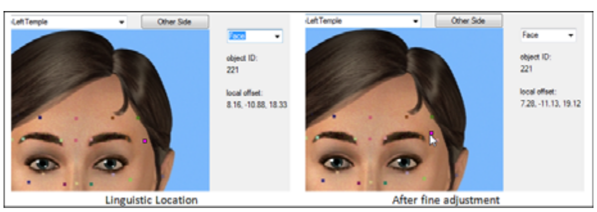



Figure 2: Fine adjustment to phonemic parameter of location

The linguistic parameter of motion caused the most difficulties. We found that the rates of change are not the same for all parameters as shown in Figure 3. In this example of the sign INFORM (National Technical Institute for the Deaf, 2000), the initial handshape (label A) transitions to the final handshape (label B) in half the time required to transition from the initial location (label A) to the final location (label C). For this reason, the sign transcriber includes facilities to designate internal timing within a sign as seen in Figure 4.

Some lexical signs require the inclusion of a facial nonmanual signal. To address this requirement, the expression builder provides access to facial elements used in the formation of nonmanual signals (Schnepp, Wolfe, & McDonald, 2010).

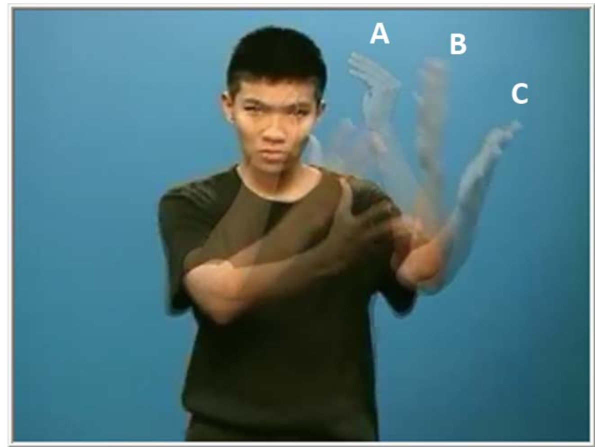



Figure 3: Varying rates of change: Handshape transition is complete before final location is achieved.


Figure 4: Timing interface

Additionally, the sign transcriber requires information about a sign's part of speech (POS). The data collected depends on the POS category. For example, agreement verbs require information about the type of agreement (object only, both subject and object), direction (backwards or forwards), and orientation agreement (Toro, 2004). The citation form is stored as data and the conjugated form is created dynamically when sentences are synthesized.

Finally the sentence generator uses a stream of text tokens as input to combine lexical items and grammar rules to generate complete sentences. The tokens can be glosses, fingerspelled words or indices. The sentence generator looks up the sign stem in the lexicon. Depending on the POS, rules modify the sign stem and may require additional information. For example, if the sentence includes an agreement verb, the user needs to specify the subject and object by designating the relevant indices. To synthesize the utterance, the sentence generator applies its grammar rules to modify the animation data, and renders the animation. Figure 5 shows the flow of data through the entire system. The signs, handshapes and nonmanual signals are represented as data, while the sentence generator is rule-based.

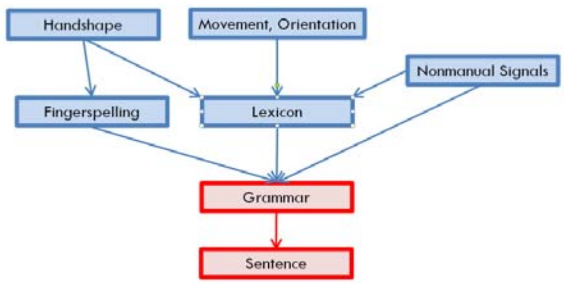



Figure 5: Flow of data towards the sentence generator. Blue indicates a data-based representation; red indicates a rule-based representation.

# 4. Synthesis methodology

Introducing a new language construct into this system is a five-step process:

1. Research the linguistic literature for descriptions and characterizations of the construct under study. Linguistic theory provides the guidance and inspiration for the algorithmic representation.
2. Observe examples of the construct in context. Study multiple signers in multiple contexts. This is similar to the motion studies animators create when planning an animation.
3. Use or modify the software to implement the construct. Synthesize signed sentences.
4. Conduct user tests with representatives of the Deaf community to gather feedback on questions such as
   a. Is the sentence comprehensible?
   b. Does the avatar sign it the way you would sign it?
   c. If it's not right, would you show us how it should be signed?
5. Analyze the feedback, propose refinements, and repeat the process.

It is at step two where corpora would be most useful. The next section discusses how different tiers in a corpus can benefit the synthesis process.

# 5. Corpora from a synthesis standpoint

Currently, we are studying processes involving indices and agreement verbs, and have been looking at the types of tiers that could support effective synthesis of sentences with these processes. Having a gloss tier is essential and ID-glosses are optimal for searching our lexicon. The start and end times for a sign are also critically useful, because they provide the average and range for a sign's duration. However, a gloss tier does not carry enough information to correctly surmise agreement verb conjugation.

Most corpora include more than a gloss tier. The following paragraphs analyze tier types and combination of tier types with respect to supporting synthesis.

## 5.1 Gloss and phonemic tiers

This approach focuses on descriptive annotation; where phonemic information is labeled, but syntactic designations are omitted to be as theory-neutral as possible. This set of tiers is useful for supporting verb conjugation because it contains specific information about the starting and ending location of the verb form. However, it is difficult to infer the identity of the referents from these locations. Per Padden (1990), the locus for a referent is not a precise geometric position. Further different verbs (SHOW vs. TELL) will assume different geometric positions while still indicating the same referent. Additionally, this approach requires a direct geometric interpretation of phonemes, which does not facilitate any fine-tuning required for naturally-flowing synthesis.

## 5.2 Gloss and syntactic tiers

In this approach, corpora contain not only glosses, but labels for POS and referents for agreement verb conjugation. A synthesizer can utilize the syntactic information to apply its rules for modifying signs. With this approach, the synthesizer makes some assumptions, placing the referents at "best guess" locations, and adjusting the verbs and nonmanual signals accordingly. Unfortunately, the synthesizer may not always make good guesses, particularly when there are more than two referents, resulting in awkward sentences.

## 5.3 Gloss, syntactic and phonemic tiers

Having access to both syntactic information about a referent as well as the phonemic information pertaining to its location gives a synthesizer everything it needs to create well-formed grammatical sentences that flow naturally. However, the prospect of tagging for syntax (which might need to be revised) and recording the detail of phonemic data is a nontrivial challenge.

## 5.4 Gloss, syntactic, and selected phonemic tiers

One possibility might be to record syntactic tags, and only a small subset of phonemic information. A synthesizer needs to know the location of a referent when it is established in the sign space, so the referent only needs to be tagged for location once in the annotation. According to Padden (1986), a location remains associated with a referent during discourse until the signer explicitly associates a new referent with the location. Since the only location data required is the first appearance of a referent, a corpus that already includes syntactic tagging would require minimal additional phonemic information.

# 6. Benefits of Standardized Tiers for Synthesis

Having a standardized set of tiers for synthesis would add flexibility. It facilitates the possibility of interchanging signing avatars or animation software and provides a test-bed for different approaches to synthesis such as mocap, procedural or manual animation. It also leaves open the possibility for changing avatars to accommodate different audiences (adults vs. children, addressing cultural sensitivities) or applications (real-time vs. higher fidelity rendering).

Maintaining the separation between the phonemic and syntactic representations of sign language makes it possible to create and modify movement algorithms for sign production without requiring re-annotation. Results of lexicographic research from projects such as iLex (Hanke, Storz, & Wagner, 2010) could be used to improve models of movement, resulting in more natural and believable sign synthesis. This approach could potentially accommodate the incorporation of signing styles (Heloir & Gibet, 2009) or to aid in the development of more natural variability in a signer's movements yielding a less robotic signing style.

## 7. Work-in-progress

We have created new algorithms for synthesizing indexing and agreements verbs based on a corpus study. For resources, we relied on the SignStream corpora (Neidle, 2002), videos from NTID (National Technical Institute for the Deaf, 2000), and our own elicited examples. The animations are viewable at http://asl.cs.depaul.edu/LREC2012. Feedback and comments are welcome.

## 8. More Questions than Answers

The considerations mentioned in this paper are only a beginning. The following are open questions:

- **Is it too soon to think about standardization for synthesis?**

With standardization comes the potential benefit of increased collaboration and the possibility of sharing resources. However, premature standardization can omit important features that are then difficult and expensive to add.

- **What other information is necessary to synthesize other language constructs?**

Although it has been posited that only a small amount of phonemic information needs to be annotated to create correct utterances involving agreement verbs, perhaps additional data is required for other cases. What other cases should be studied?

- **How can the impact of recording additional information be minimized?**

The process of annotation is expensive, and additional tagging to support synthesis will only exacerbate the situation. Are there cases where more information can be inferred from extant data?

It is hoped that this discussion will help open a dialog to consider the alternatives and ramifications for a standardization of annotation to support synthesis.

## 9. References


Crasborn, O. (2008). Open Access to Sign Language Corpora. LREC 2008 6th International Conference on Language Resources and Evaluation. Workshop Proceedings. W25. 3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora (pp. 33-38). Paris: ELRA.

Hanke, T., Storz, J., & Wagner, S. (2010). iLex: Handling Multi-Camera Recordings. LREC 2010. 7th International Conference on Language Resources and Evaluation. Workshop Proceedings. W13. 4th Workshop on Representation and Processing of Sign Languages: Corpora and Sign Language Technologies. (pp. 110-111). Paris: ELRA.

Heloir, A., & Gibet, S. (2009). A Qualitative and Quantitative Characterisation of Style in Sign Language Gestures. In M. S. Dias, S. Gibet, M. M. Wanderley, & R. Bastos (Eds.), Gesture-Based Human-Computer Interaction and Simulation: 7th International Gesture Workshop, GW 2007, Lisbon, Portugal, May 23-25, 2007, Revised Selected Papers (pp. 122-133). Berlin, Heidelberg: Springer-Verlag.

Johnston, T., & Schembri, A. (2006). Issues in the Creation of a Digital Archive of a Signed Language. In L. Barwick, & L. Thieberger (Ed.), Sustainable Data from Digital Fieldwork (pp. 7-16). Sydney, Australia: Sydney University Press.

Lillo-Martin, D., & Pichler, D. C. (2008). Development of Sign Language Acquisition Corpora. LREC 2008 6th International Conference on Language Resources and Evaluation. Workshop Proceedings. W25. 3 rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora (pp. 129-133). Paris: ELRA.

McDonald, J., Alkoby, K., Carter, R., Christopher, J., Davidson, M., Ethridge, D., et al. (2001, May). An improved articulated model of the human hand. The Visual Computer, 17(3), 158-166.

National Technical Institute for the Deaf. (2000). American Sign Language Video Dictionary and Inflection Guide. Rochester, New York, USA.

Padden, C. (1986). Verbs and Role-shifting in ASL. Proceeings of the 4th National Symposium on Sign Language Research and Teaching. Las Vegas, Nevada.

Padden, C. (1990). The Relation between Space and Grammar in ASL Morphology. In C. Lucas (Ed.), Sign Language Research: Theoretical Issues (pp. 118-132). Washington, DC: Gallaudet University Press.

Schnepp, J., Wolfe, R., & McDonald, J. (2010). Synthetic Corpora: A Synergy of Linguistics and Computer Animation . Fourth Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies LREC 2010. Valetta, Malta: ELRA.

Toro, J. (2004). Automated 3D Animation System to Inflect Agreement Verbs. Sixth High Desert Linguistics Conference. Albuquerque, New Mexico.