

# Workshop Programme

## 3<sup>rd</sup> Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora

- 08:45 – 09:00 **Workshop opening & welcome**
- 09:00 – 09:30 Diane Lillo-Martin, Deborah Chen Pichler: Development of sign language acquisition corpora
- 09:30 – 10:00 Onno Crasborn, Inge Zwitterlood: The Corpus NGT: an online corpus for professionals and laymen
- 10:00 – 10:30 Trevor Johnston: Corpus linguistics & signed languages: no lemmata, no corpus.
- 10:30 – 11:00 **Coffee break**
- 11:00 – 11:30 Lorraine Leeson, Brian Nolan: Digital Deployment of the Signs of Ireland Corpus in Elearning
- 11:30 – 12:00 Johanna Mesch, Lars Wallin: Use of sign language materials in teaching
- 12:00 – 13:30 **Poster session 1**
- 13:30 – 14:30 **Lunch**
- 14:30 – 16:00 **Poster session 2**
- 16:00 – 16:30 **Coffee break**
- 16:30 – 17:00 Onno Crasborn: Open Access to Sign Language Corpora
- 17:00 – 17:30 Adam Schembri: British Sign Language Corpus Project: Open Access Archives and the Observer's Paradox
- 17:30 – 18:00 Cat Fung H-M, Scholastica Lam, Felix Sze, Gladys Tang: Simultaneity vs. Sequentiality: Developing a transcription system of Hong Kong Sign Language acquisition data
- 18:00 – 18:45 **General discussion**
- 18:45 – 19:00 **Workshop closing**

## **Workshop Organisers**

Onno Crasborn, Radboud University Nijmegen, the Netherlands  
Eleni Efthimiou, Institute for Language and Speech Processing, Athens, Greece  
Thomas Hanke, University of Hamburg, Germany  
Ernst D. Thoutenhoofd, Virtual Knowledge Studio for the Humanities & Social Sciences,  
Amsterdam, the Netherlands  
Inge Zwitterlood, Radboud University Nijmegen, the Netherlands

## **Programme Committee**

Penny Boyes Braem, Center for Sign Language Research, Basel, Switzerland  
Annelies Braffort, LIMSI/CNRS, Orsay, France  
Patrice Dalle, IRIT, Toulouse, France  
Evita Fotinea, Institute for Language and Speech Processing, Athens, Greece  
Jens Heßmann, University of Applied Sciences Magdeburg-Stendal, Germany  
Trevor Johnston, Macquarie University, Sydney, Australia  
Lorraine Leeson, Trinity College, Dublin, Ireland  
Adam Schembri, University College London, UK  
Graham Turner, Heriot-Watt University, Edinburgh, UK  
Meike Vaupel, University of Applied Sciences Zwickau, Germany  
Chiara Vettori, EURAC, Bolzano, Italy

# Table of Contents

Title	Page
<i>Patricia Álvarez Sánchez, Inmaculada C. Báez Montero, Ana M<sup>a</sup> Fernández</i> Soneira: Linguistic, sociological and technical difficulties in the development of a Spanish Sign Language (LSE) corpus	9
<i>Louise de Beuzeville</i> : Pointing and verb modification: the expression of semantic roles in the Auslan Corpus	13
<i>Cat Fung H-M, Scholastica Lam, Joe Mak, Gladys Tang</i> : Establishment of a corpus of Hong Kong Sign Language acquisition data: from ELAN to CLAN	17
<i>Cat Fung H-M, Felix Sze, Scholastica Lam, Gladys Tang</i> : Simultaneity vs. Sequentiality: Developing a transcription system of Hong Kong Sign Language acquisition data	22
<i>Emilie Chételat-Pelé, Annelies Braffort, Jean Véronis</i> : Annotation of Non Manual Gestures: Eyebrow movement description	28
<i>Onno Crasborn</i> : Open Access to Sign Language Corpora	33
<i>Onno Crasborn, Han Sloetjes</i> : Enhanced ELAN functionality for sign language corpora	39
<i>Onno Crasborn, Inge Zwitterlood</i> : The Corpus NGT: an online corpus for professionals and laymen	44
<i>Philippe Dreuw, Hermann Ney</i> : Towards Automatic Sign Language Annotation for the ELAN Tool	50
<i>Paul Dudis, Kristin Mulrooney, Clifton Langdon, Cecily Whitworth</i> : Annotating Real-Space Depiction	54
<i>Eleni Efthimiou, Stavroula-Evita Fotinea</i> : Annotation and Management of the Greek Sign Language Corpus (GSLC)	58
<i>Thomas Hanke, Jakob Storz</i> : iLex – A database tool integrating sign language corpus linguistics and sign language lexicography	64
<i>Annika Herrmann</i> : Sign language corpora and the problems with ELAN and the ECHO annotation conventions	68
<i>Jens Heßmann, Meike Vaupel</i> : Building up digital video resources for sign language interpreter training	74
<i>Marek Hruź, Pavel Campr, Miloš Železný</i> : Semi-automatic Annotation of Sign Language Corpora	78
<i>Trevor Johnston</i> : Corpus linguistics & signed languages: no lemmata, no corpus	82
<i>Jakub Kanis, Pavel Campr, Marek Hruź, Zdeněk Krňoul, Miloš Železný</i> : Interactive HamNoSys Notation Editor for Signed Speech Annotation	88
<i>Lutz König, Susanne König, Reiner Konrad, Gabriele Langer</i> : Corpus-based Sign Dictionaries of Technical Terms – Dictionary Projects at the IDGS in Hamburg	94
<i>Markus Koskela, Jorma Laaksonen, Tommi Jantunen, Ritva Takkinen, Päivi Rainò, Antti Raike</i> : Content-based video analysis and access for Finnish Sign Language – a multidisciplinary research project	101
<i>Klaudia Krammer, Elisabeth Bergmeister, Silke Bornholdt, Franz Dotter, Christian Hausch, Marlene Hilzensauer, Anita Pirker, Andrea Skant, Natalie Unterberger</i> : The Klagenfurt lexicon database for sign languages as a web application: LedaSila, a free sign language database for international use	105

<i>Lorraine Leeson, Brian Nolan</i> : Digital Deployment of the Signs of Ireland Corpus in Elearning	<b>112</b>
<i>François Lefebvre-Albaret, Frederick Gianni, Patrice Dalle</i> : Toward a computer- aided sign segmentation	<b>123</b>
<i>Diane Lillo-Martin, Deborah Chen Pichler</i> : Development of sign language acquisition corpora	<b>129</b>
<i>Johanna Mesch, Lars Wallin</i> : Use of sign language materials in teaching	<b>134</b>
<i>Cédric Moreau, Bruno Mascret</i> : LexiqueLSF	<b>138</b>
<i>Yuji Nagashima, Mina Terauchi, Kaoru Nakazono</i> : Construction of Japanese Sign Language Dialogue Corpus: KOSIGN	<b>141</b>
<i>Victoria Nyst</i> : Documenting an Endangered Language: Creating a Corpus of Langue des Signes Malienne (CLaSiMa)	<b>145</b>
<i>Elena Antinoro Pizzuto, Isabella Chiari, Paolo Rossini</i> : The Representation Issue and its Multifaceted Aspects in Constructing Sign Language Corpora: Questions, Answers, Further Problems	<b>150</b>
<i>Siegmond Prillwitz, Thomas Hanke, Susanne König, Reiner Konrad, Gabriele Langer, Arvid Schwarz</i> : DGS corpus project – Development of a corpus based electronic dictionary German Sign Language / German	<b>159</b>
<i>Adam Schembri</i> : British Sign Language Corpus Project: Open Access Archives and the Observer’s Paradox	<b>165</b>
<i>Sandrine Schwartz</i> : Tactile sign language corpora: capture and annotation issues	<b>170</b>
<i>Jérémie Segouat, Annelies Braffort, Laurence Bolot, Annick Choisier, Michael Filhol, Cyril Verrecchia</i> : Building 3D French Sign Language lexicon	<b>174</b>
<i>Saori Tanaka, Yosuke Matsusaka, Kaoru Nakazono</i> : Interface Development for Computer Assisted Sign Language Learning: Compact Version of CASLL	<b>178</b>
<i>Inge Zwitterlood, Asli Özyürek, Pamela Perniss</i> : Annotation of sign and gesture cross-linguistically	<b>185</b>

## Author Index

Álvarez Sánchez, Patricia	9	Leeson, Lorraine	112
Báez Montero, Inmaculada C.	9	Lefebvre-Albaret, François	123
Bergmeister, Elisabeth	105	Lillo-Martin, Diane	129
Beuzeville, Louise de	13	Mak, Joe	17
Bolot, Laurence	174	Mascret, Bruno	138
Bornholdt, Silke	105	Matsusaka, Yosuke	178
Braffort, Annelies	28, 174	Mesch, Johanna	134
Campr, Pavel	78, 88	Moreau, Cédric	138
Cat Fung, H-M	17, 22	Mulrooney, Kristin	54
Chen Pichler, Deborah	129	Nagashima, Yuji	141
Chételat-Pelé, Emilie	28	Nakazono, Kaoru	141, 178
Chiari, Isabella	150	Ney, Hermann	50
Choisier, Annick	174	Nolan, Brian	112
Crasborn, Onno	33, 39, 44	Nyst, Victoria	145
Dalle, Patrice	123	Özyürek, Asli	185
Dotter, Franz	105	Perniss, Pamela	185
Dreuw, Philippe	50	Pirker, Anita	105
Dudis, Paul	54	Pizzuto, Elena Antinoro	150
Efthimiou, Eleni	58	Prillwitz, Siegmund	159
Fernández Soneira, Ana Maria	9	Raike, Antti	101
Filhol, Michael	174	Rainò, Päivi	101
Fotinea, Stavroula-Evita	58	Rossini, Paolo	150
Gianni, Frederick	123	Schembri, Adam	165
Hanke, Thomas	64, 159	Schwartz, Sandrine	170
Hausch, Christian	105	Schwarz, Arvid	159
Herrmann, Annika	68	Segouat, Jérémie	174
Heßmann, Jens	74	Skant, Andrea	105
Hilzensauerm Marlene	105	Sloetjes, Han	39
Hrúz, Marek	78, 88	Storz, Jakob	64
Jantunen, Tommi	101	Sze, Felix	22
Johnston, Trevor	82	Takkinen, Ritva	101
Kanis, Jakub	88	Tanaka, Saori	178
König, Lutz	94	Tang, Gladys	17, 22
König, Susanne	94, 159	Terauchi, Mina	141
Konrad, Reiner	94, 159	Unterberger, Natalie	105
Koskela, Markus	101	Vaupel, Meike	74
Krammer, Claudia	105	Véronis, Jean	28
Krňoul, Zdeněk	88	Verrecchia, Cyril	174
Laakson, Jorma	101	Wallin, Lars	134
Lam, Scholastica	17, 22	Whitworth, Cecily	54
Langdon, Clifton	54	Železný, Miloš	78, 88
Langer, Gabriele	94, 159	Zwitserslood, Inge	44, 185

## Editors' Preface

This collection of papers stems from the third workshop in a series on “the representation and processing of sign languages”. The first took place in 2004 (Lisbon, Portugal), the second in 2006 (Genova, Italy). All workshops were tied to Language Resources and Evaluation Conferences (LREC), the 2008 one taking place in Marrakech, Morocco. While there has been occasional attention for signed languages in the main LREC conference, the main focus there is on written and spoken forms of spoken languages. The wide field of language technology has been the focus of the LREC conferences, where academic and commercial research and applications meet. It will be clear to every researcher that there is a wide gap between our knowledge of spoken versus signed languages. This holds not only for language technology, where difference in modality and the absence of commonly used writing systems for signed languages obviously pose new challenges, but also for the linguistic knowledge that can be used in language technologies.

The domains addressed in the two previous sign language workshops have thus been fairly wide, and we see the same variety in the present proceedings volume. However, where the first and the second workshop had a strong focus on sign synthesis and automatic recognition, the theme of the third workshop concerns construction and exploitation of sign language corpora.

Recent technological developments allow sign language researchers to create relatively large video corpora of sign language use that were unimaginable ten years ago. Several national projects are currently underway, and more are planned. In the present volume, sign language linguistics researchers and researchers from the area of sign language technologies share their experiences from completed and ongoing efforts: what are the technical problems that were encountered and the solutions created, what are the linguistic decisions that were taken?

At the same time, the contributions also look into the future. How can we establish standards for linguistic tagging and metadata, and how can we add sign language specifics to well-established or emerging best practices from the general language resource community? How can we work towards (semi-) automatic annotation by computer recognition from video? These are all questions of interest to both linguists and language technology experts: the sign language corpora that are being created are needed for more reliable linguistic analyses, for studies on sociolinguistic variation, and for building tools that can recognize sign language use from video or generate animations of sign language use.

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 the programme committee that helped us reviewing the abstracts for the workshop:

Penny Boyes Braem; Annelies Braffort; Patrice Dalle; Evita Fotinea; Jens Heßmann; Trevor Johnston; Lorraine Leeson; Adam Schembri; Graham Turner; Meike Vaupel; Chiara Vettori

Finally, we would like to point the reader to the proceedings of the previous two workshops, which form important resources in a growing field of research; both works were made available as PDF files for participants of the workshop.

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.

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.

We hope the present volume will stimulate further research by making the presentations accessible for those who could not attend the workshop.

The Editors,

Onno Crasborn, Radboud University Nijmegen (NL)

Eleni Efthimiou, Institute for Language and Speech Processing (GR)

Thomas Hanke, University of Hamburg (DE)

Ernst D. Thoutenhoofd, Virtual Knowledge Studio for the Humanities & Social Sciences (NL)

Inge Zwitterlood, Radboud University Nijmegen (NL)

# Workshop Papers

# Linguistic, Sociological and Technical Difficulties in the Development of a Spanish Sign Language (LSE) Corpus

Patricia Álvarez Sánchez, Inmaculada C. Báez Montero, Ana Fernández Soneira

Universidad de Vigo – Research Group on Sign Languages<sup>1</sup>

Lagoas-Marcosende (36310) Vigo

patri.alvarez@gmail.com, cbaez@uvigo.es, anafe@uvigo.es

## Abstract

The creation of a Spanish Sign Language corpus has been, since 1995 until 2000, one of the main aims of our Sign Languages Research Group at the University of Vigo. This research has the aim of helping us in the description of LSE and developing tools for research: labeling, transcription, etc. We obtained language samples from 85 informants whose analysis raised several difficulties, both technical and sociolinguistic.

At this stage, with renewed energy, we have taken up again our initial aims, crossing the technical, linguistic and sociological obstacles that had hindered our proposal to reach its end.

In our panel we will present, apart from the difficulties that we have encountered, the new proposals for solving and overcoming them, thus, finally reaching our initial aim: to develop a public Spanish Sign Language corpus that could be consulted online.

We will go into details with the criteria of versatility and representativity which condition the technical aspects; the sociolinguistic criteria for selecting type of discourses and informants; the labels for marking the corpus and the utilities that we pretend to give the corpus, not only centered in the use of linguistic data for the quantitative and qualitative research of the LSE, but also centered in the use for teaching.

## 1. General Approach

The study of LSE should not be dealt with in a different manner to that of any other oral language. It will be mandatory to have a textual corpus. The production of a sign language has a kinetic nature. Its reception is visual, so the conversations in sign language have to be registered in video formats.

Our contribution to the congress, in the form of a panel, is divided into three sections that correspond with the three stages of the development of our corpus. Each step is marked by a general reflection.

The first stage covers our group work from 1995 until 2000 and it represents the beginning of the process. We will present subsequently, the aims set, the steps made for the actual conception of the corpus and the difficulties encountered.

The second phase goes from 2000 till 2007. It was stressed by an analysis process of the work done, and reconsiderations on our basis due to the problems at the first stage. We will here present the data obtained and the new goals that we set.

The third and last stage corresponds with the present time. It is the time of showing our advances and the decisions made on the linguistic, sociolinguistic and technical sides.

## 2. Initial Work

*“Linguistic corpora have come to fill a privileged position because they constitute a valuable source of information for the creation of dictionaries, computational lexicon and grammars. (...) As a result, a new discipline appears: CORPUS LINGUISTICS, aimed at the processing and*

*exploitation of this type of linguistic resource.” (A, Martí, 1999)*

### 2.1. Aims

Our work was focused on obtaining a LSE textual corpus of Galician signers from which to start the research on LSE. These were our initial researching aims:

- a) Starting the description of LSE
- b) Determining which are the relevant linguistic units in SL
- c) Knowing the grammatical relational processes
- d) Developing tools for research: labeling, transcription, etc

### 2.2. Corpus features

We considered these the main features for creating a corpus:

- It must contain real data
- It must constitute an irreplaceable basis for linguistic description
- It must be completed with computing support in order to make easy its use.
- It must gather:
  - a) Informants data
  - b) Different types of discourse samples
  - c) Wide range of topics depending on the type of discourse we want to obtain, etc.
- It must be transcribed in Spanish glossas (conventions adapted from Klima & Bellugi, 1979) and subtitled in written language.

### 2.3. Process stages

We have divided into seven stages the process of creating our corpus:

<sup>1</sup> <http://webs.uvigo.es/lenguadesignos/sordos>

- Tool design for the creation of a corpus
- Criteria for the selection of informants
- Creation of a database of informants' details
- Collection of language samples
- Data storage
- Data labeling and marking
- Transcription and notation systems

## 2.4. Difficulties in the process

The difficulties that aroused throughout the research process are:

- The lack of a research tradition on Sign Languages in Corpus Linguistics forces us to solve problems from the very beginning:
  - How to delimit units in sign languages.
  - How to label the different formations for their later analysis.
  - Other related issues.
- Creation of social networks in the Deaf community with the aim of avoiding the social identity of our informants to be threatened.
- Technical restrictions. We have to select appropriate material in order to avoid problems in compatibility between the different devices (video cameras, video player, computers, software...)

## 3. Analysis and Reconsiderations

*"(...) the paradox exists that once a system is available for its use, its technology becomes obsolete with regard to the one that is operative at that moment and in many cases, it must be reprogrammed"* (A, Martí, 1999)

After these first steps, it was time to analyse the gathered data. For this purpose, we created a database of informants which we are going to present now.

### 3.1. Where did we collect our data?

We have developed an interview filing card with the purpose of ascertaining the social and linguistic profile of the Galician deaf people that were later registered in videotapes.

This is the data gathered from our 85 informants:

- Identification*: name, address and phone (for future contacts);
- Origin and social environment*: place and date of birth, age of deafness occurrence, deafness degree, deaf/hearing family, job of closest family members;
- School*: degree and type of studies, special/ordinary school, use/absence of SL in school;
- Linguistic skills*: in LSE, oral Spanish, lip-reading, written Spanish;
- Place of residence*: in order to reflect and control linguistic variation.

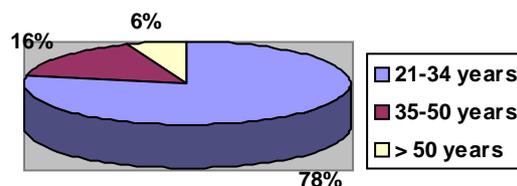


Figure 1: Distribution of informants by age group.

Distribution of informants by age group:

From 21 to 35 years: 25

From 36 to 50 years: 5

Over 50 years: 2

Total: 32 interviews.

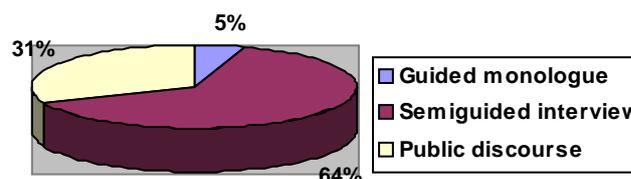


Figure 2: Distribution of language samples by gender types.

Distribution by gender types:

*Guided monologue* - 23 minutes

The signer is asked for a description of his family, his house and a short anecdote.

*Semiguided interviews* - 271 minutes

Signers are interviewed on several topics, depending on their age, sex, preferences, etc. Thus, the discourse is more spontaneous.

*Public discourse* - 130 minutes.

Conferences and round tables give us a more programmed and formal style.

### 3.2. Reconsiderations

After the research, we had to reconsiderate certain issues for a better development of our corpus. We will now sum these up:

- Revision of the projects carried out in other countries.
- Creation of social networks:

*Inside the Deaf community*:

Preparation of the members of the community for the carrying out of the interviews

*In the institutions*:

Participation in national networks for research in order to contact with the Deaf community all over Spain.

Support of the LSE Standardization Center in the creation of the corpus.

## 4. For the time being

*"If our research manages to correct mistaken or unsuitable information, we will have made a good service to linguistics; however, this type of study usually needs for certain knowledge and experiences that do not correspond with the young researcher. (López Morales, 1994, 25)"*

At this stage, with renewed energy, we have taken up again our initial aims, crossing the technical, linguistic and

sociological obstacles that had hindered our proposal to reach its end. In the following lines, we will present the advances achieved and the measures adopted for solving the problems already mentioned, in order to finally develop a public LSE corpus on-line.

#### 4.1. Advances

These are the main advances that occurred in the last years:

- We are members of a network of universities for the teaching and research on Spanish and Catalan Sign Languages (Red Interuniversitaria para la investigación y la docencia de las lenguas de señas- RIID-LLSS).
- We collaborate in the creation of a LSE Standardization Center (whose creation will be possible thanks to the pass of the Law 27/2007, 23 October, on the Use and Recognition of the Sign and the Support Media for Oral Communication).
- Our group has obtained state financing for its research project "Basis for the linguistic analysis of the Spanish Sign Language"<sup>2</sup>
- We count on three deaf teachers and four interpreters for the research and teaching tasks. We also count on specialised researchers in subtitling that will deal with the subtitling and marking tasks in the corpus<sup>3</sup>.
- In these years, several thesis and dissertations of PhD students in topics related to sign language linguistics have been published (Fernández Soneira 2004; Iglesias Lago 2006, Álvarez Sánchez 2006). Other members of this group have published research papers on grammatical aspects in reference works (Cabeza y Fernández 2004)<sup>4</sup>.

#### 4.2. Current aims

We are working for creating a textual corpus of LSE as a basis for:

- a) Development of LSE grammars. The grammatical analysis will focus on the determination of the relevant LSE units and the grammatical processes of relation.
- b) Applied research:
  - LSE interpretation
  - LSE teaching
  - Normalization and linguistic planning
  - Transcription
- c) General research:

<sup>2</sup> Basis for the linguistic analysis of the Spanish Sign Language (HUM2006-10870/FILO) funded by the Ministry of Education and Science. Length of the project: 2006-2008.

Spanish Sign Language: linguistic and philological aspects (BFF2003-05696) funded by the Ministry of Science and Technology. Length of the project: 2003-2005.

Grammatical analysis of the LSE: sociolinguistic, psycholinguistic and computational applications (PGIDT00PXIB30202PR) funded by Xunta de Galicia. Length of the project: 2000-2003.

<sup>3</sup> To consult LSE teaching staff profile, cfr. <http://www.uvigo.es/centrolinguas/index.en.htm>

<sup>4</sup> To consult the whole list of publications by the members of our research group, cfr. <http://webs.uvigo.es/lenguadesignos/sordos/publicaciones/index.htm>

Language acquisition

Linguistic universals

Other related issues

- d) Use of the corpus in the teaching platforms as a didactic element in order to provide the pupils with real language samples. These will complete the learning-teaching process started inside the classroom.

#### 4.3. Linguistic and Sociolinguistic Decisions

LSE is not an standardized language and there are very few descriptive studies on this language. This forces us to propose what kind of recordings do we want, how many people do we need in order for the corpus to be representative and real, and finally, what conclusive analysis could we obtain from it.

Taking into account these determining factors, we raise:

- Asking for the collaboration of signers of different regions of Spain for obtaining a good representation of the different geographic registers.
- Select signers that fulfill certain features: native signers of LSE, post lingual users of LSE and interpreters.
- Interview design:

Choice of deaf interviewers. Their dialogues are more natural and they obtain a higher degree of involvement from the Deaf community in this project.

Recordings should be adapted to the personality of the informants. We should take into account that most of the Deaf people don't have linguistic conscience because they have never studied their language as such. Instead, they have learnt it in a natural way as a medium for communication.

We have prepared several models of the interview, with questions that may arouse interest in the informants (on deafness, family, friends, human relationships, tobacco, etc.)

#### 4.4. Technical decisions with a view to the future

- a) Standardization of the recording format: Use of a recording set: digital cameras, similar wall background in all the recordings, identical light conditions, signers clothes, position and framing...
- b) Multiple views of the signer: face, trunk, in profile...
- c) Storage and backup of the recordings from the camera to the computer.
- d) Editing of the recordings in chapters (monologues, semi guided interviews and free conversations) for a better handling of the images.
- e) Use of the ELAN system for the notation process.
- f) Corpus labeling of grammatical features and sign configuration.
- g) Use of P2P tools for making easy the cooperation between universities or research groups with the aim of ensuring on one hand the proper management of the work teams and on the other hand, the integration of results.
- h) Enable the search and retrieval by sign configuration, grammatical aspects and signer details.
- i) Online publishing of the corpus with the aid of external financing.



Figure 3: Sample search in the future corpus

## 5. Acknowledgements

This work is part of a larger research project carried out by the Research Group on Sign Languages at the University of Vigo. Its final outcome, in the form of this paper, would not have been possible without the collaboration of Francisco Eijo Santos y Juan Ramón Valiño Freire (two of our deaf collaborators).

This research was funded by the Ministry of Education and Science, grant number HUM2006-10870/FILO. This grant is hereby gratefully acknowledged.

## 6. References

- Álvarez Sánchez, P. (2006). "La enseñanza de lengua extranjera a alumnos sordos". Diploma de Estudios Avanzados. Universidad de Vigo.
- Báez Montero, I. C. & Cabeza Pereiro, M. C. (1995). "Diseño de un corpus de lengua de señas española", XXV Simposium de la Sociedad Española de Lingüística (Zaragoza, 11-14 de diciembre de 1995).
- Báez Montero, I. C. & Cabeza Pereiro, M. C. (1999). "Elaboración del corpus de lengua de signos española de la Universidad de Vigo". Taller de Lingüística y Psicolingüística de las lenguas de signos (A Coruña, 20-21 de septiembre de 1999).
- Báez Montero, I. C. & Cabeza Pereiro, M. C. (1999). "Spanish Sign Language Project at the University of Vigo" (poster), Gesture Workshop 1999 (Gif-sur-Yvette, Francia, 17-19 de marzo de 1999).
- Cabeza Pereiro, C. & Fernández Soneira, A. (2004). "The expresión of time in Spanish Sign Language", *Sign Language and Linguistics*, vol 7/1, pp.63-82.
- Iglesias Lago, S. (2006). "Uso del componente facial para la expresión de la modalidad en lengua de signos española". Tesis doctoral inédita. Universidad de Vigo,
- Fernández Soneira, A. (2004). *La cuantificación en la lengua de signos española*, Tesis doctoral. Universidad de Vigo.
- López Morales, H. (1994). Métodos de Investigación Lingüística. Salamanca, Ediciones Colegio de España.
- Martí Antonín, M<sup>a</sup> A. (1999). "Panorama de la lingüística computacional en Europa". *Revista Española de lingüística Aplicada*, pp. 11-24.

# Pointing and Verb Modification: the expression of semantic roles in the Auslan corpus

**de Beuzeville, Louise**

Post-doctoral Research Fellow

Linguistics Department

Macquarie University NSW 2109

Australia

E-mail: [louise.debeuzeville@ling.mq.edu.au](mailto:louise.debeuzeville@ling.mq.edu.au)

## Abstract

As part of a larger project investigating the grammatical use of space in Auslan, 50 texts from the Auslan Archive Project Corpus were annotated and analysed for the spatial modification of verbs to show semantic roles. Data for the corpus comprise the Sociolinguistic Variation in Auslan Project (SVIAP) and the Endangered Languages Documentation Project (ELDP). In this paper, 20 personal narratives were analysed—10 from SVIAP and 10 from ELDP—as well as 30 retellings of two Aesop’s fables (16 of “The Boy Who Cried Wolf” and 14 of “The Hare and the Tortoise”). Each sign or meaningful gesture in the texts was identified and annotated in ELAN. These signs were then classified into word class, and the nouns and verbs tagged for whether they had the potential to be modified spatially. Next, the indicating nouns and verbs were annotated as to whether or not their spatial modification was realized. In this paper, we discuss the use of the ELAN search functions across multiple files in order to identify the proportion of sign types in the texts, the frequency with which indicating verbs are actually modified for space and the influence of the presence of a pointing sign adjacent to the verb.

## 1. Aims and Background

One of the most salient and interesting aspects of the grammar of signed languages is the use of space to track referents through discourse. One way in which this has been observed is the spatial modification of lexical verbs to show semantic roles and participants. Many previous studies have noted this and generally—when a verb has been identified as modifiable—the modification has been assumed to be obligatory (Aronoff, Meir, Padden, & Sandler, 2003; Meier, Cormier, & Quinto-Pozos, 2002; Meir, 2002; Neidle, Kegl, MacLaughlin, Bahan, & Lee, 2000; Padden, 1988; Padden, 1990). The alternative view, and one that seems to pattern better with this data, is that the modifications are gestural and the signs are a combination of morphemes and gestures (Engberg-Pedersen, 1993; Liddell, 2000, 2002, 2003b). A serious problem, however, with many previous reports on spatial modification of verbs is that they were not based on data of usage patterns, but rather native speaker intuitions.

Part of the reason for this lack of data of usage patterns was the technology available at the time of the research. Prior to the digital age, the collection of large amounts of data was difficult and expensive, as was the storage and accessibility of such data. Even more challenging, however, was the task of transcribing or annotating the data and then searching it for the relevant aspects of the grammar and their co-occurrence with other features.

These problems are now being overcome: data can easily and affordably be filmed and stored digitally; annotations can occur in software with a machine readable format; and as such, searches can be carried out by computers on single or multiple texts at the same time, thus decreasing human error in data analysis and saving countless hours of manual labour.

As part of a larger project investigating the grammatical use of space in Auslan, 50 texts from the

Auslan Corpus have been annotated and analysed for the spatial modification of verbs to show semantic roles. Using ELAN (EUDICO Linguistic Annotator) software, which allows for multiple tiers of annotations to be time aligned with multimedia files, these texts have been analysed for:

- a) the number and types of verbs used;
- b) the proportion of modifiable verbs which have actually been modified in the text; and
- c) the influence of pointing signs on the modification.

The hypothesis is that the presence of at least one adjacent pointing sign would decrease the likelihood of the sign being modified. Linked to that, it is expected that for indicating verbs with adjacent pointing signs, a lesser proportion would be modified than what occurred when there was no adjacent pointing sign.

In this paper, I discuss some of the features of ELAN that have been used to enable a search of a large amount of data with a relatively small amount of labour. I will first discuss the methodology used, then some of our previous and current results before discussing conclusions that can be drawn.

## 2. Methodology

### 2.1 Data

Data for this paper come from the Auslan Archive Project Corpus, which consists of two large corpora: the Sociolinguistic Variation in Auslan Project (SVIAP) and the Endangered Languages Documentation Project (ELDP). The SVIAP corpus is made up of films of 211 participants from all over Australia, resulting in 150 hours of edited footage of free conversation, a more formal interview, and lexical elicitation tasks. The ELDP corpus has 150 hours of edited video from 100 participants all over Australia (many the same as filmed for the SVIAP corpus). The ELDP data consists of the retelling of a

narrative, responding to formal interview questions, an attitude questionnaire, a spontaneous narrative, and some elicitation tasks for specific linguistic features. Participants were filmed by and interacted with other native signing deaf adults.

For this paper, ten spontaneous narratives were sourced from the SVIAP corpus. The second set of texts—from the ELDP corpus—consisted of 10 spontaneous personal recounts of a memorable event, as well as 30 retellings of 2 Aesop’s fables (16 of “The Boy Who Cried Wolf” and 14 of “The Hare and the Tortoise”). Participants were given an English version of the fable a week before filming and were told they would retell the story a week later.

The texts from both corpora were recorded on digital videotape, annotated using ELAN software, and analysed with ELAN and Excel. This process is explained below.

## 2.1 Analysis

In the ELAN file, users are able to specify a limitless number of tiers on which to annotate different features of a text. For this project, on the first two tiers (one for each hand), each of the texts were given a shallow gloss: that is, each sign was identified and labeled with an English “equivalent”. This was able to be done consistently due to the existence of the Auslan Lexical Database (for more information on the database and this process, see Johnston, these proceedings). This allowed for accurate counting of lexicalized signs for frequency counts of types and tokens.

In this first stage of glossing, every meaningful manual action was annotated, including: lexical signs from the Lexical Database, depicting signs, gestures, and points. Points were coded as either:

- a) a personal pronoun;
- b) a possessive pronoun;
- c) a demonstrative;
- d) a locative; or
- e) a point to a buoy handshape (Liddell, 2003).

A sign counted as a point regardless of the handshape used if it was used in a pointing manner. This was important as many point signs occurred with alternate handshapes due to the assimilation of the features of surrounding signs.

First person singular pronouns as well as points to buoy handshapes were clear as their form is different to other handshapes. However, since the form of most other points is identical regardless of whether they are referring to a non-present referent or a location, it was often impossible to be sure of the meaning of a signer’s point. Thus, in this first parse of the data, many points were coded simply as unclear.

A second tier dealt with the grammatical class of each sign as well as its spatial potential. Verbs were divided into plain verbs (those unable to be moved about or located in space), depicting verbs (classifier signs), and indicating verbs (directional or locatable). The table below defines each of these categories.

CATEGORY	EXPLANATION
depicting verb	A verb created on the spot that is not found in the dictionary or lexical database (classifier signs).
plain verb	A lexical verb that cannot physically be moved about in space; usually it is body

	anchored in some way.
indicating verb-directional	A lexical verb that can be moved meaningfully through space to show the semantic role of at least one participant.
indicating verb/noun-locatable	A lexical verb that can be located meaningfully in space, though not moved through space. Often because it has no path movement.

Table 1: Sign classes for spatial modifiability

Next there was a tier on which indicating verbs were marked for whether their spatial potential was realized: that is, were they moved meaningfully in space to show the semantic role of at least one participant. There were three possibilities: modified, not modified, or congruent—that is, citation in form, but that form was consistent with the spatial arrangement. These are explained in Table 2 below.

SPATIAL MODIFICATION	EXPLANATION
modified	The sign was modified spatially, i.e., it was not the citation form of a sign.
unmodified	The sign was spatially unmodified, i.e., it was produced in the citation form and was not congruent with the spatial framework. If it had been modified, it would/should have looked different to the citation form.
congruent	Of those unmodified forms, some were congruent with the spatial arrangement already set up. That is, any modification (if it were really there) would be ‘invisible’ because it would still look like the citation form.

Table 2: Codes for the realization of spatial potential.

Once all of the annotations were complete, search procedures were carried out through ELAN. Searches were carried out two ways: searching individual files in detail; or conducting a structured search on all 50 files at once in less detail.

In a previous round of the project (de Beuzeville et al., forthcoming; Johnston et al., 2007) searches were carried out on each file individually, for the following features:

- a) the number of annotations per file;
- b) a type/token analysis;
- c) an analysis of the frequency of each type;
- d) the number (and percentage) of each word class: and in particular, the spatial potential of nouns and verbs;
- e) the proportions of modifiable signs which were actually modified; and
- f) how often a period of constructed action (role shift) co-occurred with modified and unmodified signs.

These figures were all exported into Excel, those for all texts added together and calculations carried out. In addition, all tokens with all information attached were run through Varbrul for an analysis of statistical significance.

For this paper, the searching was conducted on all 50 files together, through the new structured search across multiple files option in ELAN. All indicating verbs were

identified, as well as the sign or gesture that occurred directly before or after. Each token of an indicating verb was also marked as modified, not modified or congruent. This data was then exported to Excel and all instances of pointing signs occurring directly before or after an indicating verb were identified and counted, in order to calculate whether the co-occurrence of a point and an indicating verb had an effect on its modification.

For this paper, the following analyses were carried out in Excel:

- a) a comparison of indicating verbs with or without a point sign adjacent and whether it influenced the modification;
- b) a comparison of all of the verb signs with point signs adjacent and whether they were more likely to occur with verbs that were modified, unmodified or congruent; and
- c) the frequency of each type of point sign.

### 3. Results

Despite the claim that indicating verbs in signed languages are obligatorily modified ('inflected') with respect to loci in the signing space in order to show person 'agreement', we found that these verbs are actually only spatially modified about a third of the time (de Beuzeville et al., forthcoming; Johnston et al., 2007).

Altogether the data being analysed contained just over 8,500 sign tokens, with about 40,000 annotations in total. Below is a figure which shows the amount of tokens for each type of verb—in terms of their spatial potential—as a proportion of all verbs. As can be seen, over half of all lexical verbs which are able to show semantic roles through spatial modification were not actually modified (61% of all indicating verbs).

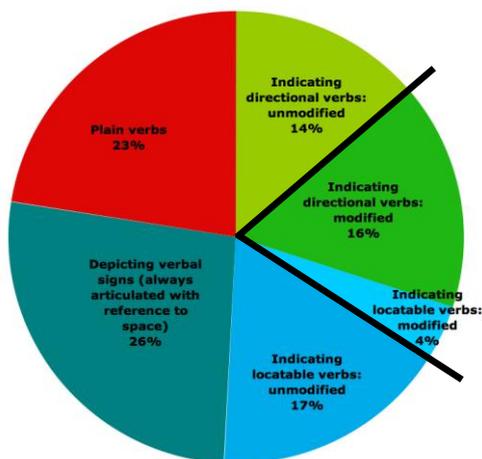


Figure 1: Proportions of types of verbs and the realization of their spatial potential.

In previous analyses of the data by Varbrul, the following factors were found to account for the variability in verb modification (de Beuzeville, et al., forthcoming; Johnston et al., 2007):

- a) indicating signs that are directional favour modification compared to locatable signs;
- b) locatable verb signs favour spatial modification compared to locatable noun signs;

- c) the five most frequent indicating verbs favour modification compared to other indicating verbs; and
- d) the presence of constructed action significantly favoured spatial modification, especially with modified verbs.

In this stage of the project, the focus is on what effect adjacent points may have on the likelihood of modification. The analysis showed that the presence of pointing did indeed have some effect on the proportion of tokens modified. As can be seen from Figure 2 below, indicating verbs were modified 41% of the time when there was no adjacent point, and this went down to 34% when there was. Unmodified indicating verbs went from 43% without a point sign to 47% with an adjacent point. These changes are in the direction predicted, but may not be statically significant. Interestingly signs that were congruent followed the pattern of unmodified signs.

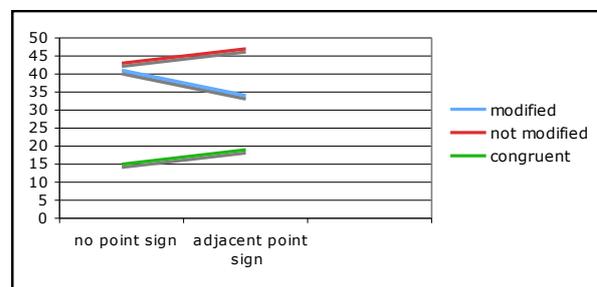


Figure 2: modification of indicating verbs with and without adjacent point signs

Further, modified signs were less likely in general to have a point sign adjacent. Figure 3 shows that of all modified indicating verbs, only 19% had an adjacent point, whereas for the indicating verbs that were not modified that figure was 24%. Congruent signs had an adjacent point sign 26% of the time, again patterning most similarly to the not modified signs.

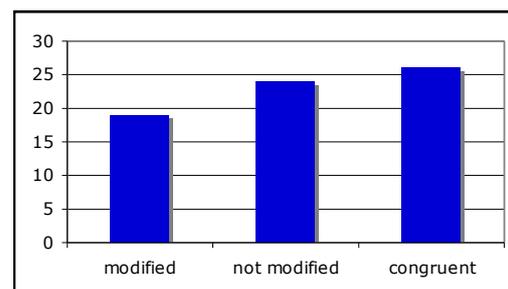


Figure 3: Proportion of modified, not modified, and congruent indicating verbs that occurred with an adjacent pointing sign.

Clearly, the type of point needs to be taken into account, since not all points give information about the semantic role of the participants of the verb. The hypothesis is that only those that do mark the semantic roles would affect modification of verbs. Figure 4 below shows the frequency of the main 3 types of points (accounting for 70% of the data): first person singular pronouns (39%), third person singular pronouns (15%) and demonstratives and locatives (16%). Approximately 14% of tokens were unclear as to their semantic function,

and the remaining 16% are made up of other types of points.

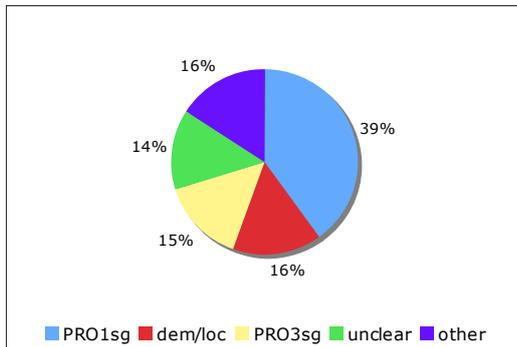


Figure 4: The frequency of different types of pointing signs

The next step in the project is to analyze the effects of individual types of points on modification, as well as to look at directional and locatable verbs separately. These are much needed analyses in order to be able to rely on the finding that adjacent points influence the modification of indicating verbs.

These results help determine where and when the spatial modification of indicating verbs is used in natural Auslan texts (and potentially other signed languages) and they indicate that the presence of an adjacent point appears to have an effect on the modification of indicating verbs, be they locatable or directional.

#### 4. Conclusion

The data presented above is an attempt to account for variability of the modification of indicating verbs. The study needs, however, to go further before any firm conclusions can be drawn.

The immediate priorities of the project are to:

- analyze the effect of different types of points on modification of indicating verbs;
- analyze the verbs that are affected according to whether they are locatable or directional; and
- carry out tests of statistical significance.

In addition, it will be necessary in the future to look at a larger environment than the sign before or after in order to see whether points further away influence modification as well, and to add more data and from non-narrative text types. It will also be necessary to decide how best to deal with signs that are congruent: that is, whether they should be assumed to be modified, treated as unmodified, or left out of the analysis as ambiguous examples.

Whatever the factors that affect the modification of indicating signs, the fact remains that they are not modified obligatorily. Thus, the data are not compatible with the view that spatial codings are highly grammaticalised or a system of verb agreement, since such systems of agreement should allow for referential cohesion and referent tracking, be head marked versus dependent-marked, obligatory and grammaticalised (that is, bleached of meaning).

Based on this data, we suggest that 1) the degree of grammaticalization of indicating verbs may not be as great as once thought and 2) the apparent non-obligatory or variable use of spatial modifications may be partly

accounted for by the presence of pointing signs—very frequent in signed texts—before or directly after the verb.

#### 5. Acknowledgements

Australian deaf community native signer participants contributing to the corpus; Associate Professor Trevor Johnston and Dr Adam Schembri for access to data and assistance with analysis; and research assistants and additional ELAN annotators: Julia Allen, Donovan Cresdee, Karin Banna, Michael Gray, and Della Goswell. This work was supported by Australian Research Council grants #LP0346973 & #DP0665254 and an Endangered Languages Documentation Project grant #MDP0088.

#### 6. References

- Aronoff, M., Meir, I., Padden, C. A., & Sandler, W. (2003). Classifier constructions in two sign languages. In K. Emmorey (Ed.), *Perspectives on classifier constructions in sign languages* (pp. 53-84). New Jersey: Lawrence Erlbaum Associates.
- de Beuzeville, L., Johnston, T. A., & Schembri, A. (forthcoming) *The use of space with lexical verbs in Auslan*.
- Engberg-Pedersen, E. (1993). *Space in Danish Sign Language* (Vol. 19). Hamburg: SIGNUM-Verlag.
- Johnston, T. A., de Beuzeville, L., Schembri, A., & Goswell, D. (2007). *On not missing the point: indicating verbs in Auslan*. Paper presented at the 10<sup>th</sup> International Cognitive Linguistics Conference, Krakow, Poland, 15<sup>th</sup> – 20<sup>th</sup>, 2007
- Liddell, S. K. (2000). Indicating verbs and pronouns: Pointing away from agreement. In K. Emmorey & H. Lane (Eds.), *The signs of language revisited: An anthology to honor Ursula Bellugi and Edward Klima* (pp. 303-320). New Jersey: Lawrence Erlbaum Associates.
- Liddell, S. K. (2002). Modality effects and conflicting agendas. In D. F. Armstrong, M. A. Karchmer & J. V. Van Cleve (Eds.), *The study of signed languages: Essays in honor of William Stokoe* (pp. 53-81). Washington, DC: Gallaudet University Press.
- Liddell, S. K. (2003b). *Grammar, gesture and meaning in American Sign Language*. Cambridge, England: Cambridge University Press.
- Meier, R. P., Cormier, K. A., & Quinto-Pozos, D. (2002). *Modality and structure in signed and spoken language*. Cambridge: Cambridge University Press.
- Meir, I. (2002). A cross-modality perspective on verb agreement. *Natural Language and Linguistic Theory*, 20(2), pp. 413-450
- Neidle, C., Kegl, J., MacLaughlin, D., Bahan, B., & Lee, R. G. (2000). *The Syntax of American Sign Language: Functional categories and heirarchical structure*. Cambridge, MA: MIT Press.
- Padden, C. A. (1988). *The interaction of morphology and syntax in American Sign Language*. New York: Garland.
- Padden, C. A. (1990). The relation between space and grammar in ASL verb morphology. In C. Lucas (Ed.), *Sign language research: Theoretical issues* (pp. 118-132). Washington, DC: Gallaudet University Press.

# Establishment of a corpus of Hong Kong Sign Language acquisition data: from ELAN to CLAN

Cat Fung H-M, Scholastica Lam, Joe Mak, Gladys Tang

Centre for Sign Linguistics and Deaf Studies

203, Academic Building #2, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

E-mail: cat\_cslds@cuhk.edu.hk, schola\_cslds@cuhk.edu.hk, joemakkl@yahoo.com.hk, gtang@arts.cuhk.edu.hk

## Abstract

This paper introduces the Hong Kong Sign Language Child Language Corpus currently developed by the Centre for Sign Linguistics and Deaf Studies, the Chinese University of Hong Kong. When completed, the corpus will include both longitudinal and cross-sectional data of deaf children acquiring Hong Kong Sign Language. Our research team has decided to establish a meaning-based transcription system compatible with both the ELAN and CLAN programs in order to facilitate future linguistic analysis. The ELAN program, which allows multiple-tier data entries and synchronization of video data with glosses, is an ideal tool for transcribing and viewing sign language data. The CLAN program, on the other hand, has a wide range of well-developed functions such as auto-tagging and the 'kwal' function for data search and they are extremely useful for conducting quantitative analyses. With add-on programs developed by our research team and additional functions in CLAN developed by the CHILDES research team, the transcribed data are transferable from the ELAN format to CLAN format, thus allowing researchers to optimize the use of both programs in conducting different types of linguistic analysis on the acquisition data.

## 1. Introduction

The establishment of the Hong Kong Sign Language Child Language Corpus began in 2002 as one of the research outputs of two RGC-funded research projects entitled "Development of Hong Kong Sign Language by Deaf Children" and "Acquisition of Classifiers in Hong Kong Sign Language by Deaf Children". The major goal of this corpus is to collect, transcribe and tag acquisition data of Hong Kong Sign Language (hereafter HKSL) that would facilitate the long-term development of sign language acquisition research. When completed, the corpus will contain acquisition data collected both longitudinally and cross-sectionally. The transcription system of the corpus is based on the CHAT format with additional symbols for properties specific to sign languages, thanks to the assistance and advice from the research team of the Child Language Data Exchange System (CHILDES) headed by Brian MacWhinney. The finalized transcriptions are compatible with the CLAN program of CHILDES as well as the ELAN program developed by Max Planck Institute for Psycholinguistics Nijmegen, The Netherlands. A major strength of this transcription system is that researchers can have full access to the existing features or functions of both programs. On the other hand, researchers can compare signed and spoken acquisition data with ease using the CLAN interface. This paper describes the procedures we went through in transcribing the HKSL acquisition data: how the data were first transcribed in ELAN and then exported to a format compatible with CLAN. In Section 2 we will briefly introduce our transcription system. Section 3 describes the initial transcription procedure. Section 4 explains the technical steps involved in exporting the data from ELAN to CLAN. Section 5 discusses the difficulties

we encountered in the process of transferring the data. Section 6 is the conclusion.

## 2. Transcription system developed by the Hong Kong Sign Language acquisition research team

Our research team aimed at achieving the following goals when developing the transcription system of the Hong Kong Sign Language Child Language Corpus:

- (a) The transcription system must be transparent enough for easy viewing. That is, the transcribed data should be accompanied with an appropriate amount of linguistic information presented in an easy-to-read format.
- (b) The transcription system should be compatible with other well-established computerized corpora so that researchers can make full use of the functions of these programs and solicit technical support from the developers of these programs when necessary.
- (c) The transcription system should facilitate cross-linguistic and cross-modal comparative studies.

For the ease of data viewing, all lexical signs are glossed with English word(s) which bear the closest possible meanings, e.g. *BOOK*, *FATHER*, *DANGEROUS* (see Figure 1). If more than one English word is needed to stand for the meaning of a sign, an underscore is used to connect these English words, as in *NAME\_SIGN* and *BIRTHDAY\_CAKE*.<sup>1</sup> If there are several synonyms in

---

<sup>1</sup> In the sign language literature, English words that are used to gloss the meaning of a sign are usually connected by hyphens.

English that can match the meaning of a sign, only one is chosen to ensure the consistency and accuracy of data coding. Supplementary codings are adapted from the CHAT specification to mark grammatical properties specific to sign languages. For example, the gloss of a spatial verb is followed by a hyphen and a small letter that indicates the locative affixes, as in *PUT-a*, *PUT-b* and *PUT-c*. Note that at this initial stage of transcription the letters ‘a’, ‘b’ and ‘c’ are abstract in nature – they do not represent specific locations in the signing space. Rather, they simply show that locative marking is present with the glossed sign (see Figure 2 and 3).



Figure 1: lexical sign for *BOOK* in HKSL

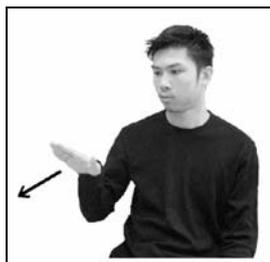


Figure 2: citation form of *PUT* in HKSL<sup>2</sup>

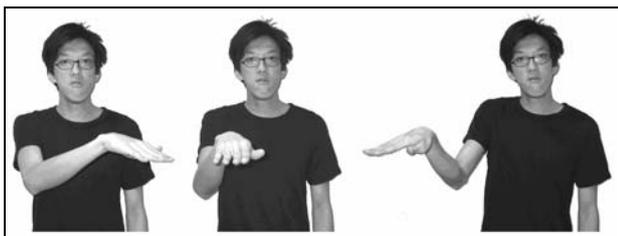


Figure 3: Spatial verb *PUT* with loci marked as *a*, *b* and *c*; glossed as *PUT-a*, *PUT-b* and *PUT-c*

In our acquisition corpus, lexical signs, gestures and simple classifier predicates are glossed on a single glossing tier (i.e. gloss 1), with the exceptions of simultaneous constructions involving independent morphological units produced separately by two manual articulators. In the latter case, the signs produced by the two hands would be glossed on the ‘gloss 1’ tier and ‘gloss 2’ tier (‘g1’ and ‘g2’ in short form) respectively.<sup>3</sup>

However, this convention is in conflict with the existing annotation convention of CHILDES. We therefore replaced underscores with hyphens.

<sup>2</sup> Photos in Figure 1 and 2 are taken from Tang (2006).

<sup>3</sup> Details of the glossing system for signs in general and

### 3. Transcription procedures

#### 3.1 Initial Transcription in the ELAN program

Viewing of sign language transcription is relatively more convenient in the ELAN program than in the CLAN program because in the former multiple-tier annotations with time alignment are possible and the annotations are synchronized with the video images. It is therefore decided that the transcription of the HKSL acquisition data be done with the ELAN program first. The transcription is done by deaf researchers who are native signers of HKSL. Delimiters are also added by the deaf researchers at the last annotation of each sentence/utterance.

#### 3.2 A table of glosses for consistency check and tagging

To check the consistency of the glosses, an add-on program is developed by our research team to examine the transcriptions on the ‘gloss 1’ and ‘gloss 2’ tier. Error messages are generated if the program notices any formatting typos in the annotations, such as a gloss with an open square bracket ‘[’ but not a close square bracket ‘]’. When all errors spotted by the program are corrected, a table containing all the glosses in the data will be generated for the purpose of consistency check, substitution and tagging. (See Figure 4)<sup>4</sup> The table consists of four columns: Glosses, Grammatical Category, Substitution and Files. Information of the first and the fourth column is generated by the add-on program. For the column of Glosses, the same English glossing items found in a selected set of files will only appear once. For instance, as shown in Figure 4, the sign *IX\_I* appears in the ELAN file ‘CC02017.eaf’ and ‘CC030713.eaf’ respectively. The entry *IX\_I* appears once only in the table, with the names of the files containing the sign listed in the fourth column. The researchers would need to go through this table with naked eyes to check the consistency of the English glosses. For example, it has been decided that in our transcription system the V-handshape sign should be glossed as *SEE* but sometimes it may be mistakenly glossed as *LOOK\_AT*. This type of inconsistency is unavoidable because the data transcription has been done by more than one deaf researcher.<sup>5</sup> When this happens, the researchers can type in *SEE* in the Substitution column for the gloss entry

simultaneous constructions involving two manual articulators will be given in another oral presentation from our colleagues.

<sup>4</sup> Since the add-on program is developed in an early stage of the establishment of the corpus, the program can only generate the glosses for the transcription using the internal transcription coding.

<sup>5</sup> When two or more English words match the meaning of a sign, the one with a more general meaning will be chosen, for example, we have chosen ‘MALE’ instead of ‘MAN’ or ‘BOY’. When two or more signs with the same meaning can only be translated with one English word, we use *\_1*, *\_2*, etc. to denote different signs, such as *LIGHT\_1* for brightness and *LIGHT\_2* for weight.

*LOOK\_AT*. By the same token, if typos are found in the glosses, e.g. *BOOK* is spelt as *BOO* by mistake, the researcher can enter the correct form in the Substitution column. The table with filled information on the substitution column will then be processed by the add-on program again and the substitutions will be performed automatically by the program in the selected ELAN files.

As for the column of Grammatical Category, the researchers will need to input the grammatical categories of all the gloss entries in the table manually. For example, *PUT* is a spatial verb and it is tagged as ‘v:sp’, whereas *IX\_1* is tagged as ‘n:pro’ to show that it is a pronoun.<sup>6</sup> The completed table will become part of the source code for tagging in the future. The following figure shows the outlook of the table:

Glosses	Grammatical category	Substitution	ELAN files	
LOOK_AT	v:agr	SEE	CC040621.eaf	
BOO	n	BOOK	CC030713.eaf	
PUT	v:sp	<sub>	CC030523.eaf	CC040621.eaf
IX_1	n:pro	<sub>	CC020617.eaf	CC030713.eaf

Figure 4: Table generated by the add-on program for consistency check and tagging

#### 4. Utterance and morphosyntactic tier – from ELAN to CLAN

##### 4.1 Generation of utterance tier and morpho-syntactic tier

When the consistency of the glosses is checked, the two tiers for glosses in the ELAN format and the glossing table will be processed by the add-on program to generate an utterance tier and a morphosyntactic tier for each signing participant.<sup>7</sup> The add-on program will automatically combine the glosses on the two glossing tiers to form an utterance tier. Sentence/utterance boundaries are detected on the basis of the delimiters added earlier on the two glossing tiers. The majoring of codings and symbols on the utterance tier are generated automatically by the add-on program, but a few more require manual input. The utterance tier becomes the main line of the transcription (*\*BRE* in Fig 5). At the same time, the information on the grammatical categories listed in the glossing table will be used to generate the morphosyntactic tier, in which each single gloss will be mapped with its corresponding tag. When the utterance and morphosyntactic tier are completed, the Elan files including all of the transcription tiers will then be exported to a CLAN readable format. The following figure shows the transcription of a sentence by a deaf

<sup>6</sup> Pronouns in Hong Kong Sign Language are indexical signs represented by ‘IX’ in the corpus. ‘1’, ‘2’, ‘3’ represent 1st, 2nd and 3rd person respectively.

<sup>7</sup> Some of our earlier files were transcribed with a glossing system incompatible with the CHAT specifications. Another function in the add-on program was designed to convert these glosses into forms compatible with the CHAT format.

adult in the ELAN program:

*BRE	IX_1 ELDER_SISTER IX_3 ELDER_SISTER SAME_2
%mor@BRE	n:pro IX_1 n ELDER_SISTER n:pro IX_3 n ELDER_SISTER adj SAME_2
%xg1@BRE	IX_1 >  IX_3 >  SAME_2
%xg2@BRE	ELDER_SISTER   ELDER_SISTER

Figure 5: A sample of the transcription in the ELAN program

(meaning: “You address both me and her as ‘elder-sister’.”)

##### 4.2 Coding for CHAT format

As the glosses correspond to individual signs only, certain utterance-level information, e.g. whether an utterance involves repetition or imitation, cannot be coded clearly on the two glossing tiers. In the process of generating the utterance tier, the add-on program can recognize certain set patterns of annotations, such as repetition of a sequence of signs. For example, if the signer produces the sign sequence ‘A B, A B’, additional symbols ‘< > [/]’ matching the CHAT specification are added automatically by the add-on program to result in ‘<A B> [/] A B’ on the utterance tier. Another auto-formatting generated by the add-on program is the switch ‘[+ imit]’, which marks imitation of a whole utterance on the utterance tier of the deaf child. For example, the deaf adult produces a sequence of signs and the deaf child produces the same sequence of signs by imitation. Each of the imitated sign on the glossing tier of the deaf child is followed by ‘[”]’. The add-on program, when generating the deaf child’s utterance tier, will recognize these symbols and automatically add ‘[+ imit]’ to the end of the imitated utterance. In the CLAN program, researchers can decide whether these utterances should be included in their analysis or not.

However, a number of additional codings for different types of simultaneous constructions need to be added manually by the researchers. For example, in certain simultaneous constructions, the two manual articulators produce signs that do not combine syntactically to form a phrasal category (e.g. the co-articulation of *IX\_2* and *LIE* as in Figure 6). On the utterance tier additional symbols ‘<A~B> [% sim]’ are added to indicate that the sequence of signs enclosed by the angle brackets does not reflect the actual order of appearance, i.e. the two signs are produced simultaneously rather than sequentially.

*BRE	<IX_2~LIE> [% sim] IX_obj~GIVE-2S&Obj IX_obj [= sweets] SWEETS NOT
%mor@BRE	n:pro IX_2 v:pl LIE v:agr GIVE-2S&Obj ix:obj IX_obj n SWEETS neg NOT
%xg1@BRE	IX_2  IX_obj~GIVE-2S&Obj  SWEETS  NOT
%xg2@BRE	LIE  IX_obj [= sweets]  >

Figure 6: Representation of simultaneous signing in ELAN interface

(meaning: “You lie then I won’t give you any sweets.”)

In some cases, a sign is first held in the signing space for a prosodic function and is then re-activated again to form a larger morphosyntactic complex with the co-occurring signs. In Figure 7 below, the *TWO\_LIST* is first held by the weak hand and is reactivated again later and combines with *IX\_TWO* to form a noun phrase. Two sets of symbols, namely, ‘&{l=SIGN’ and ‘&}l=SIGN’, are added on the utterance tier to indicate the duration for which the sign *TWO\_LIST* is held in the signing space.<sup>8</sup>

*CHI		TWO_LIST+IX_ONE&{l=TWO_LIST}NOT TWO_LIST+IX_TWO NOT TWO_LIST+IX_ONE	
%mor@CHI		num+num TWO_LIST+num IX_ONE neg NOT num+num TWO_LIST+num IX_TWO	
%xg1@CHI		TWO_LIST >	
%xg2@CHI		IX_ONE   NOT   IX_TWO   NOT	
(cont'd)	*CHI	TWO_LIST+IX_TWO&{l=TWO_LIST}RED [*]	
%mor@CHI		neg NOT num+num TWO_LIST+num IX_ONE num+num TWO_LIST+num IX_TWO adj RED	
%xg1@CHI		TWO_LIST >	
%xg2@CHI		IX_ONE   IX_TWO	

Figure 7: Representation of simultaneous reactivation in sign holding in ELAN interface – CC 3;5;23<sup>9</sup>

(meaning: “There are two: this one is not, that one is not; this and that are red.”)

### 4.3 Exporting the data from ELAN to CLAN

After the utterance tier is generated and the additional codings are included manually, the transcribed ELAN data will be exported to a CLAN readable format by using the function ‘ELAN2CHAT’ in the CLAN program. Using the ‘CHAT2ELAN’ function, data from CLAN files can also be transferred back to the ELAN program. Any changes in the ELAN/CLAN file can be converted back to the CLAN/ELAN interface using these two functions. The following table shows the outlook of the exported files in the CLAN format.

*BRE:	IX_1 ELDER_SISTER IX_3 ELDER_SISTER SAME .
%mor:	n:prj IX_1 n ELDER_SISTER n:prj IX_3 n ELDER_SISTER adj SAME .
%xgl1:	IX_1 .
>	.
IX_3 .	
SAME .	
%xgl2:	ELDER_SISTER .
ELDER_SISTER .	

Figure 8: Representation of the tiers corresponding to one signer in the HKSL acquisition corpus in CLAN interface

(meaning: “You address both me and her as ‘elder-sister’.”)

<sup>8</sup> Note that in our corpus, classifier predicates are glossed according to the adjective/verb root of the predicates and the handshape morphemes only. Other morphemic units, such as locatives, are not yet included in the glosses at this stage. Below is the transcription of an example that involves a two-handed classifier predicate meaning “a cup on the table”:

utt: put+CL\_hand:cup+be\_located+CL\_sass:table [= a cup on the table]  
 g1: put+CL\_hand:cup [= a cup on the table]  
 g2: be\_located+CL\_sass:table

<sup>9</sup> CC is short for the name of a longitudinal subject in the Hong Kong Child Language Corpus. This data is taken from the corpus in which ‘CHI’ stands for the subject ‘child’ in the data.

Note that in Figure 8 the bullet at the end of each line corresponds to the video clip linked to the utterance or the sign on the same line. The video clip will be played in the CLAN video-player when the button is clicked. .

One major advantage of our transcription system and add-on program is that the functions/features of both ELAN and CLAN are made accessible to the researchers. As ELAN allows multiple-tier entries and synchronization of video data with glosses, it is an ideal tool for transcribing and viewing sign language data. The CLAN program, on the other hand, has a wide range of functions, like auto-tagging and ‘kwal’ function for searching data, which can facilitate the linguistic analysis of sign language data. Exporting the sign language data to a readable format in CLAN also allows researchers to compare the acquisition data between spoken language and sign language.

## 5. Problems encountered in the course of setting up the corpus

We encountered a number of problems in the course of establishing the current acquisition corpus. Generating morphosyntactic tiers with the add-on program requires the tagging table as mentioned in Section 3.2. The grammatical categories are input manually for the whole batch of data and the process is repeated when a new batch of data is transcribed. To facilitate the tagging process, the research team is now switching to CLAN using the auto-tagging function and the establishment of the HKSL lexicon is now in progress.

On the other hand, the add-on program can only generate the tagging table for the transcription data following the internal transcription system which is used in an earlier stage. Despite of the transferring function of the add-on program, the research team is now switching to the CHAT transcription system on ‘gloss 1’ and ‘gloss 2’ tiers. Further development of the add-on program is required in order to support the existing ‘substitution’ function.

## 6. Conclusion

At present, the transcriptions in the Hong Kong Sign Language Child Language Corpus consist of glosses for the manual articulators and the data are convertible between CLAN and ELAN. The development of such a transcription system and the add-on program makes the functions/features of both ELAN and CLAN accessible to the researchers. On the other hand, as the data are readable in the CLAN format, researchers can make use of the functions and other child language data in the CHILDES to conduct cross-linguistic and cross-modal comparisons.

## 7. Acknowledgement

The development of the Hong Kong Sign Language Child Language Corpus was initially supported by the Research Grant Council (RGC Grant #CUHK4278/01H; Direct Grant #2081006). Our corpus has been supported by Mr. Alex K. Yasumoto’s private donation since 2006. In

addition, we would also like to thank the four deaf research assistants, Kenny Chu, Phippen Wong, Anita Yu and Brenda Yu, for data collection and transcription. Our thanks also go to Brian MacWhinney for his valuable advice in helping us devise a transcription system compatible with the CHAT format of the CHILDES corpus.

## 8. References

- MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- Tang, G. (2006) *A Linguistic Dictionary of Hong Kong Sign Language*. Hong Kong: Chinese University Press.
- Vermeerbergen, M., Demey, E. (2007). Comparing aspects of simultaneity in Flemish Sign Language to instances of concurrent speech and gesture. In M. Vermeerbergen, L. Leeson & O. Crasborn (Eds) *Simultaneity in Sign Languages: Form and Function*. Amsterdam/Philadelphia: John Benjamins, pp.257--282.
- MacLaughlin, D., Neidle, C., Greenfield, D. (2000). *Sign Stream TM User's Guide (Version 2.0)*. University of Boston.

# Simultaneity vs Sequentiality: Developing a transcription system of Hong Kong Sign Language acquisition data

Cat Fung H-M, Felix Sze, Scholastica Lam, Gladys Tang

Centre for Sign Linguistics and Deaf Studies

203, Academic Building #2, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

E-mail: cat\_cslds@cuhk.edu.hk, felix\_cslds@cuhk.edu.hk, schola\_cslds@cuhk.edu.hk, gtang@arts.cuhk.edu.hk

## Abstract

Sign languages are characterized with a wide range of constructions which encode information of various linguistic levels simultaneously in different autonomous channels. Specifically, the signs produced by the two manual articulators may exhibit a varying degree of relatedness or integration with respect to their semantic, morphological, or syntactic characteristics. In a two-handed lexical sign, the two hands form a single morphemic unit which cannot be further decomposed morphologically. In a typical two-handed classifier construction that is made up of two independent classifiers, the handshape, movement, and location of each of the two hands bear a morphemic status and these morphemes are put together to form a larger morphosyntactic complex. In a signing discourse, it is not uncommon to see the whole or part of a completed sign to be held in space in one hand, while another sign is produced by the other hand. In some cases, the held sign may bear no morphosyntactic relation with the co-occurring sign and its presence only serves a discourse or prosodic function. In some other cases, however, the held sign may combine with the co-occurring sign to constitute a larger morphosyntactic unit. This paper discusses how we devise a consistent transcription system to capture and differentiate these different types of simultaneity for our Hong Kong Sign Language Child Language Corpus in a way that would facilitate not only the viewing of the glosses, but also the analysis of morphosyntactic complexities of deaf children's signing production.

## 1. Introduction

It is a well-known fact that sign languages are characterized with a wide range of simultaneous constructions that make use of the availability of two manual articulators to form complex polymorphemic constructions. This paper discusses the transcription system we develop for the Hong Kong Sign Language Child Language Corpus, with specific focus on how simultaneous constructions involving the two manual articulators are glossed. Our discussion will proceed as follows. In Section 2 we will briefly introduce the basic features of our acquisition corpus. Section 3 discusses the types of simultaneous constructions we attempt to code and differentiate in our corpus. Section 4 presents our transcription system. Section 5 is the concluding remarks.

## 2. Hong Kong Sign Language Child Language Corpus: A basic description

We are currently developing a Hong Kong Sign Language (hereafter HKSL) acquisition corpus in which the data are transcribed with ELAN (EUDICO Linguistic Annotator), the multimedia annotation tool developed by the Max Plank Institute of Psycholinguistics, Nijmegen, The Netherlands.<sup>1</sup> The corpus contains digitized video recordings and transcriptions of sign language production by deaf children acquiring HKSL and the signing adults who interacted with them. At this initial stage of development, the corpus includes two tiers of basic glosses, an utterance tier which mainly serves to mark sentence/utterance delimitations and a morphosyntactic

tier that contains information about the grammatical categories of the signs. The tiers for glosses and morphosyntactic information require manual input, whereas the utterance tier is basically generated via an interface program that can systematically and automatically combine information from the glossing and morphosyntactic tiers in a format transferable and readable in CLAN, the data analyzing programme in CHILDES.<sup>2</sup> The symbols and features we use in the transcription system are compatible with CHILDES in order to facilitate cross-platform sharing of the data once the corpus is completed.<sup>3,4</sup>

## 3. Simultaneous constructions involving two manual articulators

In a signing discourse, signs produced by the two manual articulators may exhibit a varying degree of relatedness or integration with respect to their semantic, morphological, or syntactic characteristics. In a two-handed lexical sign, the two hands form a single morphemic unit which cannot be further decomposed morphologically. In addition, signers may produce a lexical sign and a gesture at the same time. Signers may also simultaneously produce two lexical signs which are usually presented sequentially. For

<sup>1</sup> The first batch of our transcribed data will be released in CHILDES by the end of this year.

<sup>2</sup> The acronym CLAN stands for 'Computerized Language Analysis'. It is a program that is designed specifically by Leonid Spektor at Carnegie Mellon University to analyze data transcribed in CHAT, the format of the Child Language Data Exchange System (CHILDES).

<sup>3</sup> For example, symbols that stand for repetition and substitution in our data are adopted from the CHAT format of CHILDES.

<sup>4</sup> Details of the utterance tier will be given in another poster presentation from our colleagues.

instance, instead of signing *IX\_3 MALE* ('that man') (i.e. a pointing determiner followed by a lexical noun), a signer may produce these two one-handed signs at the same time. These two lexical signs are free morphemes in and of themselves, but are syntactically related as they combine to form a noun phrase. In a typical two-handed classifier construction, for example, *put+CL\_hand:cup+CL\_sass:table* [= a cup on a table], the handshape, movement, and location of each of the two hands bear a morphemic status and these morphemes are put together to form a larger morphosyntactic complex that represents a single, static event.<sup>5,6</sup> In a signing discourse, it is also not uncommon to see the whole or part of a completed sign to be held in space in one hand, while another sign is produced by the other hand. In some cases, the held sign may bear no morphosyntactic relation with the co-occurring sign and its presence only serves a discourse or prosodic function. In some other cases, however, the held sign may combine with the co-occurring sign to constitute a larger morphosyntactic unit. What complicates the picture further is that the held sign may remain dormant for some time, but become active again later in the discourse. These several types of simultaneity with respect to the two manual articulators show a varying degree of complexities at different linguistic levels, and such information are of great value when researchers probe into the sign language development of deaf children. In constructing a sign language acquisition corpus, we therefore deem it necessary to differentiate and code them explicitly in our transcription system.

#### 4. Representation of Simultaneity in the Hong Kong Sign Language Child Language Corpus

##### 4.1 The two glossing tiers for the two manual articulators

In the sign language literature, diverse labels have been adopted to name the glossing tiers that transcribe the linguistic information encoded by the two manual articulators, e.g. left-hand vs right-hand (e.g. Nyst, 2007; Anna-Lena Nilsson, 2007; Vermeerbergen and Demey, 2007), dominant-hand vs non-dominant hand (Leeson & Saeed, 2007), or main gloss vs non-dominant hand gloss (MacLaughlin, Neidle and Greenfield, 2000). In our corpus, however, we have decided to use 'gloss 1' and 'gloss 2' instead of these commonly-used labels due to the following reasons.

<sup>5</sup> In our corpus, classifier handshapes are divided into four types in the HKSL acquisition data, including (i) *CL\_sem* for semantic classifier handshapes; (ii) *CL\_sass* for size and shape classifiers; (iii) *CL\_hand* for handling classifiers; and (iv) *CL\_body* for both bodypart classifiers (i.e. handshape that stands for a body part) and body classifiers (i.e. the signer's body represents a referent's body).

<sup>6</sup> At this initial stage of data transcription, only verb roots and classifier handshapes of classifier predicates are coded explicitly. We plan to include other morphemic units, such as location and manner, in the future development of the corpus.

The division of left-hand and right-hand may be a good option for transcribing situations in which each of the two manual articulators produces independent morphological units, e.g. one-handed lexical sign or classifier predicate, but it cannot effectively label two-handed lexical signs. Researcher may need to set up a separate tier, e.g. both-hand, for coding two-handed lexical signs, or state the same gloss twice, one on the left-hand tier and the other the right-hand tier. The first option creates an extra tier in the transcription system, and this makes viewing of glosses difficult and inconvenient because the glosses would be scattered among three different tiers. Representing the same gloss twice is equally problematic, because this may mistakenly lead to an impression that the deaf child is producing two morphologically independent units and as such over-estimate a child's language development if quantitative analyses such as frequency count or MLU are conducted. Most importantly, except for a few signs (e.g. *LEFT* and *RIGHT*), handedness of a sign is usually linguistically insignificant, at least in HKSL. As our corpus aims at representing the grammatical development of deaf children rather than the phonetic interaction of the two manual articulators, we leave the left-hand/right-hand dichotomy to later research.

The dominant/non-dominant hand distinction, on the other hand, may be useful in representing the phonetic relation between the dominant and weak hand in a two-handed lexical sign, two-handed classifier constructions consisting of a figure (i.e. dominant hand) and a ground (i.e. non-dominant hand), or situations in which a sign is produced by the active signing hand (i.e. dominant) in the presence of the maintenance of a previously completed sign in the non-active hand (non-dominant). Yet this pair of labels cannot be used to transcribe classifier constructions in which both hands represent figures actively involved in the predicate, or in cases where both hands are independent morphemes which are of equal significance morphosyntactically, as in the simultaneous production of two lexical signs, *IX\_3* and *MALE* (i.e. that man).

In view of the above problems, and in order to encompass as many types of simultaneous phenomena as possible, we have decided to dispense with these commonly used labels and adopt 'gloss 1' and 'gloss 2' instead, which are theoretically more neutral.

##### 4.2. Use of the 'gloss 1' tier in the transcription system

In our HKSL acquisition corpus, two separate tiers – 'gloss 1' and 'gloss 2' ('g1' and 'g2' in short form) – are set up for each signing participant to gloss the meaning of individual signs. A lexical sign, if produced independently without any co-occurring constituent, will be coded on the 'gloss 1' tier. It is glossed with the English words bearing the closest possible meaning.<sup>7</sup> Classifier

<sup>7</sup> Note that additional symbols are adapted from the CHAT specification of CHILDES for coding grammatical properties

predicates signifying the motion or locative property of a single referent are also coded on the ‘gloss 1’ tier.<sup>8</sup> Apart from the meaning of entire predicate, the verb root and the classifier handshape are only marked explicitly. For example, a classifier predicate which means “a person walks forward” is glossed as *walk+CL\_sem [= a person walks forward]*. Gestures, if produced manually, are coded on the ‘gloss 1’ tier, too. For instance, a hand-waving gesture signers commonly use to call other people’s attention is glossed as *gesture [= get someone’s attention]*.<sup>9</sup> Note that glosses for gestures are in small letter to distinguish them from lexical signs. The meaning of the gestures and the classifier predicates are enclosed in square brackets containing an equality symbol ‘[= ]’. Whether these lexical signs, classifier predicates and gestures are one-handed or two handed, left-handed or right-handed, is not a matter of concern in the transcription.

#### 4.3 Use of the ‘gloss 2’ tier in the transcription system

The ‘gloss 2’ tier is only invoked when the two manual articulators produce signs which are morphologically independent from each other. As discussed in Section 3, there are several types of simultaneous constructions which are differentiated and coded in our transcription system. They will be discussed one by one in the following sub-sections.

##### 4.3.1 Simultaneous production of a lexical sign plus a gesture

The first type of simultaneous construction that invokes the use of the ‘gloss 2’ tier involves the production of a gesture plus a lexical sign, as in the following example:

Example (1): “It is ashamed for you to become angry.”

*BRE	<IX_2~gesture [= get someone's attention]> [% sim] ANGRY ASHAMED IX_2
%mor@BRE	n:pr[IX_2~n:pr]gesture adj ANGRY adj ASHAMED n:pr IX_2
%xg1@BRE	IX_2   ANGRY   ASHAMED   IX_2
%xg2@BRE	gesture [= get someone's attention]

Figure 1: Example for simultaneous articulation of a lexical sign plus a gesture

Note that the lexical sign and the gesture are not related morphologically and syntactically. On the utterance tier, they are separated by a tilde and are enclosed in angle brackets followed by ‘[% sim]’. This notation indicates that they are produced simultaneously but are morphosyntactically independent of each other.

##### 4.3.2 Simultaneous production of two lexical signs

The second possible type of simultaneous constructions involves two independent lexical signs produced simultaneously. The two lexical signs may or may not combine and form a larger syntactic constituent.

Example (2): “After being bitten (by the dog), (the cat) was frightened, in pain and (its body) bled.”

*CHI	BITE [*] gesture [= bite]<AFRAID~PAINFUL> [% sim] BLEED [*]
%mor@CHI	v:agr BITE ges gesture adj AFRAID adj PAINFUL n BLEED
%xg1@CHI	BITE [*]   AFRAID   BLEED [*]
%xg2@CHI	gesture [= bite]   PAINFUL

Figure 2: Example for simultaneous articulation of two lexical signs – CC 4;6;21<sup>10</sup>

In example (2) above, the child produces two lexical signs - *AFRAID* and *PAINFUL* - at the same time.<sup>11</sup> Although they are simultaneously produced, they represent two coordinated adjectival predicates that do not combine to form a larger syntactic constituent. On the utterance tier, these two signs are separated by a tilde and are enclosed in angle brackets followed by ‘[% sim]’. This notation indicates that the two lexical signs are produced simultaneously but are morphosyntactically independent of each other.

Example (3): “You just begin (to learn how to ride a bicycle). The bicycle will move along a zigzag path when you ride it on your own.”

specific to sign languages. Examples include agreement markings (e.g. *GIVE-IS&Sub* stands for the sign GIVE inflected for 1st person singular agreement) and spatial markings on verbs (e.g. *PUT-a*, *PUT-b*, *PUT-c* stands for three instances of the spatial verb PUT at location ‘a’, ‘b’ and ‘c’) and mouthing for spoken words.

<sup>8</sup> Occasionally a classifier predicate denoting a single referent may involve two hands. For instance, in *swim+CL\_body:jelly\_fish [= a jelly fish swims by moving its tentacles]*, the classifier for the jellyfish consists of a spread-5 handshape with flexed fingers representing the top, and another spread-5 handshape with laxly flexed fingers representing the tentacles. In cases like this, the classifier predicate is still given a single gloss on the ‘gloss 1’ tier.

<sup>9</sup> Gestures which are included in the transcription include those related to discourse information only, such as head nod indicating a reply. These gestures subjects to appear independently on the ‘gesture’ tier at the next stage of development.

<sup>10</sup> CC is the short form for the name of a longitudinal subject in the corpus.

<sup>11</sup> Note that *AFRAID* is placed on the ‘gloss 1’ tier because its onset time is earlier than that of *PAINFUL*. If two signs begin at the same time, they will be placed on the ‘gloss 1’ and ‘gloss 2’ tier by random.

*KEN	BEGIN <IX 2-SELF> [% sim] move+CL_sem [= a bicycle moves along a zigzag path] IX 2 SELF .
%mor@KEN	v:pl BEGIN n:pro IX 2-n:pro SELF cl +v move+c CL_sem n:pro IX 2 n:pro SELF .
%xg1@KEN	BEGIN (IX 2) move+CL_sem [= a bicycle moves along a zigzag path]   IX 2   SELF .
%xg2@KEN	(SELF)

Figure 3: Example for simultaneous articulation of two lexical signs

In example (3) above, the indexical sign is a determiner. It combines with the lexical noun *SELF* form a noun phrase. On the utterance tier, the two signs are joined together by a plus sign '+' to indicate that they are produced simultaneously and they combine to form a larger syntactic constituent.

### 4.3.3 Classifier predicates involving classifiers for two independent referents

The third situation invoking the 'gloss 2' tier is complex classifier predicates involving two classifiers for independent referents. The two classifiers may combine together to form a single event, as in example (4), or they represent separate but coordinated events, as in example (5):

Example (4): "Put the tea bag into the cup; pour water into the cup and (the water) changes to brown."

*BRE	TEA+CL_hand:tea_bag+CL_hand:cup [= put a tea bag into the cup] &[=CL_hand:cup_pour+
%mor@BRE	n TEA cl +v put+c CL_hand+cl CL_hand:cl +v pour+c CL_hand:cl CL_hand n WATER
%xg1@BRE	TEA  put+CL_hand:tea_bag [= put a tea bag into the cup]   pour+CL_hand:pot [= pour some water
%xg2@BRE	be_located+CL_hand:cup   CL_hand:cup
(cont'd)	*BRE CL_hand:pot+CL_hand:cup [= pour some water into the cup] WATER become_fall+CL_sass:
%mor@BRE	cl +v become_fall+c CL_sass+c CL_hand v:pl CHANGE 1 adj BROWN.
%xg1@BRE	into the cup   WATER   become_fall+CL_sass:water [= water in the cup becomes full]
%xg2@BRE	>   CL_hand:cup
(cont'd)	*BRE water+CL_hand:cup [= water in the cup becomes full] CHANGE 1-a BROWN &[=CL_hand:cup
%mor@BRE	
%xg1@BRE	CHANGE 1-a   BROWN.
%xg2@BRE	>

Figure 4: Example of two independent classifiers which combine together to form a single event

Example (5): "Many birds flew together with the plane."

*KEN	fly+CL_sem:plane &[=CL_sem:plane HAVE MANY BIRD <fly+CL_sem:plane- fly+
%mor@KEN	cl +v fly+c CL_sem v:pl HAVE quant MANY n BIRD
%xg1@KEN	fly+CL_sem:plane [= a plane fly in the sky]   >
%xg2@KEN	HAVE    MANY
*KEN	CL_sem:birds [= many bird fly together with the plane]> [% sim] &[=CL_sem:plane.
%mor@KEN	cl +v fly+c CL_sem-cl +v fly+c CL_sem
%xg1@KEN	fly+CL_sem:plane.
%xg2@KEN	BIRD   fly+CL_sem:birds [= many birds fly together with the plane].

Figure 5: Example for two independent referents forming a single event (from Tang et al. 2007)

In example (4), two handle classifiers are produced to stand for the tea-bag and the cup. They are listed as *put+CL\_hand:tea\_bag* [=put a tea bag into the cup] and *be\_located+CL-hand:cup* on the 'gloss 1' and 'gloss 2' tier respectively. On the utterance tier, these two glosses are linked up by '+' to indicate that they combine to form a complex classifier predicate.

In example (5), two classifier predicates – *fly+CL\_sem:plane* and *fly+CL\_sem:birds* [=many birds fly together with the plane] – are produced at the same time to represent two co-temporal events. As these two classifier predicates are structurally independent from each other, the two glosses are linked by a tilde and are enclosed in angle brackets followed by a comment '[% sim]'. In other words, their representation is the same as the simultaneous articulation of two independent lexical signs. They may be perceived as conjoined constructions.

### 4.3.4 A phonetic suspension of a completed sign in the presence of other morphosyntactically unrelated signs

The fourth type of simultaneity to be coded by the two glossing tiers involves the suspension of the handshape of a completed sign in one hand while the other hand continues to sign. In the literature, the phonetic maintenance of a completed sign is commonly marked by an arrow sign '>', but this symbol does not specify whether the held sign is morphosyntactically related to the co-occurring signs. As our transcription system aims at capturing the morphosyntactic complexities of the sign language production of deaf children, we restrict the use of '>' to a suspension of a completed sign which does not relate morphosyntactically to the co-occurring signs, as in example (6) below:

Example (6): "There is (a person wearing) a headscarf. There is a witch."

*CHI	HAVE [*] ONE tie+CL_hand [= a headscarf] HAVE WITCH HAVE .
%mor@CHI	v:pl HAVE num ONE cl +v tie+c CL_hand v:pl HAVE n WITCH v:pl HAVE .
%xg1@CHI	HAVE [*]   ONE   CL_hand [= a headscarf]   HAVE >   HAVE .
%xg2@CHI	WITCH

Figure 6: Example of phonetic suspension of a completed sign – CC 4;6;21

In the above example, *HAVE* is held by one hand while the other hand signs *WITCH*, which is morphosyntactically independent from *HAVE*. Note also that in the transcription the beginning of the gloss entry for '>' overlaps with the ending of *HAVE*. This kind of simultaneity is not specifically highlighted on the utterance tier.

### 4.3.5 A previously held sign being reactivated and combining morphosyntactically with the co-occurring sign

If a previously-held sign is reactivated again and

combines with another co-occurring sign to form a larger morphosyntactic unit, an entirely new gloss with semantic and/or syntactic content will be provided. An example is provided below to illustrate this situation:

Example (7): “The man (thief) is shot by the police and the bullet streaks towards him.”

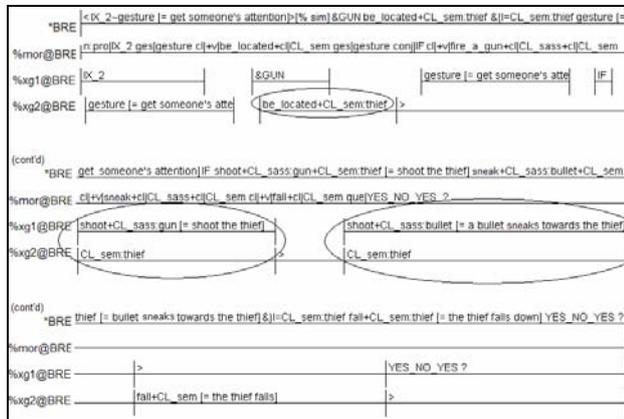


Figure 7: Example of a previously suspended sign being reactivated and combining morphosyntactically with the co-occurring sign

In example (7) above, the semantic classifier that stands for the thief (i.e. *be\_located+CL\_sem:thief*) is first held phonetically in the signing space while the signer continues to produce a gesture and the lexical sign *IF*. This kind of suspension is indicated by ‘>’ in the transcription. After that the same semantic classifier becomes an argument of the predicate ‘shoot the thief’ (i.e. *shoot+CL\_sass:gun*). As the semantic classifier is morphosyntactically active, it is glossed again in the transcription. The same classifier is also a component of the predicate “a bullet streaks towards the thief” (*shoot+CL\_sass:bullet*) and is therefore glossed once again. In order to show the articulatory continuity of *CL\_sem:thief*, the five consecutive glosses, namely, *be\_located+CL\_sem:thief*, >, *CL\_sem:thief*, > and *CL\_sem:thief*, are connected to each other without any separation. One advantage of this method of representation is that we can capture the fact that a sign, when being held in space, may perform different morphosyntactic functions in relation to other co-occurring constituents. Note further that when a held sign forms a morphosyntactic unit with a co-occurring sign, the two gloss entries will be time aligned. One more example is given below:

Example (8): “Put the tea bag into the cup; pour the water and it changes to brown.”

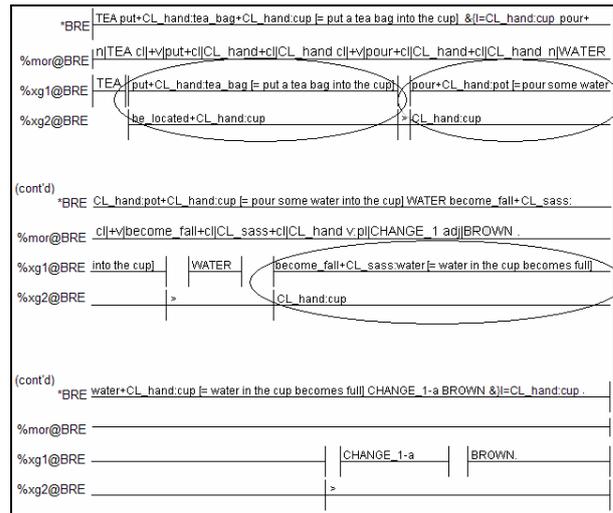


Figure 8: Example of a previously suspended sign being reactivated and combining morphosyntactically with the co-occurring sign

In the above example, the sign *CL\_hand:cup* is glossed again when it is morphosyntactically re-activated to be part of the predicate of “pour some water into the cup” and “water in the cup becomes full”. The six glosses on the ‘gloss 2’ tier are connected to each other in order to show the articulatory continuity of *CL\_hand:cup*.

In our transcription, if a certain sign is held in space and is reactivated some time later, two sets of symbols – ‘&{I=sign’ and ‘&}I=sign’ are used to delimit the scope of its phonetic persistence. In example (7), the semantic classifier for the thief (i.e. *CL\_sem:thief*) is held in space for a string of predicates. On the utterance tier, the first appearance of *be\_located+CL\_sem:thief* is followed by *&{I=CL\_sem:thief*, indicating that the classifier handshape is held in space. The holding of the semantic classifier ends before *fall+CL\_sem:thief*, which is preceded by *&}I=CL\_sem:thief* on the utterance tier.

### 5. Conclusion

Our transcription system can clearly capture and distinguish between different types of simultaneous constructions produced by the two manual articulators. Two glossing tiers are used whenever the signs produced by the two manual articulators form separate morpheme(s). If two co-occurring signs are syntactically related, that is, they combine to form a larger syntactic constituent, the two signs are linked up by a ‘+’ sign on the utterance tier. If the two signs only co-exist temporally without any morphosyntactic relation, they are enclosed in angle brackets on the utterance tier.

Note further that in our proposed glossing system, ‘>’ is restricted to suspension of a sign which does not interact morphosyntactically with other co-occurring signs. A new gloss is provided if a previously-held sign is reactivated in combination with other co-occurring signs to form a larger morphosyntactic unit. Such a coding system can

draw a distinction between spatially held signs with active morphosyntactic content and those whose maintenance in space only serve a discourse or prosodic function. This system can also capture the fact that a sign, when held in space, may perform different morphosyntactic or discourse roles depending on the type of co-occurring signs the held sign enter into a relationship with.

One major disadvantage of our proposed transcription system is that a sign which is held in space may be split up into several glosses. Although the articulatory continuity is still indicated by the timing connection of the gloss entries, researchers who are interested in how signs are held in discourse cannot rely on the search function of ELAN to extract the quantitative information on this phenomenon, e.g. how long is a sign held in space, how often are signs held in space, etc. This has to be done manually.

Another inadequacy of our current transcription system is that not all simultaneously presented morphemic units are coded explicitly at this stage of development. For example, the locative or manner morphemes are left unspecified. Hopefully these types of missing information will be coded as we continue to develop our corpus in the future.

## 6. Acknowledgements

The development of our acquisition corpus was initially supported by the Research Grant Council (RGC Grant #CUHK4278/01H; Direct Grant #2081006). Our corpus has been supported by Mr. Alex K. Yasumoto's donation since 2006. In addition, we would also like to thank the four deaf research assistants, Kenny Chu, Phippen Wong, Anita Yu and Brenda Yu, for data collection and transcription. Our thanks also go to Brian MacWhinney for his valuable advice in helping us devise a transcription system compatible with the CLAN format of the CHILDES corpus

## 7. References

- Leeson, L., Saeed, J. (2007). Conceptual blending and the windowing of attention in simultaneous constructions in Irish Sign Language. In M. Vermeerbergen, L. Leeson, & O. Crasborn (Eds) *Simultaneity in Sign Languages: Form and Function*. Amsterdam/Philadelphia: John Benjamins, pp.55--72.
- Nyst, V. (2007). Simultaneous construction in Adamorobe Sign Language (Ghana). In M. Vermeerbergen, L. Leeson & O. Crasborn (Eds) *Simultaneity in Sign Languages: Form and Function*. Amsterdam/Philadelphia: John Benjamins, pp.127--145.
- Nilsson, A.-L. (2007). The non-dominant hand in a Swedish Sign Language Discourse. In M. Vermeerbergen, Lor. Leeson & O. Crasborn (Eds) *Simultaneity in Sign Languages: Form and Function*. Amsterdam/Philadelphia: John Benjamins, pp.163--185.
- Tang, G. (2006) *A Linguistic Dictionary of Hong Kong Sign Language*. Hong Kong: Chinese University Press.
- Tang, G., Sze, F., Lam, S. (2007) Acquisition of simultaneous constructions by deaf children of Hong Kong Sign Language. In M. Vermeerbergen, L. Leeson & O. Crasborn (Eds) *Simultaneity in Sign Languages: Form and Function*. Amsterdam/Philadelphia: John Benjamins, pp. 283--316.
- Vermeerbergen, M., Demey, E. (2007). Comparing aspects of simultaneity in Flemish Sign Language to instances of concurrent speech and gesture. In M. Vermeerbergen, L. Leeson & O. Crasborn (Eds) *Simultaneity in Sign Languages: Form and Function*. Amsterdam/Philadelphia: John Benjamins, pp.257--282.
- MacLaughlin, D., Neidle, C., Greenfield, D. (2000). *Sign Stream TM User's Guide (Version 2.0)*. University of Boston.

# Annotation of Non Manual Gestures: Eyebrow movement description

Emilie Chételat-Pelé, Annelies Braffort & Jean Véronis

LIMSI/CNRS, Orsay, France

Provence University, Aix-en-Provence, France

Emilie.chetelat@limsi.fr, annelies.braffort@limsi.fr, Jean@Véronis.fr

## Abstract

Our study tackles Non Manual Gestures (NMGs) annotation within the context of Sign Language (SL) research and more particularly within the context of automatic generation of French Sign Language (LSF). Present descriptions need instantiation for the animation software. Thus, we propose a new annotation methodology, which allows us precise description of NMGs and which takes into account the dynamic aspect of LSF. On the video corpus, we position points on elements to be annotated, to obtain their coordinates. These coordinates are used to obtain precise position of all NMGs frame by frame. These data are used to evaluate the annotation by means of a synthetic face, for numerical analysis (by using curve), and, finally, to obtain numerical definition of each symbol of our annotation system based on arrows

## 1 Introduction

This paper deals with non manual gestures (NMGs) annotation involved in Sign Language (SL) within the context of automatic generation of SL. Many researches in SL emphasize the importance of NMGs at different language levels (lexical, syntactical, pragmatic...) and recognize that NMGs are essential for the message comprehension. However, the NMGs structure knowledge is limited. Our purpose is to refine the knowledge of NMGs structure and their roles. To acquire this knowledge, it is necessary to have precise NMG descriptions. These descriptions are obtained from the observation and annotation of a video corpus. Depending on the degree of precision we need, the first step is to conceptualize an annotation methodology. We suggest in this paper a methodology, which allows us a numerical annotation of NMGs for a precise description of NMGs structure. This study is based on French Sign Language (FSL) but can be used for another SL.

The next section presents the context of this study: the available descriptions and transcriptions of NMGs and the presentation of our purposes. In the third section, we suggest a new annotation methodology, which allows us to study the NMG movement dynamics.

## 2 Problematic

At present, descriptions of NMGs are symbolical. Transcription systems like HamNoSys (Prillwitz and Zienert, 1989), D'Sign (Jouison, 1995) or SignWriting (Sutton, Gleaves, 1995), describe the NMG posture with more or less iconical graphical forms (Figure 1: "eyebrows high" transcribes by different systems).

This type of description is not suitable for automatic generation systems because they do not contain numerical indication. Moreover these descriptions relate to a given instant and do not allow us to describe the movement intensity and dynamics. For example, for a description such as "Eyebrows high", we would like to know the movement intensity and the raising duration. Thus, these systems are not accurate enough to study the importance of these elements in the meaning transmission.

In this article, we suggest a new methodology applied to the eyebrow and eye movements. This allows us to study the NMG movements with the aim to provide precise descriptions of these movements. Describing NMGs precisely imply a rigorous annotation of the different NMG movements that can be observed on a video corpus. The methodology must provide the means to describe all the phenomena and the study of the NMG movement dynamics. The methodology has also to provide a formal definition of NMG structure.

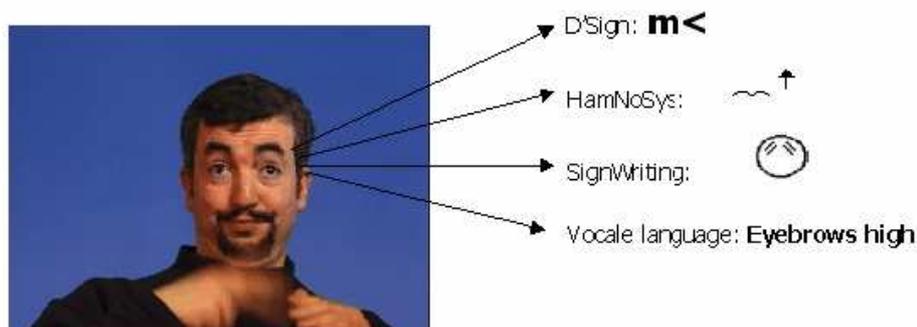


Figure 1: Many transcriptions of "eyebrows high".

### 3 Methodology

This part presents an application of our methodology on eyebrows and eyes movements. For annotation, we used LS-Colin corpus (Braffort et al, 2001; Segouat, Braffort, Martin, 2006). The video quality and the close-up shot are particularly precious for our study. Moreover, we used Anvil software because this software offers the possibility to annotate with personal icons and colors, which is of great help for a visual perception of phenomena.

Moreover, Anvil allows us to directly annotate on the video frames by means of points. Their coordinates can then be exported for further treatments.

The first section (3.1) presents how the video was annotated based on the FACS system (Facial Action Coding System). In a second section (3.2), we explain in detail the annotation data processing. Then, the last three sections (3.3, 3.4, and 3.5) present three data uses that permit to analyze and evaluate the annotation.

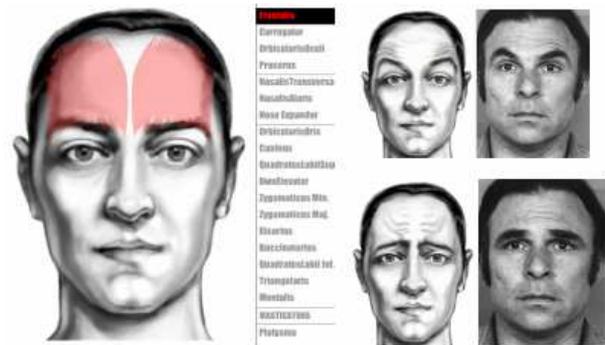


Figure 2: Frontal muscle frontal and the associated eyebrows AUs: outer extremity rise (AU1) and inner extremity (AU2). Pictures extracted from the Artnatomy<sup>1</sup> website (Contreras Flores, 2005) and the FACS manual<sup>2</sup>.

The corrugator supercillii muscle, the orbicularis oculi muscle and the procerus muscle allow lateral movement of the eyebrows, which is inducing a variation of the distance between the eyebrows (Figure 3).

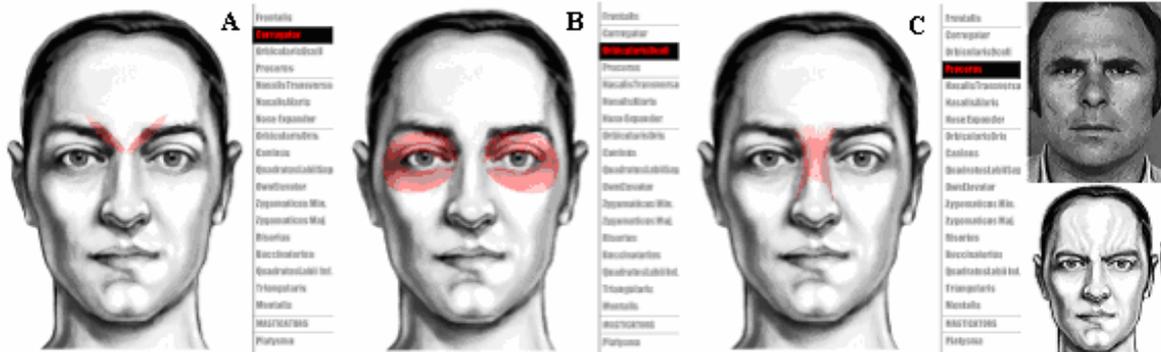


Figure 3: The corrugator supercillii muscle (picture A), the orbicularis oculi muscle (picture B) and the procerus muscle (picture C), responsables of the AU4. Pictures extract of Artnatomy (Contreras Flores, 2005) and FACS manual.

#### 3.1 Annotation on the videos

For the eyebrows movement description we use the FACS system, which has been designed for the description of emotion mimics. FACS is a description system of facial expression, which is based on facial muscles (Ekman and Friesen, 1978). Actually, Ekman and Friesen use these muscles as a base for the definition of all face movements. FACS measurement units are the Action Units (AUs), which represent the muscular activity that produces momentary changes in facial appearance.

For the eyebrows, Ekman and Friesen distinguish four muscles allowing three actions: rise of eyebrow inner (AU1), rise of eyebrow outer (AU2) and eyebrow lowering (AU4).

The frontal muscle (Figure 2) is responsible of the rise of the eyebrow inner and outer extremities.

The figure 4 shows three AUs combination: AUs 1 with 4 (inner rise and eyebrow lowering), AUs 1 with 2 (inner and outer rises), and AUs 1, 2 and 4 (inner and outer rises and eyebrow lowering).

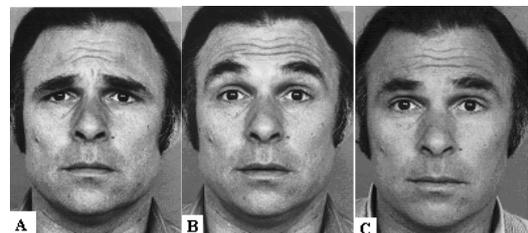


Figure 4 : Three AUs combinations. Picture A: AU1 + AU4 ; Picture B: AU1 + AU2 ; Picture C: AU1 + AU2 + AU4

<sup>1</sup> [www.artnatomia.net](http://www.artnatomia.net)

<sup>2</sup> <http://www.face-and-emotion.com/dataface/facs/manual/TitlePage.html>

Theses pictures show that the size of the eyebrow can change according to the AUs and their combinations. Moreover, the middle of the eyebrow rises with a bigger amplitude than its outer extremity, implying a more important perception of movement in this area.

FACS is a formal coding system is useful for facial expression description. However, it does not allow a description of dynamics (temporal analysis...). Then we only use FACS as a base, from which we have elaborate our own methodology.

For the eyebrows, FACS distinguishes two points (inner and outer extremities), which can move on horizontal and lateral axes. We retain these points for the video annotation. But because of its greater movement amplitude, we also consider the middle of the eyebrow and annotate it.

Moreover, to limit the annotation imprecision involved by the eyebrow thickness we double the extremity points (inner and outer) for each eyebrow and triple the middle point, the most difficult to accurately position.

Finally, to determinate the eyebrows movements independently of the head movement, we consider reference positions: the two extremities of each eye.

Thus, we position 18 points on each frame of the video (25 frames by second). The figures 5 and 6 show the location of each point.

After having annotated the whole video, we export the 2d coordinates x and y of each point. Calculations on theses coordinates give us precise data of the eyebrows movements.

### 3.2 Calculation on the 18 point coordinates

For the data processing, we used Scilab<sup>3</sup> software, free software for scientific calculation, which allows us, within a script, the automation of calculations. The input is coordinates of each point. These data are used for calculations to compute the position of each point independently of the head movement, frame by frame:

1. First, we calculate the average coordinates of the extremity and middle of each eyebrow for each frame (2-3, 7-8, 11-12, 16-18 for the extremities, and 4-5-6, 13-14-15 for the middles).
2. The news coordinates are used to calculate the distance (D) between these 3 points of each eyebrow to the extremity points of the eyes (for example, the distance between the point 1(x1,y1) and the average point 2-3 (x2,y2)):  $D = \sqrt{(x1 - x2)^2 + (y1 - y2)^2}$ .
3. We calculate the variation (V) of the position at the frame (n) by means of the Distance (D):  $V(n) = D(n) - D(n-1)$ . This variation can be positive (for a rise) or negative (for a lowering).
4. Then, the variation (V) allows us to calculate the position (P) of each element, independently of the head movements, for each frame of the video:  $P(n) = V(n) + P(n-1)$ .

These final data are used for the annotation evaluation and analysis.

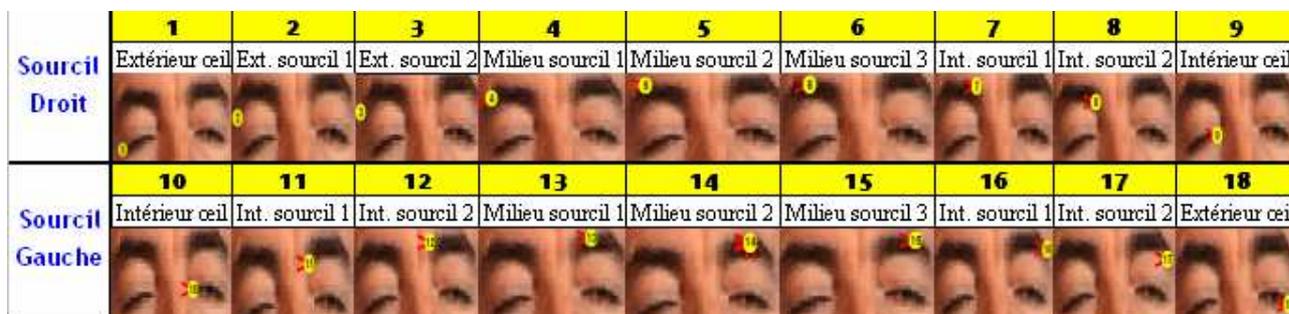


Figure 5: Site of 18 points



Figure 6: Corpus video extract with points.

### 3.3 Intermediate evaluation

These numerical data allow us to automatically generate the eyebrows animation on a synthetic face. For the generation, we used the Xface<sup>4</sup> software (Balci, 2006). Xface is a 3D talking head and was built for vocal production; not for SL production.

This automatic generation allows us to have a first qualitative evaluation of our annotation. We can compare the video and the Xface production simultaneously and evaluate if all phenomena are presents. Thus, we can adjust the annotation (for example put one more point) if necessary.

<sup>3</sup> <http://www.scilab.org/>  
<sup>4</sup> <http://xface.itc.it/>

In figures 7 and 8, the left picture is extracted from our corpus (LS-Colin corpus, Braffort et al, 2001; Segouat, Braffort, Martin, 2006), and the right one from some Xface productions generated from our annotation.

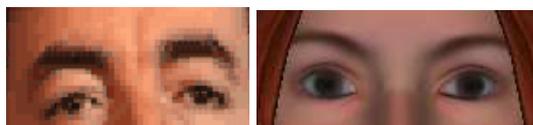


Figure 7: Standard position



Figure 8: higher eyelid and distance between eyebrows lowered

Moreover, playing the Xface production and the video at the same time allows us to evaluate the synthetic face. We have yet identified the limits of the Xface face model and we can propose ameliorations for the synthetic faces used for automatic generation of LSF. For example, we observe that Xface do not have wrinkle and does not provide enough amplitude for the movements of eyebrow and eyelid. These limits induce perception problems for deaf users because it is very difficult to determinate the of eyebrow position. Thus, we can establish a list of necessary elements for synthetic face to produce realistic LSF.

This first use of the data allows us qualitative evaluation of the methodology. Data are then used for NMGs analysis.

### 3.4 Structural analysis of NMGs

Numerical data allows us to analyze the movement structure. For example, the curve presented figure 9 informs us of the amplitude of the eyebrow inner point and allows us a classification of the rises.

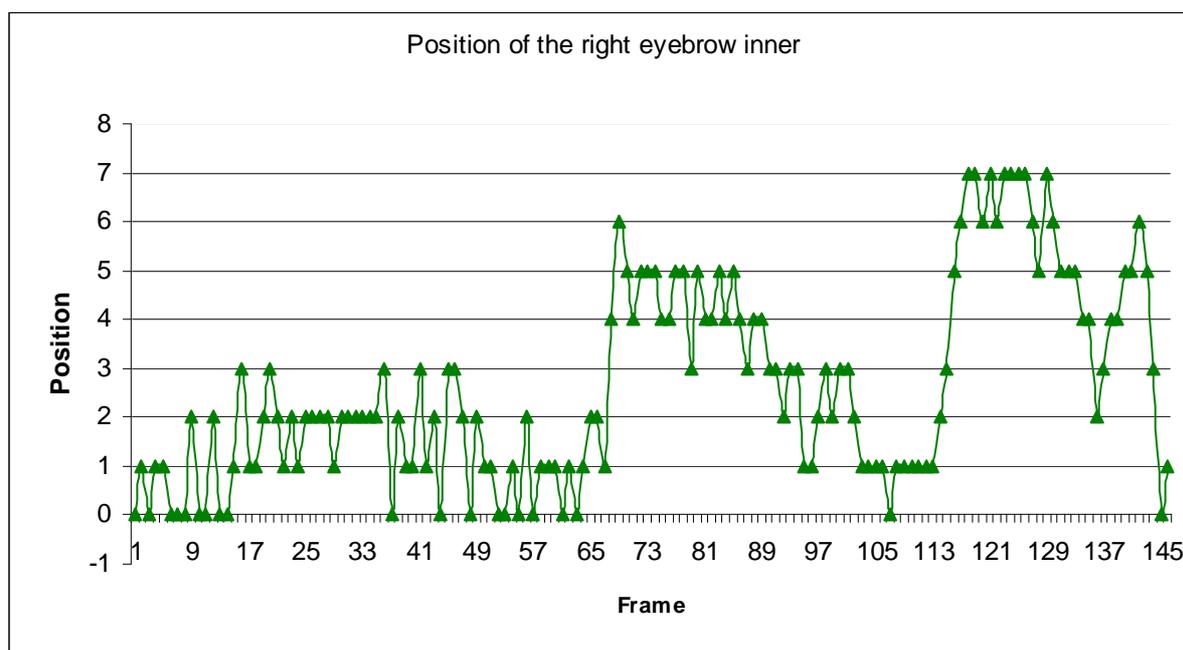


Figure 9: Position of the right eyebrow inner.

This curve shows three rise amplitudes for the eyebrow inner point: one small rise (1 unit for this person), one medium rise (2 units) and one high rise (3 units). These rises can be defined related to the small rise: a medium rise is two times higher than a small. A high rise is three times higher than a small rise. The precise numerical value of the rises amplitude can vary but the number of rise classes and their proportions are always the same. Then, a very high rise (7 units on the curve) is adopted in several steps: several rises of different degrees successively.

As show this example, the curves allow us to analyze the structure of the NMGs movements.

### 3.5 Formalization evaluation

These numerical data also allow us a validation and a numerical instantiation of the formal description based on arrows that we had presented in a previous paper (Chételat-Pelé, Braffort, Véronis, 2007). This system is based on four properties:

- Movement description (instead of posture description): For example: "eyelid lowering" instead of "low eyelid".
- Movement decomposition: For example, the diagonal movement of shoulders is described with horizontal movement and vertical movement separately;
- Element decomposition: For example, we separate higher eyelid and lower eyelid;
- The use of a set of symbols rather than words (Figure 10). One symbol can describe many phenomena (for example with use of colors for the movement intensity, figure 11).



Figure 10: Set of symbols used.



Figure 11: Different degrees of intensity

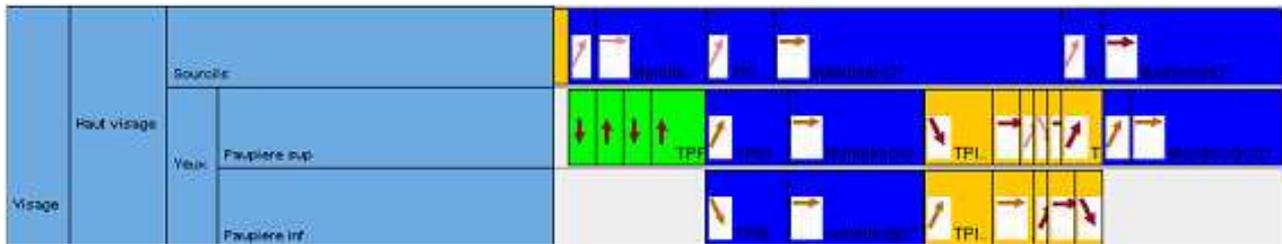


Figure 12: Annotation extract.

This description is simple and the use of the colors allows us to identify quickly the present phenomena (Figure 12). Our methodology allows us to define numerical values for each symbol. Moreover, we can automatically produce the annotation by means of the numerical data and validate our system. The numerical data have confirmed that there are three degrees of eyebrow movement (Figure 9). Applied on the whole arrow system we can determinate the pertinence of each symbol.

#### 4 Conclusion

This study takes place within the context of automatic generation of SL and aims at enhancing of the NMGs structure knowledge to ameliorate the animation capacities of automatic generation system.

We have presented, in this paper, a system allowing a accurate numerical description of some NMGs. This system is based on the annotation of each video frame. Moreover, it allows us to obtain precise positions of the eyebrows, independently of the head movements.

The annotation will be extended on other video to validate our first observations. Moreover, the synthetic face evaluation will be extended to identify the properties that the faces have to respect to produce precise and understanding LSF.

#### Références

- Balci K. (2004). MPEG-4 based open source toolkit for 3d facial animation. In AVI04, working Conference on Advanced Visual Interfaces. Gallipoli, Italie, 25-28 Mai 2004.
- Braffort A., Choisier A., Collet C., Cuxac C., Dalle P., Fusellier I., Gherbi R., Jausions G., Jirou G., Lejeune F., Lenseigne B., Monteillard N., Risler A., Sallandre M.-A. (2001). Projet LS-COLIN. Quel outil de notation pour quelle analyse de la LS ?. In Journées Recherches sur la langue des signes. UTM, Le Mirail, Toulouse.
- Chételat-Pelé E., Braffort A., Véronis J. (2007). Mise en place d'une méthodologie pour l'annotation des gestes non manuels. In TALS 2007. Traitement Automatique des Langues des Signes 2007 : atelier de Traitement Automatique des Langues Naturelles 2007. Toulouse.
- Chételat-Pelé E., Braffort A., Véronis J. (2008). Signs Language Corpus Annotation: Toward a New Methodology. In LREC 2008. Conference on Language resources and Evaluation. Marrakech. (A paraître).
- Victoria Flores V. (2005). Artnatomy/Artnatomia. url ([www.artnatomia.net](http://www.artnatomia.net)). Spain.
- Ekman P., Friesen W. V. (1978). Facial Action Coding System (FACS). Manuel Palo Alto : Consulting Psychologists Press.
- Jouison P. (1995). Ecrits sur la Langue des Signes française. Garcia, B. (éd). Paris : L'Harmattan, Paris
- Kipp M. (2004). Gesture Generation by Imitation - From Human Behavior to Computer Character Animation" Boca Raton Florida: Dissertation.com
- Prillwitz S., Zienert H. (1989). Hamburg Notation System for Sign Language: Development of a sign writing with computer application. Allemagne : S. Prillwitz & T. Vollhaber (Eds.): Current trends in European Sign Language Research Signum.
- Segouat J., Braffort A., Martin E. (2006). Sign language corpus analysis: synchronisation of linguistic annotation and numerical data. LREC 2006. Fifth International Conference on Language Resources and Evaluation Genoa Italy : 2006
- Sutton V., Gleaves R. (1995). SignWriter – The world's first sign language processor. Deaf Action Comittee for SignWriting, La Jolla, CA.

## Open access to sign language corpora

Onno Crasborn

Department of Linguistics, Radboud University Nijmegen

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

E-mail: o.crasborn@let.ru.nl

### Abstract

This paper sketches recent developments in internet publishing and related copyright issues, and explores how these apply to sign language corpora. As a case study, the *Corpus NGT* project is characterised, which publishes a systematic collection of sign language video recordings and annotations online as open access data. It uses Creative Commons licenses to make explicit the restricted copyright rules that apply to it.

### 1. Background

While native intuitions of Deaf informants have played some role in the linguistic study of signed languages, linguistic studies since Tervoort (1953) and Stokoe (1960) have mostly used film and video recordings. Descriptions and transcriptions of these video recordings were made on paper until the 1980s; since then, the transcriptions were increasingly made in office software like word processors, spreadsheets and databases. It was not until the 1990s that digital video became commonplace, and only since around the year 2000 it has become easy to process and store large amounts of video recordings in desktop computers. Only since the venue of multimedia annotation tools like SignStream, Transana and ELAN sign language researchers can use a direct link between their transcriptions and video recordings.

This paper will not go into the technical aspects of these developments, but aims to describe the ongoing shift in accessibility to sign language data by researchers. Many sign language researchers and research groups used to have shelves full of video tapes, but were not able to use the data very often after an initial transcription or analysis was made, simply because of the extremely time-consuming process of locating a specific point in time on a video tape, let alone comparing different signers on different tapes. With the use of modern technology, a direct link can be established between an instance of a transcription value and a time segment in a particular video file, and data that are already transcribed can easily be double-checked or shown to colleagues. This is commonly seen as leading to a potential increase in quality of one's own research.

We are currently at the brink of a next step in our use of sign language data, as data can be exchanged over internet and even published online. In this way, it can become easier to also check data used for linguistic publications by other investigators; access to not only the linguistic

analysis but also the data at the base of that analysis could lead to a further increase in reliability of linguistic research. This may appear to be obvious, as linguistic analysis typically do include written examples of the data under discussion for languages like English or Spanish, or phonetically transcribed examples of unwritten languages. The situation is a bit different for signed languages, where there is no conventional writing system that in use throughout deaf communities, and moreover, there is very little standardisation on the transcription of sign language data, whether for gloss annotations or for phonetic transcriptions. This holds both for manual and non-manual activity.<sup>1</sup> Access to the original data therefore has a relatively large value in evaluating linguistic claims.

Aside from the technological difficulty in creating digital video files, there are privacy issues related to the publication of video material, as not only what is said can be accessed, but also (and more unequivocally so than with audio data of speakers) what the identity of the speaker or signer is. This paper describes the ongoing developments in the publication of data on internet (section 2), and then discusses the nature and role of privacy protection of online publications of video recordings (section 3). Section 4 characterises one particular system of user licenses that is being developed especially for online publications, 'Creative Commons'. As a case study, section 5 discusses the construction and open access publication of the *Corpus NGT*, a linguistic corpus of video recordings of Sign Language of the

---

<sup>1</sup> SignWriting (<http://www.signwriting.org>) is used as a writing system in parts of some deaf communities and HamNoSys (<http://www.sign-lang.uni-hamburg.de/hamnosys>) and FACS ([http://face-and-emotion.com/dataface/facs/new\\_version.jsp](http://face-and-emotion.com/dataface/facs/new_version.jsp)) are stable phonetic annotation systems, but as yet, none of them is actually used by a substantial part of the research community.

Netherlands, which makes use of Creative Commons licenses to protect the data from undesired types of use.

## 2. Internet publishing developments

The publication of speech resources for spoken language research is quite common, and text data have been an object of study since the earliest stage of computer technology. There are now several organisations that offer online speech resources and associated tools for sale, including the Linguistic Data Consortium (LDC)<sup>2</sup> and the Evaluations and Language resources Distribution Agency (ELDA)<sup>3</sup>. Increasingly, spoken language data are also recorded and published on video, to be able to study non-verbal behaviour of speakers in addition to speech. The organisations above typically sell copies of data sets to researchers, rather than simply publishing them on a server for everyone to access for free. The intent is not necessarily to make profit from these sales; sometimes, the goal is merely to cover the costs that are made in creating hardcopies of data and manuals and sending them to someone.

One of the current developments on internet more generally is the increasing attention for ‘open content’: data of all kinds, whether text, images or video, are made publicly available, without charging a fee. While there may be restrictions on the type of use that is allowed, selling content and strictly protecting it under copyright laws appears not desirable necessary for some types of content. For example, many (starting) artists benefit from the wide distribution of their creative output without wanting to sell specific instances of works of art. For new art forms that crucially depend on computer access, including some multimedia productions, free internet access is a crucial component of their work. In addition to audiovisual and graphic arts, text distribution can also profit from open access even though traditionally, essays would be published in journals or books that could only be obtained by purchasing them.

Traditional publications of reproducible work in hardcopy, whether on paper, CD or DVD, or any other medium, would typically be accompanied by a message stating that “all rights are reserved”. When computer technology made the copying of for example music purchased on a CD easier, this statement did not so much apply to the unauthorised copying of parts of a text in another text, but to creating actual copies of the material. The venue of digital information distribution over internet was accompanied by new means of protection, referred to as ‘digital rights management’ (DRM).

<sup>2</sup> <http://www ldc.upenn.edu>

<sup>3</sup> <http://www.elda.org>

By contrast to these commercial publications, there are now many publications on internet where the explicit goal of the author is not to prohibit copying and usage, but rather to encourage use by others. This development is sometimes characterised as a change from ‘copyright’ to ‘copyleft’: rather than stating that “all rights are prohibited”, people are encouraged to use materials for their own benefit.

The same change in perspective can also be witnessed in science. Rather than being protective of one’s own data, it is becoming more and more common to publish research data, hoping that others will profit from it and do the same with their own data. The European Research Council, founded in 2006, explicitly encourages open access to research data, noting that while hundreds of repositories exist for the medical and natural sciences, the humanities are in a different position:

“With few exceptions, the social sciences & humanities (SSH) do not yet have the benefit of public central repositories for their recent journal publications. The importance of open access to primary data, old manuscripts, collections and archives is even more acute for SSH. In the social sciences many primary or secondary data, such as social survey data and statistical data, exist in the public domain, but usually at national level. In the case of the humanities, open access to primary sources (such as archives, manuscripts and collections) is often hindered by private (or even public or nation-state) ownership which permits access either on a highly selective basis or not at all.” (ERC, 2006)

‘Open access’ does not necessarily imply that no restrictions apply, nor that anyone can view materials without registration or subscription; thus, in the area of science, archive access may well be restricted to people who register as researchers or who work at research institutes. The Creative Commons licenses discussed in section 4 constitute one way of restricting the use of materials, but imply no assumption on whether one needs to register to use the materials.

## 3. Ethical concerns in the publication of sign language data

As was already indicated above, the publication of sign language data on video implies inevitably that the message content can be connected to the identity of the signer. Even without explicitly adding the name or other details of the signer’s identity to the video clip in metadata, people can easily be identified on the basis of their face. The chance that this will happen as well as its potential consequences are relatively large given the small size of Deaf communities in most countries. For

example, in the case of the Auslan corpus that is currently being constructed at Macquarie University, Sydney, the 100 people in the corpus form 1.7% of the Australian Deaf community, estimated to be about 6,000 (Johnston 2004).<sup>4</sup>

The open access publication of a sign language corpus implies providing information on who is and who is not recorded for scientific data, which in such a small community can be a sensitive matter in itself. The wide range of possible uses of a corpus of a substantial subset of signers might also have an influence of the language, the signing in the corpus being considered a standard of some form, or the signers being considered role models for second language learners. These type of issues will not be further discussed here, but they are considered as meriting further attention in any corpus construction project and any publication of sign language data.

The recording of signers for any linguistic research typically does not involve special ethical reviews for dealing with human subjects, which are common in (international) grant applications: there is no risk of (physical or psychological) harm to the signer, participation is voluntary and signers typically receive payment for their contribution, they just need to be treated with respect. Moreover, people typically sign a form to give the researcher a proof of their ‘informed consent’, which means that (1) the person has the legal capacity to give consent (so that parents should give consent for participation of their children), (2) the person gives consent on a voluntary basis, not being pressured to participate, and (3) the person is able to make an informed decision. It is exactly this last point that warrants some further attention.

Firstly, depending on the type of data that are being recorded and published, a lot of personal information can be revealed in discussions and conversations. While it is attractive to use free conversation data as instances of spontaneous language use, the risk of including personal information (whether about oneself or about others) increases, and it is not always possible to monitor this before publication of the material, neither by the signer nor by the researcher.

A document guidelines for research ethics of linguistic studies from McGill University (Canada) characterises most linguistic data collection as being ‘low-risk’ in the sense that “the information being collected is not of a sensitive or potentially private nature, i.e. people would not reasonably be embarrassed by other people knowing about it” (McGill 2008). The problem with online publication of sign language videos is thus that the nature of the data cannot always be well established, but moreover, that publication on internet cannot be undone. While a publisher can in

principle try to withdraw a publication by finding back all copies of books or CDs, this is virtually impossible with electronic open access material once it has been downloaded or re-distributed by others.

The irrevocable nature of the publication of sign language video data could also become a problem when signers decide in the future to withdraw their participation. Although the consent form has given the researcher the legal right to publish the material, for a good relation with the participant and the Deaf community in more general terms, it could be desirable to indeed withdraw items from a corpus that is already published.

Secondly, it is debatable whether anyone can make an informed decision on publication of video recordings on internet given the high speed of the development of computer technology. As publication entails possible availability forever, new technologies can imply uses of the video data that we cannot yet foresee. Although one can decide to not use names or initials in any of the metadata accompanying videos (as was done in the Corpus NGT, see section 5), if face recognition software should become available as part of the average desktop operating system and when automatic sign recognition technology allows translation of signed discussions to text (in whatever language), discussion content and identity can easily be matched and linked to further information on individuals that is available online. Thus, even though at present signers may be perfectly happy with the publication of video recordings, it is not unlikely that this will change in the future.

On the other hand, we currently also see a rapid change in what is considered as privacy-sensitive information now that people massively publish their own materials online. Aside from discussions in message boards and mailing lists, many people do not hesitate to publish large sets of family pictures online, and community web sites like Facebook<sup>5</sup> or Hyves<sup>6</sup> elicit wide participation from people who appear to be eager to share a lot of personal information with the whole world.

The question remains whether this is a sign of a permanent change in (Western) culture, or whether people will be dissatisfied with it in ten or twenty years time. Where people voluntarily take part in the publication of personal information about themselves, one might expect that this is not so much an issue, although one may still debate whether anyone can estimate the impact of exposing details of one’s private life online. However, in the case of sign language corpus construction and open access publication, the decision to publish something online is very

<sup>4</sup> <http://www.ling.mq.edu.au/centres/sling/research.htm>

<sup>5</sup> <http://www.facebook.com>

<sup>6</sup> <http://www.hyves.net>

indirect: it is not a concrete activity of a signer at his own computer, but the signing that was recorded was not inspected by the signers, and was only published online a few months after the event. It will remain important to monitor and discuss these developments in the future.

#### 4. Creative Commons licenses

Although copyright law cannot completely prevent abuse of published material, it can encourage people to treat materials with respect. Creative Commons is a recent initiative that explicitly aims to allow publishers of online material to apply some restrictions to the (re)use of online content, by declaring the applicability of a license with one or more conditions to a specific work that is published online. The international organisation Creative Commons was founded in 2001 as a bridge between national copyright laws and open content material on internet. All licenses have been translated to the national languages of more than thirty countries and have been adapted where necessary to national copyright laws in these countries, yet they all seek to stay as close as possible to the US originals to ensure that the licenses will be regarded as an international standard.

There are currently three types of restrictions, and some new developments are underway. The first restriction that can be applied is dubbed “BY”, and requires the user to refer to the original author of the work when re-publishing or using the work. The second restriction concerns the prohibition of commercial use of the work, and is dubbed “NC” (no commercial use). The third restriction concerns the modification of the work, and states that the work has to be reproduced in the same form (“ND”, no derivative works) or that modifications are allowed but have to be shared under the same conditions (“SA”, share alike).

The Creative Commons licenses are available in various forms: a plain language statement (as in the previous sentences), a formal legal text, and a machine-readable version for use by software. Reference to the licenses on internet is typically done by including an images with symbols for the different license conditions, some of which are illustrated in Figure 1. The image then links to the text of the actual license, or explicit reference to the URL of the license text can be included.

A large advantage of using these licenses is that creators of any type of work can publish materials themselves, and enter in an agreement with the user about the types of use that are allowed. Traditionally, various types of publishers acquired the rights for distribution, promotion, sales, et cetera, and these publishers then entered into agreements with the end users (here too, the term ‘license’ was sometimes used). Thus, using the

Creative Commons licenses, creators can retain more responsibility over what happens to their material, and at the same time profit from the relatively cheap production and distribution channels that are now offered on internet. All rights remain with the creator of a work.

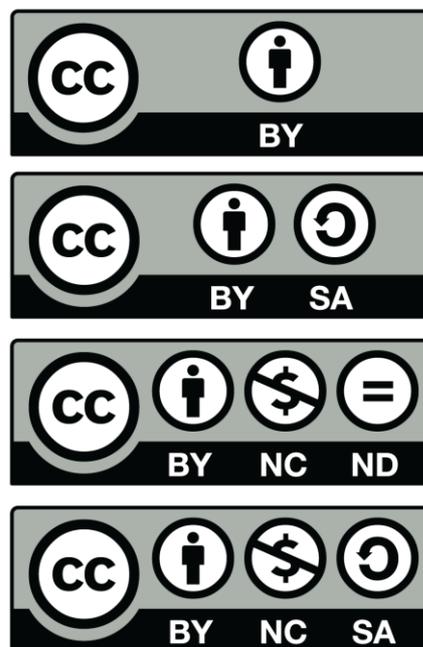


Figure 1. Examples of Creative Commons license buttons

#### 5. A case study: the *Corpus NGT*

The Creative Commons licenses form a very attractive way of protecting the use of the sign language videos in the *Corpus NGT*, a sign language corpus of Sign Language of the Netherlands (Nederlandse Gebarentaal, NGT; Crasborn & Zwitserlood, this volume).

For this corpus, a total of 100 signers will be recorded; most of these will be available in the first release in May 2008. These signers produced around 75 hours of interactive language material, divided in more than 2,000 video segments. The wish to publish this material not only for research purposes (its primary goal, cf. the funding from the Netherlands Organisation for Scientific Research) stems from its large possible value for various parties in the Netherlands: deaf signers themselves, second language learners of sign language, interpreting students, etc.

As was discussed above, a central problem in publishing sign language data online is privacy protection. In the *Corpus NGT*, we try to protect the privacy of the informants in several ways: we urge people to not reveal too much personal information about themselves or about others in

their stories and discussions, we limit the amount of metadata that we publish online (leaving out many of the standard fields from the IMDI metadata standard), and nowhere we mention or refer to the name or the initials of the signers. Personal information about family background and signing experience that we did collect will in principle be made available for other researchers, who will have to sign a license form declaring not to publish information on individuals. The nature of this license is not yet established, but we might consider copying such agreements from endangered languages documentation projects such as DOBES.<sup>7</sup>

We chose to apply the Creative Commons 'BY-NC-SA' license to all of the movie files in the *Corpus NGT* (symbolised by the last image in Figure 1). This license states that people may reuse the material provided they refer to the authors, that no commercial use be made, and that (modifications of) the material are distributed under the same conditions. As opposed to the 'no derivative works' condition, the latter condition allows users to use segments of clips for their own web sites, to add subtitling or other graphics to it, et cetera. While these types of modification will not frequently be interesting to scientific users, they do broaden the possible educational uses of the material.

Although the permission for the licensed open access publication is requested of the signers in the corpus, it was discussed above that we can not guarantee that signers can foresee the consequences at the time of recording. Will future technologies allow easy face recognition on the basis of movies and thereby obliterate the privacy protection measures that have been taken? What will the (normative) effect of publishing signing of a group of 100 signers from a small community be? There is a clear risk in the publication of sign language data without an answer to these questions. The 'solution' taken in the *Corpus NGT* project is to invest substantial time and energy in publicity within the deaf community, to explain the goal and nature of the corpus online, and to encourage use by deaf people.

The plain language version of the licenses is attached to every movie in the *Corpus NGT* by a short text preceding and following every movie file, thus allowing relatively easy replacement should future changes in policy require so (Figure 2). We expect to offer a signed version of the licenses in the near future as well.

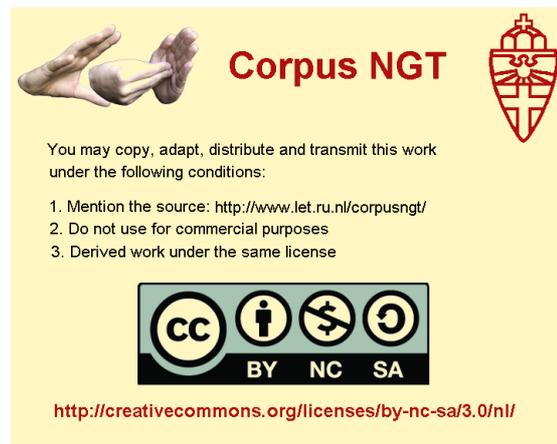


Figure 2. Reference to Creative Commons licenses in the *Corpus NGT* movies

## 6. Conclusion

The possibilities offered by current internet and video technologies together with new forms of licensing agreements offer attractive possibilities for the archiving of sign language research material, at the same time offering access to these materials for the language community itself and other interested public parties. This paper has tried to emphasise that the possibilities also raise new ethical issues that should receive attention at the same time. The traditional research ethics of informed consent and respecting ones informants will not be sufficient for internet publishing. The recently founded Sign Language Linguistics Society,<sup>8</sup> which is currently setting up a code of conduct for sign language research, might play a role in the discussion of these developments.

## Acknowledgements

This paper was written as part of the *Corpus NGT* project and is supported by grant 380-70-008 from the Netherlands Organisation for Scientific Research (NWO).

## References

- ERC (2006) ERC Scientific Council Statement on Open Access. <http://erc.europa.eu/pdf/open-access.pdf>. Document accessed in March 2008.
- Crasborn, O. & Zwitserlood, I. (this volume)
- Johnston, T. (2004) W(h)ither the deaf community? Population, genetics and the future of Auslan (Australian Sign Language). *American Annals of the Deaf*, 148(5), pp. 358-375.
- McGill University (2008) Department of Linguistics Procedures for Ethical Review of Research on Human Subjects.

<sup>7</sup> <http://www.mpi.nl/DOBES>

<sup>8</sup> <http://www.slls.eu>

[http://www.mcgill.ca/files/linguistics/research\\_ethics.pdf](http://www.mcgill.ca/files/linguistics/research_ethics.pdf). Document accessed in March 2008.

Stokoe, W. (1960) Sign language structure. An outline of the visual communication systems of the American Deaf (1993 Reprint ed.). Silver Spring, MD: Linstok Press.

Tervoort, B. (1953) Structurele analyse van visueel taalgebruik binnen een groep dove kinderen. Amsterdam: Noord-Hollandsche Uitgevers Maatschappij.

## Enhanced ELAN functionality for sign language corpora

Onno Crasborn, Han Sloetjes

Department of Linguistics, Radboud University Nijmegen  
 PO Box 9103, NL-6500 HD Nijmegen, The Netherlands  
 Max Planck Institute for Psycholinguistics  
 PO Box 310, 6500 AH Nijmegen, The Netherlands  
 E-mail: o.crasborn@let.ru.nl, han.sloetjes@mpi.nl

### Abstract

The multimedia annotation tool ELAN was enhanced within the *Corpus NGT* project by a number of new and improved functions. Most of these functions were not specific to working with sign language video data, and can readily be used for other annotation purposes as well. Their direct utility for working with large amounts of annotation files during the development and use of the *Corpus NGT* project is what unites the various functions, which are described in this paper. In addition, we aim to characterise future developments that will be needed in order to work efficiently with larger amounts of annotation files, for which a closer integration with the use and display of metadata is foreseen.

### 1. The *Corpus NGT* project<sup>1</sup>

#### 1.1 General characterisation

The *Corpus NGT* that was published in May 2008 is one of the first large corpora of (semi)spontaneous sign language use in the world, and the first to become publicly available online. It is targeted primarily at linguistic researchers, but due to its open access policy can also be used for other purposes, whether scientific, educational, or private. The corpus consists a large collection of sign language video recordings with linguistic annotations and audio translations in Dutch. Recordings were made of nearly 100 signers communicating in pairs. This resulted in 2,000 segments totaling 75 hours. The use of multiple cameras for four different angles resulted in a collection of  $\pm 15,000$  media files.

The four different angles can be displayed in sync by the ELAN annotation tool; for this purpose, an annotation file was created for every time segment. These documents were created from a template containing multiple (empty) tiers for glosses, translations and remarks. Over 160 files were actually annotated with gloss annotations on four different tiers, one for each hand of each of the two signers. In total, over 64,000 gloss annotations were added to these files. As two-handed lexical items receive a separate gloss for the left and for the right hand (each with their own alignment), the number of annotations cannot be blindly equated with the number of signs.

Further technical and linguistic information on the *Corpus NGT* can be found in Crasborn & Zwitserlood (this volume) and Crasborn (this volume), as well as on the corpus web site: [www.let.ru.nl/corpusngt/](http://www.let.ru.nl/corpusngt/). The corpus is currently hosted at the corpus server of the Max Planck Institute for Psycholinguistics, and part of their Browsable Corpus.<sup>2</sup>

#### 1.2 Use of standards and tools

<sup>1</sup> The *Corpus NGT* project was made possible by an investment grant from the Netherlands Organisation for Scientific Research (NWO), grant no. 380-70-008.

<sup>2</sup> <http://corpus1.mpi.nl>

The *Corpus NGT* makes use of open standards for its publication, aiming to guarantee long-term availability:

- Media files conform to the various MPEG standards (MPEG-1, MPEG-2, MPEG-4), rather than popular commercial formats such as Adobe Flash video.
- Metadata descriptions are made conforming to the IMDI scheme (Wittenburg, Broeder & Sloman, 2000; IMDI Team, 2003).<sup>3</sup> While this format may not be used in ten years time, its widespread use in linguistics and the publication of the whole corpus as part of a larger set of IMDI corpora at the Max Planck Institute for Psycholinguistics ensures that the corpus will be part of larger conversion efforts to conform to future standards.
- The annotation files were all created with ELAN and thus conform to the specification for EAF files (Brugman & Russell 2004).<sup>4</sup>

### 2. Developments in the ELAN software

The *Corpus NGT* project involved annotating many hours of video and a large number of annotation documents. The first aim of the technological goal of software improvement in the *Corpus NGT* project was to ease annotation. A second aim was to facilitate the use of annotation documents, in its widest sense: browsing, searching, and data analysis.

The functions described in this section appeared in a series of releases between versions 2.6 and 3.4. Specifications were set up by the *Corpus NGT* project and the ELAN developers. For guidelines on how to use the functions, including the location in menus and keyboard shortcuts, we refer to the ELAN manual.<sup>5</sup>

#### 2.1 Extension of the EAF specification and a change in the preferences format

- The property ‘annotator’ has been added in the specification of tiers, allowing groups of researchers to separate which tier has been filled by whom. It is expected that this property will become a selection criterion in the

<sup>3</sup> [http://www.mpi.nl/IMDI/schemas/xsd/IMDI\\_3.0.xsd](http://www.mpi.nl/IMDI/schemas/xsd/IMDI_3.0.xsd)

<sup>4</sup> <http://www.mpi.nl/tools/elan/EAFv2.5.xsd>

<sup>5</sup> <http://www.lat-mpi.eu/tools/elan/manual/>

search mechanism in a future release of ELAN.

- Preferences are no longer stored in binary .pfs files, but in user-readable XML files. The current preferences settings can be exported to a file and imported in (i.e. applied to) any other document; in this way, the ordinary user without knowledge of XML can also copy settings from one document to another. In this way, it has become easy to homogenise the layout of larger sets of ELAN documents and modify this ‘style sheet’.

## 2.2 New functionality

- The ‘**duplicate annotation**’ function was created to facilitate the glossing of two-handed signs in cases where there are separate tiers for the left and the right hand: copying an annotation to another tier saves annotators quite some time, and prevents misspellings. A disadvantage of using this function turned out to be that

signs. While the hands often do not start and end their articulation of a sign at the same time, the ‘duplicate annotation’ function makes it attractive to classify a sign as a phonologically two-handed form, even though the phonetic appearance can show differences between the two hands. Moreover, larger timing differences between the two hands have shown to play a role in many levels of the grammar of signed languages beyond the lexicon (Vermeerbergen, Leeson & Crasborn 2007). It will depend on the user’s research goal whether or not detailed timing differences are important to annotate correctly.

In addition to this quick annotation duplication shortcut some more generic copy and paste actions have been added. An annotation can be copied to the clipboard either as single annotation or as a group with all ‘dependent’ annotations. Pasting of an annotation or a group of annotations is not restricted to the same time segment (i.e.

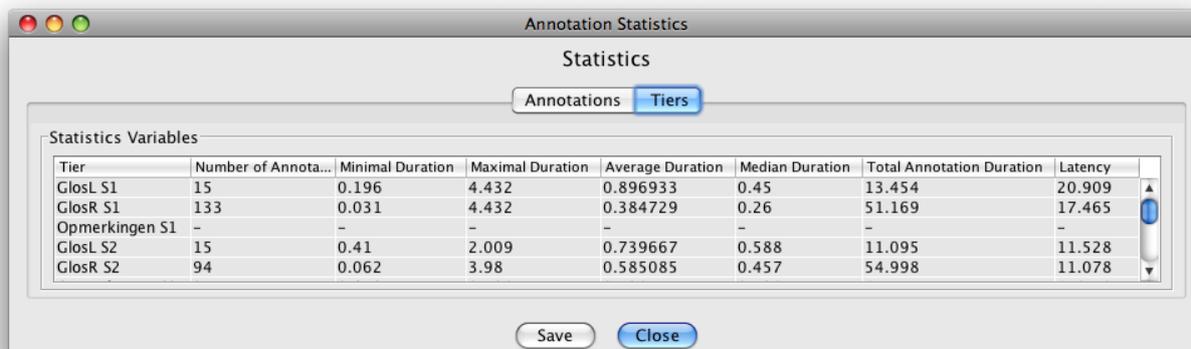


Figure 1. Tier statistics

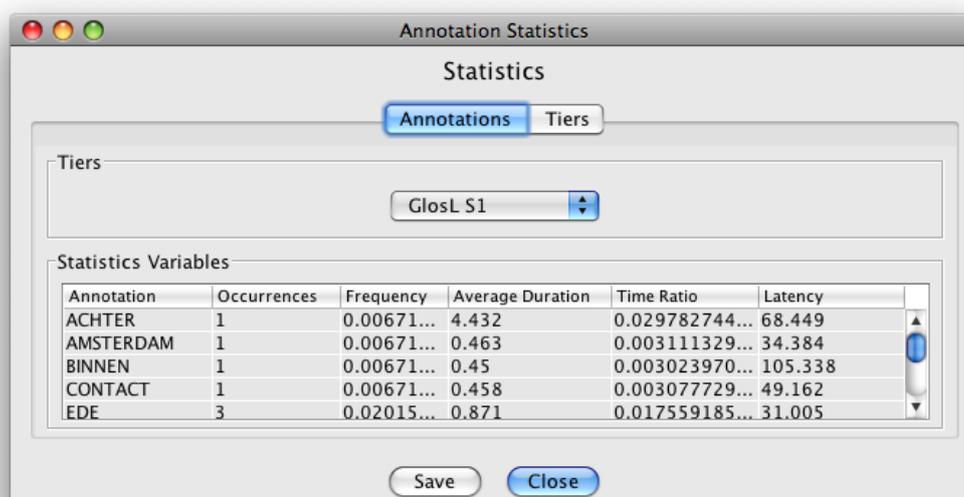


Figure 2. Annotation statistics

annotators may no longer pay close attention to the timing differences between the two hands in two-handed

an annotation can be pasted at a different position in the timeline) or to the same document.

- A new variant of **‘multiple file search’** was implemented. In addition to the pre-existing ‘simple text search’ in multiple files, now structured searches combining search criteria on different tiers can be carried out in a subset of files that can be compiled by the user. The matching types ‘exact match’, ‘substring match’ and ‘regular expression match’ are available and the search can be restricted to a certain time interval. It is also possible to specify a minimal and/or maximal duration for matches.

The results can be displayed in concordance view, with a variable size of the context, or in frequency view, showing the absolute number of occurrences of each hit as well as the relative number (percentage). The results can be exported to a tab-delimited text file with multiple columns.



As a special case, a search for n-gram patterns can be executed, where the pattern should be found either within (multiword) annotations or over annotations on the same tier.

- The **segmentation** function was further developed so that annotations with a fixed, user definable duration can be created by a single key stroke while the media files are playing. The keystroke can either mark the beginning of an annotation or the end. Keyboard navigation through

the media has been made in accordance with this function in the main window.

- A function has been added to flexibly **generate annotation content** based on a user definable prefix and an index number. Indexing can be performed on the annotations of a single tier or on those of multiple tiers.

- A panel can be displayed that lists **basic statistics** for all tiers in an annotation document (Fig. 1): the number of annotations, the minimum, maximum, average, median and total annotation duration per tier, and the latency (start time of the first annotation on that tier). This helps the user getting a better grip on the content in an annotation document and can be helpful in data analysis.

In the same window, a panel can be displayed with a list of unique annotation values for a user-selectable tier (Fig. 2): their number of occurrences and frequency as a fraction of the total number of annotations, the average duration, the time ratio, and the latency (time of first occurrence in the document).

Both panels can be saved as a text file with tab-separated lists.

- The **annotation density viewer** can now also be set to only show the distribution of annotations of a single, selectable tier. The label of a tier in the timeline viewer can optionally show the current number of annotations on that tier.

- The list of existing export options has been enriched by an option to **export a list of unique annotation values** or **a list of unique words** from multiple annotation documents. In the latter case, annotation values are tokenized into single words before evaluating their uniqueness.

- The media files that are associated to a document could already be inspected, added and removed by the ‘linked files’ viewer in the ‘Edit’ menu. Now, easy interactive **hiding and showing of any of the associated video files** is possible, without having to remove the media file association altogether (Figure 3). The maximum number of videos that can be displayed simultaneously is four. But it is possible to add more than four videos to a document and by interactively hiding or showing videos any combination of them can be shown. Temporarily hiding one or more videos can also be useful to improve



Figure 4. New structure of the menu bar

playback performance, especially on less powerful computers.

- A click on a video image copies the **x and y coordinates of the mouse pointer** to the clipboard. The coordinates can then be pasted into any annotation. This can be useful e.g. to record the position of body parts at various moments in time. There are three variants in the format of the coordinates. The reason for this is the ambiguity of dimension and aspect ratio in some popular media formats. As a result, media frameworks can differ in their interpretation of the video dimensions. This has to be taken into account when files are transferred between platforms, ELAN being a multi-platform application running on Windows, Mac OS X and Linux.

### 2.3 User interface

In addition to new functionality, a large number of user interface improvements have been implemented, including the following.

- There is an improved, more intuitive layout of the main menu bar. Due to the increase of functionality, reflected in the growth of the number of items in the menus, some menus had become overpopulated and inconvenient. The key concepts in ELAN ‘Annotation’, ‘Tier’ and ‘Linguistic Type’, were promoted to their own menu in the main menu bar (Figure 4).
- Many additional keyboard shortcuts have been added. The list of shortcuts is logically subdivided into groups of functionally related items and can now be printed.
- A recent files list has been added.
- Easy keyboard navigation through the group of opened documents/windows is now possible.
- There has been a subtle change in the background of the timeline viewer, facilitating the perception of the distinction between the different tiers by the use of lighter and darker horizontal bars (a ‘zebra’ pattern; Figure 5).
- With the use of a new preferences system in version 3, users can now set the **colour of tier labels** in the timeline viewer, thus allowing the visual grouping of related tiers

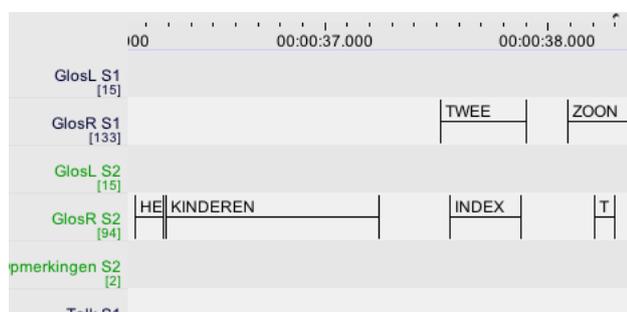


Figure 5. Striped background of the timeline viewer; tier labels with identical colours

in documents containing many tiers by setting the same colour for multiple tiers (as can also be seen in Figure 5). It is also possible to select a preferred font per tier; a Font Browser is included to simplify selection of a suitable font.

- Previously, video windows could only be enlarged (e.g. to view details) or reduced (e.g. to have more screen space for other viewers) by detaching video windows one by one, and adapting the size of each. A function has been added whereby the video windows that are displayed can all be made smaller or larger by dragging a double arrow on the right hand side of the window above the time line viewer. All other viewers automatically resize accordingly, to keep the size of the window constant.

## 3. Future developments

Within ongoing projects, several new needs have become clear which all relate to the fact that suddenly the number of annotation documents that linguists can work with has increased from a small number that one can handle by hand to a huge number (around 2,000 for the *Corpus NGT*). Special attention is needed to keeping the collection well-organised (section 3.1) and to trying to use the available IMDI metadata descriptions to get a grip on the data (section 3.2). In addition, collaborative work with ELAN files is discussed in section 3.3.

### 3.1 Manipulating collections of files

Although enhanced search functionalities and templates facilitate working with multiple ELAN documents, it is not yet possible to ‘manage’ a set of ELAN files systematically in any way. For the specific files and needs of the *Corpus NGT*, Perl scripts were developed in order to add tiers and linguistic types to a set of documents, to change annotation values in multiple documents, and to generate ELAN and preferences files on the basis of a set of media files and existent annotation and preferences files.

For future users, it would be beneficial if such kind of functionality would become available in a more stable and integrated way, whether in ELAN, in the IMDI Browser, or in a stand-alone tool that can manage EAF files.

### 3.2 Use and display of IMDI metadata in ELAN

Current collaboration between the ELAN developers at the Max Planck Institute for Psycholinguistics and the sign language researchers at Radboud University are targeted at enhancing search facilities and facilitating team work between researchers using large language corpora containing ELAN documents.

Currently, annotation files that are included in an IMDI corpus can be searched using ANNEX, the web interface to annotation documents<sup>6</sup>, after a subset of metadata sessions has been selected through an IMDI search. For example, one can first search for all IMDI sessions that include male signers, and then search in all EAF files that are linked to the resulting set of IMDI sessions. In this way, metadata categories and annotations can be combined.

However, currently, ANNEX cannot be used for many tasks: annotations cannot be added, edited or deleted, and the synchronous playback of multiple video streams is not

<sup>6</sup> <http://www.lat-mpi.eu/tools/annex/>

accurate. A separate two-step search is thus being developed for local corpora and the stand-alone version of the IMDI Browser.

Searching is a useful way to combine data and metadata categories, but it implies that one knows what one is looking for. Browsing through an annotation document can also be useful for many types of research, but in that case, metadata information is not available unless one knows it by heart. While the gender of the signer/speaker can be easily established by looking at the video, this does not hold for many other categories: regional or dialect background of the participant, deafness, precise age, recording date, etc. It is therefore important to have quick access to the metadata information linked to an annotation document. This requires that an IMDI metadata description is present, and that the EAF file is linked to the IMDI session. Currently, different ways of displaying metadata information in ELAN are being investigated. Some form will be available in a future version of ELAN in 2008.

### 3.3 Collaborative annotation

Larger collections of files are typically used not by single researchers but by research groups, and stored not on a local drive but on network drives or integrated in a corpus. This requires some type of systematic ‘collaborative annotation’ to ensure that changes made by one person are also available to others. Moreover, one could imagine that people add different values to annotations, that are simultaneously present and can be compared. This would be particularly useful for different translations or analyses of the same construction. Brugman et al. (2004) already discussed ways in which users at different locations look at and edit annotation documents together. We expect this concept to be further developed in the future.

## 4. Conclusion

A corpus building project like the present one clearly provides a fruitful collaboration between software developers and the users of the software. Although the fact that the *Corpus NGT* project was carried out on the same campus as the Max Planck Institute for Psycholinguistics facilitated collaboration, one can certainly imagine that future corpus projects reserve budget for similar collaborations between software developers and linguists. In this way, relatively small software tools can gradually be developed to match the needs of large groups of users.

## 5. References

- Brugman, H., O. Crasborn & A. Russell (2004) Collaborative annotation of sign language data with peer-to-peer technology. In: Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation, M.T. Lino et al., eds. Pp. 213-216.
- Brugman, H. & A. Russell (2004). Annotating Multi-media / Multi-modal resources with ELAN. In: Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation.

- IMDI Team (2003), IMDI Metadata Elements for Session Descriptions, Version 3.0.4, MPI Nijmegen.  
[http://www.mpi.nl/IMDI/documents/Proposals/IMDI\\_MetaData\\_3.0.4.pdf](http://www.mpi.nl/IMDI/documents/Proposals/IMDI_MetaData_3.0.4.pdf)
- Vermeerbergen, M., L. Leeson & O. Crasborn (Eds.). (2007). Simultaneity in signed languages: form and function. Amsterdam: John Benjamins.
- Wittenburg, P., D. Broeder & B. Sloman (2000), EAGLES/ISLE: A Proposal for a Meta Description Standard for Language Resources, White Paper. LREC 2000 Workshop, Athens.  
[http://www.mpi.nl/IMDI/documents/Proposals/white\\_paper\\_11.pdf](http://www.mpi.nl/IMDI/documents/Proposals/white_paper_11.pdf)

# The *Corpus NGT*: an online corpus for professionals and laymen

Onno Crasborn, Inge Zwitserlood

Department of Linguistics, Radboud University Nijmegen

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

E-mail: o.crasborn@let.ru.nl, i.zwitserlood@let.ru.nl

## Abstract

The *Corpus NGT* is an ambitious effort to record and archive video data from Sign Language of the Netherlands (Nederlandse Gebarentaal: NGT), guaranteeing online access to all interested parties and long-term availability. Data are collected from 100 native signers of NGT of different ages and from various regions in the country. Parts of these data are annotated and/or translated; the annotations and translations are part of the corpus. The *Corpus NGT* is accommodated in the Browseable Corpus based at the Max Planck Institute for Psycholinguistics. In this paper we share our experiences in data collection, video processing, annotation/translation and licensing involved in building the corpus.

## 1. Introduction

As for most sign languages, NGT resources are scant. Still, such resources are direly needed for several purposes, sign language research not the least. The aim of the *Corpus NGT* is to provide a large resource for NGT research in the shape of movies of native NGT signers. The signed texts include several different genres, and the signers form a diverse group in age and regional background. Besides the movies, crude annotations and translations form (a small) part of the corpus, so as to ease access to the data content. The corpus is made publicly available to answer the need for NGT data (e.g. by NGT teachers and learners and interpreters).

## 2. Data collection

### 2.1 Participants

The initial aim was to record 24 native signers, divided over two regions where two different variants of NGT are reported to be used. The plan was changed in its early stages so as to include a much larger number of participants, spread over all five reported variant regions. Moreover, by including participants from different ages, it was possible to record older stages of NGT, even male and female variants in these older stages. Altogether, this ensures a good sample of the current state of the language

The participants were invited to take part in the recordings by announcements on Deaf websites, flyers and talks at Deaf clubs, and by ‘sign of hand’. Interestingly, when the project became familiar in the Deaf community, many older people wanted to participate, in order to preserve their own variant of NGT. Because most signers are familiar with the use of contact varieties combining signs with spoken Dutch and because the variation in the form of such contact varieties is very large,

participants were selected who are deaf from birth or soon after and who started to use NGT at a very early age (preferably before the age of 4). Also, we tried to eliminate standardised NGT (an artificial variant of NGT, recently constructed on request of the Dutch government; Schermer 2003).

### 2.2 Tasks and materials

In building the corpus, we followed the project design developed by the constructors of the Auslan corpus project<sup>1</sup>, although adaptations were made to match the situation in the Netherlands. This means that a subset of the tasks given to the participants of the Auslan project were used, using the same or similar stimuli. These included narratives based on cartoons (the Canary Row cartoon of Tweety & Sylvester), fable stories presented to the signer in NGT, comic stories (including the Frog Story children book), and TV clips (e.g. funniest home videos). Besides elicitation of such monologue data, (semi-)spontaneous conversation and discussion forms a substantial part of the *Corpus NGT*. Using the advice from the Auslan experience, the elicitation materials that were used contained as little written text as possible. The participants were all asked to briefly introduce themselves and to tell about one or more life events they experienced. Most importantly (in terms of quantity and content), they were asked to discuss a series of topics introduced to them in NGT movies concerning Deaf and sign language issues. Finally, they engaged in a task where they had to spot the differences between two pictures they had in front of them. In addition to these tasks, occasional free conversation was also recorded.

### 2.3 Recording situation

The participants were recorded in pairs, to encourage ‘natural’ signing as much as possible.

---

<sup>1</sup> <http://www.hrelp.org/grants/projects/index.php?lang=9>

Beforehand, the purpose of the corpus and the tasks and proceedings were explained to them by a native Deaf signer, who also led the recordings. Explanation and recording took approximately 4 hours, and resulted in  $\pm 1.5$  hours of useable signed data per pair. Some recordings were made at the Radboud University and the Max Planck Institute for Psycholinguistics, both in Nijmegen. However, most recordings were made in Deaf schools, Deaf clubs or other places that were familiar to the Deaf participants. All recordings from the northern region (Groningen) were made at the Guyot institute for the Deaf in Haren.<sup>2</sup> As a result of the different sizes and light circumstances of the rooms, there is some variation in the recordings. All recordings were made with consumer quality cameras; no additional lighting equipment was used.

In a recording session, the participants were seated opposite each other, preferably in chairs without armrests as these might hamper their signing. An upper body view and a top view of each signer were recorded. This situation is illustrated in Figure 1. In combination, these front and top views approximate a three-dimensional view of the signing. Previous research has shown that such a view can give valuable information on the use of space and even on the shape of signs, if these are not completely clear from the front view (Zwitserlood, 2003). The top views were recorded with two Sony DV cameras on mini-DV tapes. The cameras were attached with bolts to metal bookstands that could be easily attached to the ceiling above the seated participants. The front views were recorded using two Sony High Definition Video (HDV) cameras on mini-DV tapes; these were mounted on tripods. The upper body view was recorded slightly from the side. This had the advantage of a better view of the signing (since a recording straight from the front does not always give reliable or clear information on the location and handshape(s) in particular signs). Also, when one looks at the front view recordings of both participants in a session, the slight side view gives a better impression of two people engaged in conversation, rather than two people signing to cameras.

We chose to use HDV recordings for the front views because of the high resolution (the full HD recording includes 1920x1080 pixels in contrast to normal digital video, with a format of 720x568 pixels for the European PAL format), resulting in recordings that are very detailed in comparison to standard PAL video. Furthermore, we wanted to provide detailed information on facial expressions; the HDV resolution allowed cutting out a view on

<sup>2</sup> We thank Annemieke van Kampen for her work in finding participants and in leading all the recording sessions in Groningen.

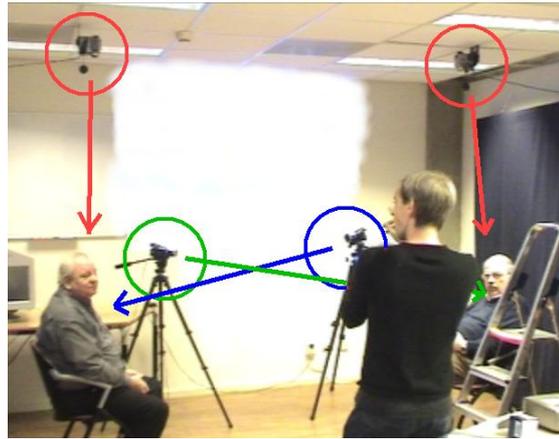


Figure 1: Recording situation

the face, rather than having to use two additional cameras that could be zoomed in on the face.

The recording sessions were lead by a Deaf native signer, who would explain the aims of the project and the procedure beforehand to the participants, allowing ample time for questions, and who stressed the fact that we were especially interested in normal signing, viz. they should try not to sign “neater” or “more correct” than usual. Every new task was explained in detail, and if necessary, the session leader would give examples or extra information during the execution of a task.

For each pair of participants, there were three one-hour recording sessions. In between there were breaks in which the participants could rest and chat, and the tapes were replaced by new ones. Since the cameras were not connected to each other electronically and since switching the four cameras into recording mode by a remote control proved unreliable, each camera was switched on by hand. When all four cameras were running, there would be three loud hand claps, that would show in all the recordings and could thus be used to synchronise the four video streams afterwards.

### 3. Data processing

We took the following steps in processing the recorded data: data capturing, editing, and compression. These are explained in the following sections.

#### 3.1 Capturing and editing

For capturing and editing of the recorded tapes, the video processing programme Final Cut Pro (version 5.3.1, later version 6.0.2) was used. This is a professional video editing programme and the only one that, at the time, was able to handle HDV format video as well as normal DV video. The content of the videotapes was captured in Apple computers (using OS X version 10.4, later 10.5). A Final Cut project contains the four tapes of a

recording session, that are then synchronised on the basis of the clap signal. Subsequently, as many fragments as possible were selected for further use (even those where signers were grunting about a particular task), and all other bits in between were cut out (where a participant was looking at the stimuli or the session leader was explaining something). The selected fragments were assigned a specific “session code” (e.g. CNGT0018) with a postscript indicating the signer (S001 to S100) and the viewpoint of the camera (‘t’ for top view, ‘f’ for face and ‘b’ for body) exported to Quicktime movies in DV and HDV format, respectively. These ‘raw DV’ files were too large to be included in the corpus or to be used productively in applications such as ELAN; for that reason, all movies were compressed to different MPEG formats.

### 3.2 Compression

The project aimed at providing movies that can be used for different purposes and in different applications; moreover, the video should still be accessible in a few decades from now. For this reason, we followed the policy of the data archive at the Max Planck Institute for Psycholinguistics (which is also the location of the DOBES archive) to use MPEG-1 and MPEG-2 video files. The latter keeps the original PAL image size, while the former reduces the size to about one quarter, often 352x288 pixels. The various MPEG standards are a publicly defined and accessible standard, and are not a commercial format promoted and protected by a company (such as the Flash video standard is owned by Adobe).

The resulting movie clips can be easily used in various software applications such as the annotation tool ELAN (see section 5.1). The combination of the MPEG-1 format and the segmenting of video recordings into smaller clips ensures that the movies can be readily downloaded or viewed online. The MPEG-2 version of the top view movies are also included in the corpus for those who need a higher quality image; and also as a relatively unreduced original that can be converted to future video standards in the future. The hosting of the whole corpus at MPI ensures that the material in the corpus will be converted to future standards along with the many other corpora in the corpus in the future.

For the body and face views, a different procedure was followed. In the first stages of the project (late 2006), we were not able to find a compression technique that was able to maintain the full resolution of the HDV recordings. Although the H.264 compression method that is part of the MPEG-4 standard should in principle be able to maintain the full spatial and temporal resolution at highly reduced data rates, we were not able to produce such files. Since both this standard and the

HDV recording techniques were only just appearing on the (consumer) market, we decided to wait with a decision on the high-quality archive format of the HDV recordings. For now, such high-resolution recordings will not be frequently used anyway, given the infrequent use of high-resolution displays: the 1920x1200 resolution is equal to the better graphic cards and (22” and 23”) monitors on the market nowadays, and few computer setups will be used with two such displays side by side (needed to play back the conversations in the corpus). At the end of the project in May 2008, we still have not yet decided what to use as a full-resolution format; the ‘raw’ HDV exports from Final Cut Pro will be included in the corpus for future processing. They can be played back in Quicktime-compatible video players, but are not yet de-interlaced.

To be able to use the recordings productively, we decided to create two MPEG-1 movies from every HDV file. Since the aspect ratio of MPEG-1 (4x3) does not match that of HDV (16x9), cropping was necessary anyway; different cropping settings were used to create cut-outs of the face and of the whole upper body plus head; in addition, the face versions were scaled down less than the upper body versions. Thus, for every section of the recordings, we have six MPEG-1 movie files, three for each signer.

At the start of the project, Apple’s Compressor (version 2) appeared to be unreliable for the compression to MPEG-2 format. Therefore, the programme MPEG Encoder Mac 1.5 from MainConcept was used for this type of compression initially. This program has proved to produce good quality MPEG-1 and MPEG-2 movies. However, its disadvantage is that there is no easy way to compress large numbers of movies in an easy batch mode; all settings have to be re-applied for every movie. Because of the large numbers of movies in the corpus, this was too labour-intensive. Midway the project, when Compressor version 3 proved to have a reliable MPEG-2 compression option, we switched to that programme for the production of both MPEG-1 and MPEG-2 versions.

In all parts of the corpus, even in the ‘monologue’ story-telling, two signers interact. For a good understanding of the signing one therefore needs the movies of both participants, and they should be played in synchrony. While this is a standard function of the ELAN annotation software (see section 5), most common movie players that are integrated in web browsers are not built to play separate movies simultaneously. Therefore, we also provide movies in which the MPEG-1 movies of the front view of both signers are combined into one wide-screen image. These combined movies also have MPEG-1 compression settings, but the aspect ratio is that of two juxtaposed MPEG-1

movies. This process was carried out by the Ffmpeg and Avisynth tools for Windows.

Finally, after the MPEG-1 and MPEG-2 movies have been published online as part of the MPI Browsable Corpus, in the near future all movies will be converted into streaming MPEG-4 clips and made accessible through MPI's streaming server. In this way, movies can be easily accessed by online tools such as ANNEX<sup>3</sup>.

## 4. Access

### 4.1 Metadata

The *Corpus NGT* is published by the MPI for Psycholinguistics, as part of their growing set of language corpora. We follow the IMDI standard for creating metadata descriptions and corpus structuring.<sup>4</sup> These metadata concern information about the type of data (narrative, discussion, retelling of cartoon, etc) and about the participants. Although all data are freely accessible and the participants are clearly visible in the movies, their privacy is protected as much as possible by restricting the metadata for the participants in the corpus to their age at the time of the recording, their age of first exposure to NGT, their sex, the region where they grew up, and their handedness. Researchers who need more information (e.g. about the fact whether there are deaf family members) can request such information from the corpus manager. Names or initials are not used anywhere in the metadata description for the participants.

### 4.2 Access for all

Although the corpus is mainly intended for linguistic research, the data can have several other uses. Because of the need of NGT data indicated earlier, we are happy to share the data in the corpus with other people who need such data or are interested in them, providing open access to all video and annotation documents. Other interested scientists may be psychologists, educators, and those involved in constructing (sign) dictionaries. Deaf and hearing professionals in deaf schools and in the Deaf community may want to use the material, including NGT teachers, developers of teaching materials, and NGT interpreters. Many hearing learners of NGT will benefit from open access to a large set of data in their target language. Deaf people themselves may be interested in the discussion on deaf issues that forms part of every recording session.

All participants in the corpus signed a consent form that explicitly mentions the online publication and the open access policy. The forms in Dutch were explained by the Deaf person leading the

recording session. Most importantly, the publication and possible use of the material was explained to the signers before they agreed to come and participate. During the actual recordings, signers were encouraged to limit the amount of personal information they might reveal in their discussions. In a few cases, we decided to leave out privacy-sensitive segments after the recordings, often in conformance with requests from the participants.

Since the construction of large sign language corpora is a recent phenomenon, we hope that our experiences will be valuable for other projects. Therefore, the project's open access policy extends beyond the video data to the annotations, workflows and guidelines for tools that have been used, which will all be published online.

Although everyone has free access to the data in the MPI Browsable Corpus that is available via the internet,<sup>5</sup> searching and finding interesting movies in the large corpus is not an easy or quick task. Therefore, we are currently designing a few websites for specific target groups (e.g. NGT teachers, deaf children and their parents, NGT interpreters), from which websites selected movies are easily accessible.

### 4.3 Licensing

The use and reuse of the data are encouraged *and* protected at the same time by Creative Commons licenses (see Crasborn, this volume, for further discussion). Creative Commons offer six types of protection, ranging from restrictive to highly accommodating. We chose the combination BY-NC-SA:

1. Attribution: when publishing (about) data of this corpus, mention the source (BY);
2. Non-commercial: no part of this corpus can be used for commercial purposes (NC);
3. Share alike: (re)publishing (parts of) data of this corpus should be done under the same licenses (SA).

The first two licenses are self-explanatory. The third license is meant to encourage other people to make use of the data and to share new data based on data from the corpus with others (while, again, protecting the new data). For example, an NGT teacher may want to use a part of a movie to point out particular grammatical phenomena to her students, or provide a movie with subtitles, and share the new movie with colleagues. Alternatively a researcher interested in a particular aspect of NGT may use an annotation file, add new annotations and share the enriched file with other researchers. The licenses are mentioned in the metadata. Also, the licenses are part of all the movies in the corpus: a short message in Dutch and

<sup>3</sup> <http://www.lat-mpi.eu/tools/annex/>

<sup>4</sup> <http://www.mpi.nl/imdi/>

<sup>5</sup> <http://corpus1.mpi.nl>

English is shown at the start and end of each movie.

## 5. Accessibility of the data

Not all people who may be interested in the data of the corpus are fluent in NGT. For these users, the corpus provides ways to gain (better) access to at least parts of the data, viz. annotations and translations.

### 5.1 Annotation

For annotation the annotation tool ELAN (Eudico Linguistic Annotator), developed at the MPI for Psycholinguistics, was used.<sup>6</sup> This program is currently widely in use for the annotation of various kinds of linguistic and other communicative data. This tool allows online transcription where the original data (a sound or video file) can be played and annotations can be aligned with the audio of video signal. Originally used for annotation of gesture, it has been improved substantially since it also started to be used for sign language annotation. Based on experiences in previous projects (e.g. ECHO<sup>7</sup>) and desired functionality in the corpus project, various new features were formulated and implemented in the software (see Crasborn & Sloetjes, this volume). The extension of ELAN as well as the integration of ELAN and IMDI (the data and metadata domains) formed a substantial part of the project.

Annotation is an enormously time-consuming process. Due to time and budget limitations (the project was funded for two years), and as we invested in more recordings than originally planned which left less time for annotation, it was only possible to provide crude gloss annotations of a small subset of the data. Four Deaf assistants were assigned this job, on a part-time basis to avoid health problems because of the intensive use of the computer. They were trained to use ELAN (showing only a front view of both participants) and to gloss the signs made by the left and right hand with a Dutch word, or a description if there was no appropriate Dutch word available. They could use a bilingual Dutch-NGT dictionary holding approximately 5000 lemmas and Dutch (picture and normal) dictionaries to check Dutch spelling, as well as a reference list with the gloss conventions to be used. These conventions were based on and adapted from the conventions used in ECHO; see Nonhebel *et al.* 2004. At the end of the project, 160 movies were annotated, totalling almost 12 hours of signing and 64.000 glosses. Unfortunately, the assistants' skills in Dutch appeared to be quite poor, resulting in a rather large

amount of spelling and writing mistakes in the annotations. In addition, they did not remember conventions well enough and/or seemed to be reluctant in looking up information that they needed. Also, it appeared to be a hard task to focus solely on the manual component of signing in determining a gloss text, annotators almost automatically look at the meaning of the whole construction, including facial expression and other non-manual behaviour. Because of that, several other mistakes occur in the annotations, including misalignments between start and end of many signs and their annotations. We corrected the most salient spelling mistakes and diacritics used in the wrong places. Furthermore, some of the ELAN files were corrected by a Deaf signer experienced in the use of ELAN and in annotation. Still, the current annotations should not be blindly relied upon. We plan to do further corrections and to provide more and more detailed annotations in future research projects.

### 5.2 Translation

Annotations are very helpful in doing linguistic research. However, besides researchers, the data are also made available to other interested parties. In order to make as much of the data set accessible to a large audience, parts of the data are provided with a voice-over translation, done by interpreters and interpreter students. For this, empty ELAN files were created, only showing front view movies of two participants for the data to be translated. The interpreters were instructed in the navigation of ELAN and in the use of a Sony minidisc recorder with one or two microphones (depending on whether the movies to be translated involved monologues or dialogues). Their job was to look at a particular movie one or two times, if necessary to discuss difficult parts with a colleague, switching on the minidisc recorder and give a running translation while watching the movie. The audio files on the minidisks were processed into WAV files, aligned with the movies and connected to ELAN files.

The interpretation of the (often unknown) participants in discussion turned out to be a challenging task. The option to play back the movie is almost irresistible to interpreters if they know that they may not have fully understood every detail. As sign language interpreters are rarely in such a position, typically doing simultaneous interpreting in events where they have little control over things like signing rate, the voice interpreting for the *Corpus NGT* was an attractive task, precisely because of the option to replay and discuss the interpretation with their colleague. The nature of this process can be considered a mix between interpretation and translation. On average, the interpreting process (including administrative

<sup>6</sup> <http://www.lat-mpi.eu/tools/elan/>

<sup>7</sup> <http://www.let.ru.nl/sign-lang/echo/>

and technical tasks related to the recording procedure) took ten times realtime (thus, one hour of signing took ten hours to record on minidisc). Because of the increase in recorded hours of signing with respect to the original plan, it was not possible to provide a voice-over with all video recordings.

Originally we had hoped for the possibility to transfer the speech signal of the interpreters into written Dutch using speech recognition software. However, this appeared not to be possible because of a combination of factors. First, most speech recognition programs need to be trained to recognize the speech of the interpreters; it appeared to be impossible to set this up logistically. Second, speech recognition software that we could use does not need the auditory signal for training, but instead, uses word lists. However, the wide range of lexical items and spontaneous nature of the spoken translations appeared to be too variable for reliable transfer to written text. Taking into account the post-hoc corrections that would be necessary, it is probably cheaper and more reliable to use typists. This is clearly an option for the future.

## 6. Future developments

It is clear from the programme of the present workshop alone that we can expect rapid developments in the field of corpus studies in signed languages. There is an enormous increase in the data that linguists have at their disposal, which will enable deeper insights in the linguistic structure and in the variability of signing within a community. Even though the *Corpus NGT* explicitly aimed to exclude signers that only used some form of sign supported Dutch, the influence of Dutch appears to vary greatly across age groups, an observation that has not yet received any attention in the literature.

In order to carry out such linguistic studies, we need clear standards for annotation and transcription in sign language research. While there have been some efforts in the past, for example as collected in the special double issue of *Sign Language & Linguistics* (issue 4, 2001), there is very little standardisation for common phenomena such as gloss annotations. We hope that the increasing use of shared open source tools such as ELAN that use published XML file formats will increase the possibilities for exchanging data between research groups and countries, and promote standardisation among linguists.

In terms of technology, progress is slowly being made in automatic sign recognition. Having tools that enable some form of automatic annotation would constitute a next large jump in the construction and exploitation of sign language corpora. Recording and publishing video data online is now possible, but the Achilles heel in

using them remains in accessing the large amounts of data: search tools need to be enhanced, but for these tools just as for the linguistic eye, annotations remain crucial, yet require an enormous investment in time and money. For the *Corpus NGT*, we hope that its use by various researchers in the near future will slowly increase the 15% of the data that have been glossed until now.

## 7. Acknowledgements

The *Corpus NGT* project is funded by the Netherlands Organisation for Scientific Research (NWO) by a two-year grant, nr. 380-70-008. We thank the participants in the recordings for their valuable contributions.

## 8. References

- Nonhebel, A., Crasborn, O. & Van der Kooij, E. (2004) Sign language transcription conventions for the ECHO project. Version 9, 20 January 2004. Radboud University Nijmegen.
- Schermer, T. (2003). From variant to standard. An overview of the standardization process of the lexicon of Sign Language of the Netherlands (SLN) over two decades. *Sign Language Studies*, 3(4), 96-113.
- Zwitsersloot, I. (2003) Classifying Hand Configurations in Nederlandse Gebarentaal (Sign Language of the Netherlands). Utrecht, LOT Dissertation Series 78.

# Towards Automatic Sign Language Annotation for the ELAN Tool

*Philippe Dreuw and Hermann Ney*

Human Language Technology and Pattern Recognition Group,  
RWTH Aachen University, Aachen, Germany  
{dreuw, ney}@cs.rwth-aachen.de

## Abstract

A new interface to the ELAN annotation software that can handle automatically generated annotations by a sign language recognition and translation framework is described. For evaluation and benchmarking of automatic sign language recognition, large corpora with rich annotation are needed. Such databases have generally only small vocabularies and are created for linguistic purposes, because the annotation process of sign language videos is time consuming and requires expert knowledge of bilingual speakers (signers). The proposed framework provides easy access to the output of an automatic sign language recognition and translation framework. Furthermore, new annotations and metadata information can be added and imported into the ELAN annotation software. Preliminary results show that the performance of a statistical machine translation improves using automatically generated annotations.

## 1. Introduction

Currently available sign language video databases were created for linguistic purposes (Crasborn et al., 2004; Neidle, 2002 and 2007) or gesture recognition using small vocabularies (Martinez et al., 2002; Bowden et al., 2004). An overview of available language resources for sign language processing is presented in (Zahedi et al., 2006). Recently, an Irish Sign Language (ISL) database (Stein et al., 2007) and an American Sign Language (ASL) database (Dreuw et al., 2008) have been published.

Most available sign language corpora contain simple stories performed by a single signer. Additionally, they have too few observations for a relatively large vocabulary which is inappropriate for data driven and statistically based learning methods. Here we focus on the automatic annotation and metadata information for benchmark databases that can be used for analysis and evaluation of:

- linguistic problems
- automatic sign language recognition systems
- statistical machine translation systems

For storing and processing sign language, a textual representation of the signs is needed. While there are several notation systems covering different linguistic aspects, we focus on the so called gloss notation, being widely used for transcribing sign language video sequences.

*Linguistic research* in sign language is usually carried out to obtain the necessary understanding regarding the used signing (e.g. sentence boundaries, discourse entities, phonetic analysis of epenthetic movements, coarticulations, or role changes), whereas *computer scientists* usually focus on features for sign language recognition (e.g. body part tracking of head and hands, facial expressions, body posture), or on post-processing and additional monolingual data for statistical machine translation to cope with encountered sign language related statistical machine translation errors.

Therefore some common important features and search goals for these different research areas are e.g.

- body part models and poses, hand poses, facial expressions, eye gaze, ...

- word spotting and sentence boundary detection
- pronunciation detection and speaker identification

In particular, statistical recognition or translation systems rely on adequately sized corpora with a rich annotation of the video data. However, video annotation is very time consuming: in comparison to the annotation of e.g. parliamentary speech, where the annotation real-time-factor (RTF) is about 30 (i.e. 1 hour of speech takes 30 hours of annotation), the annotation of sign language video can have a annotation RTF of up to 100 for a full annotation of all manual and non-manual components.

## 2. Baseline System Overview & Features

Figure 1 illustrates the components of our proposed recognition and annotation system.

The recognition framework and the features used to achieve the experimental results have been presented in (Dreuw et al., 2007a). The baseline automatic sign language recognition (ASLR) system uses appearance-based image features, i.e. thumbnails of video sequence frames. They give a global description of all (manual and non-manual) features that have been shown to be linguistically important. The system is Viterbi trained and uses a trigram language model (Section 2.4.) which is trained on the groundtruth annotations of main glosses.

The ASLR system is based on the Bayes' decision rule: for a given sign language video input sequence, first features  $x_1^T$  are extracted to be used in the global search of the model which best describes the current observation:

$$\begin{aligned} & \arg \max_{w_1^N} \{Pr(w_1^N | x_1^T)\} \\ & = \arg \max_{w_1^N} \{Pr(w_1^N) \cdot Pr(x_1^T | w_1^N)\} \end{aligned} \quad (1)$$

The word sequence  $w_1^N$  (i.e. a gloss sequence) which maximizes the language model (LM) probability  $Pr(w_1^N)$  and the visual model probability  $Pr(x_1^T | w_1^N)$  will be the recognition result.

Statistical machine translation (SMT) is a data-driven translation method that was initially inspired by the so-called noisy-channel approach: the source language is interpreted

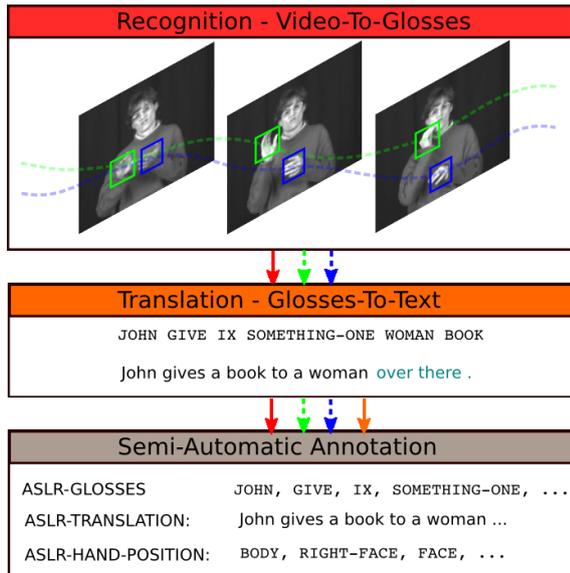


Figure 1: Complete system setup with an example sentence: After automatically recognizing the input sign language video, the translation module has to convert the intermediate text format (glosses) into written text. Both system outputs and features can be used to automatically generate annotations.

as an encryption of the target language, and thus the translation algorithm is typically called a decoder. In practice, statistical machine translation often outperforms rule-based translation significantly on international translation challenges, given a sufficient amount of training data.

A statistical machine translation system presented in (Dreuw et al., 2007b) is used here to automatically transfer the meaning of a source language sentence into a target language sentence. Following the notation convention, we denote the source language with  $J$  words as  $f_1^J = f_1 \dots f_J$ , a target language sentence as  $e_1^I = e_1 \dots e_I$  and their correspondence as the a-posteriori probability  $\Pr(e_1^I | f_1^J)$ . The sentence  $\hat{e}_1^I$  that maximizes this probability is chosen as the translation sentence as shown in Equation 2. The machine translation system accounts for the different grammar and vocabulary of sign language.

$$\hat{e}_1^I = \arg \max_{e_1^I} \{ \Pr(e_1^I | f_1^J) \} \quad (2)$$

$$= \arg \max_{e_1^I} \{ \Pr(e_1^I) \cdot \Pr(f_1^J | e_1^I) \} \quad (3)$$

For a complete overview of the translation system, see (Mauser et al., 2006).

### 2.1. Body Part Descriptions

The baseline system is extended by hand trajectory features (Dreuw et al., 2007a) being similar to the features presented in (Vogler and Metaxas, 2001). Similar as presented in (Bowden et al., 2004; Yang et al., 2006), features such as the relative position and pose of the body, the hands or the head could be extracted. The proposed system can be easily extended by other feature extraction methods which could

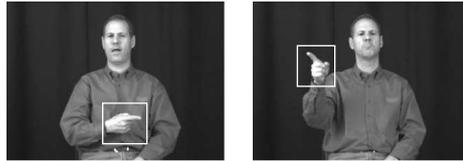


Figure 2: Sample frames for pointing near and far used in the translation.

extract further user specific metadata information for the annotation files.

To enhance translation quality, we propose to use visual features from the recognition process and include them into the translation as an additional knowledge source.

### 2.2. Pronunciation Detection and Speaker Identification

Given dialectal differences, signs with the same meaning often differ significantly in their visual appearance and in their duration. Each of those variants should have a unique gloss annotation.

Speakers could e.g. be identified using state-of-the-art face detection and identification algorithms (Jonathon Phillips et al., 2007).

### 2.3. Sentence Boundary Detection and Word Spotting

Temporal segmentation of large sign language video databases is essential for further processing, and is closely related to sentence boundary detection in speech recognition (ASR) and tasks such as video shot boundary detection (Quenot et al., 2003).

In addition to audio and video shot boundary detection, which is usually done just at the *signal level*, we could use the hand tracking information inside the virtual signing space from our sign language recognition framework to search for sentence boundaries in the signed video streams (e.g. usage of neutral signing space). Due to the different grammar in sign language, a word spotting of e.g. question markers (e.g. so called ONSET, OFFSET, HOLD or PALM-UP signs (Dreuw et al., 2008)) could deliver good indicators for possible sentence boundaries.

### 2.4. Language Models

Due to the simultaneous aspects of sign language, language models based on the (main gloss) sign level versus independent language models for each communication channel (e.g. the hands, the face, or the body) can be easily generated using e.g. the SRILM toolkit (Stolcke, 2002) and added as metadata information to the annotation files.

## 3. Automatically Annotating ELAN Files With Metadata Information

The ELAN annotation software<sup>1</sup> is an annotation tool that allows you to create, edit, visualize, and search annotations for video and audio data, and is in particular designed for the analysis of language, sign language, and gesture. Every ELAN project consists of at least one media file with its corresponding annotation file.

<sup>1</sup><http://www.lat-mpi.eu/tools/elan/>

Our proposed automatic annotation framework is able to:

- convert and extend existing ELAN XML annotation files with additional metadata information
- automatically annotate new media files with glosses (video-to-glosses), translations (glosses-to-text), and metadata information from the automatic sign language recognition framework

The richness of gloss annotation can be defined by different user needs (e.g. sentence boundaries, word spotting, main glosses, facial expressions, manual features, etc.) (c.f. Section 2.), and can depend on the confidence of the sign language recognition or translation framework: the linguist might search for a specific sign and would need high quality annotations, whereas the computer scientist could only import annotations with low confidences and erroneous recognition or translation for a fast analysis and correction of the automatically generated annotations in order to use them for a supervised retraining of the system.

Currently our proposed framework converts a recognizer output file with its corresponding word confidences generated by the `sclite` tool<sup>2</sup> from the NIST Scoring Toolkit (Fiscus, 2007) into a tab-delimited text file, which can be imported by the recently released ELAN 3.4.0 software. The file contains for each tier the begin, end, and duration times of each annotation value.

#### 4. Experimental Results

An independent multi-channel training and recognition will allow automatic annotation of e.g. head and hands. The current whole-word model approach only allows for complete main gloss annotations. However, in another set of experiments presented in (Dreuw et al., 2007b), for incorporation of the tracking data, the tracking positions of the dominant-hand were clustered and their mean calculated. Then, for deictic signs, the nearest cluster according to the Euclidean distance was added as additional word information for the translation model. For a given word boundary, these specific feature informations can be added as an additional tier and imported to the ELAN tool (see ASLR-HAND tiers in Figure 3).

For example, the sentence JOHN GIVE WOMAN IX COAT might be translated into *John gives the woman the coat* or *John gives the woman over there the coat* depending on the nature of the pointing gesture IX (see ASLR-TRANSLATION tier in Figure 3). This helped the translation system to discriminate between deixis as distinctive article, locative or discourse entity reference function.

Preliminary results for statistical machine translation with sign language recognizer enhanced annotation files have been presented in (Dreuw et al., 2007b; Stein et al., 2007). Using the additional metadata, the translation improved in performance from 28.5% word-error-rate (WER) to 26.5% and from 23.8% position-independent WER to 23.5%, and shows the need for further metadata information in corpora annotation files.

<sup>2</sup><http://www.nist.gov/speech/tools/>

Preliminary annotation results for word boundaries, sentence boundaries, and head/hand metadata information are shown in Figure 3. Depending on a word confidence threshold of the recognition system, the amount of automatically added glosses can be controlled by the user (see ASLR-GLOSSES and ASLR-CONFIDENCES tier in Figure 3). This also enables to search for pronunciations (if modeled as e.g. in (Dreuw et al., 2007a)). Furthermore body part and spatial features as proposed in (Stokoe et al., 1965; Bowden et al., 2004) can be added as additional information streams (see ASLR-HAND and ASLR-FACE tiers in Figure 3).

#### 5. Summary & Conclusion

Here, we presented and proposed an automatic annotation extension for the ELAN tool which can handle automatically generated annotations and metadata information from a continuous sign language recognition and translation framework.

Challenging will be multiple stream processing (i.e. an independent recognition of hands, faces, body, ...), pronunciation detection, and speaker identification, as well as the extraction of better visual features in order to improve the quality of the automatically generated annotation files. It will enable to automatically add rich annotations (e.g. head expression/position/movement, hand shape/position/movement, shoulders, eye brows/gaze/aperture, nose, mouth, or cheeks) as already partly manually annotated in (Neidle, 2002 and 2007).

Interesting will be unsupervised training, which will improve the recognition and translation performance of the proposed systems. The implicitly generated ELAN annotation files will allow for fast analysis and correction.

A helpful extension of the ELAN software would be an integrated video annotation library (e.g. simple box drawing or pixel marking) which would allow to use ELAN as a groundtruth annotation tool for many video processing task, and would furthermore allow for a fast and semi-automatic annotation and correction of sign language videos.

#### 6. References

- R. Bowden, D. Windridge, T. Kadir, A. Zisserman, and M. Brady. 2004. A linguistic feature vector for the visual interpretation of sign language. In *European Conf. Computer Vision*, volume 1, pages 390–401.
- Onno Crasborn, Els van der Kooij, and Johanna Mesch. 2004. European cultural heritage online (ECHO): publishing sign language data on the internet. In *Theoretical Issues in Sign Language Research*, Barcelona, Spain, October.
- P. Dreuw, D. Rybach, T. Deselaers, M. Zahedi, and H. Ney. 2007a. Speech recognition techniques for a sign language recognition system. In *Interspeech 2007*, pages 2513–2516, Antwerp, Belgium, August. ISCA best student paper award of Interspeech 2007.
- P. Dreuw, D. Stein, and H. Ney. 2007b. Enhancing a sign language translation system with vision-based features. In *International Workshop on Gesture in Human-Computer Interaction and Simulation*, pages 18–20, Lisbon, Portugal, May.

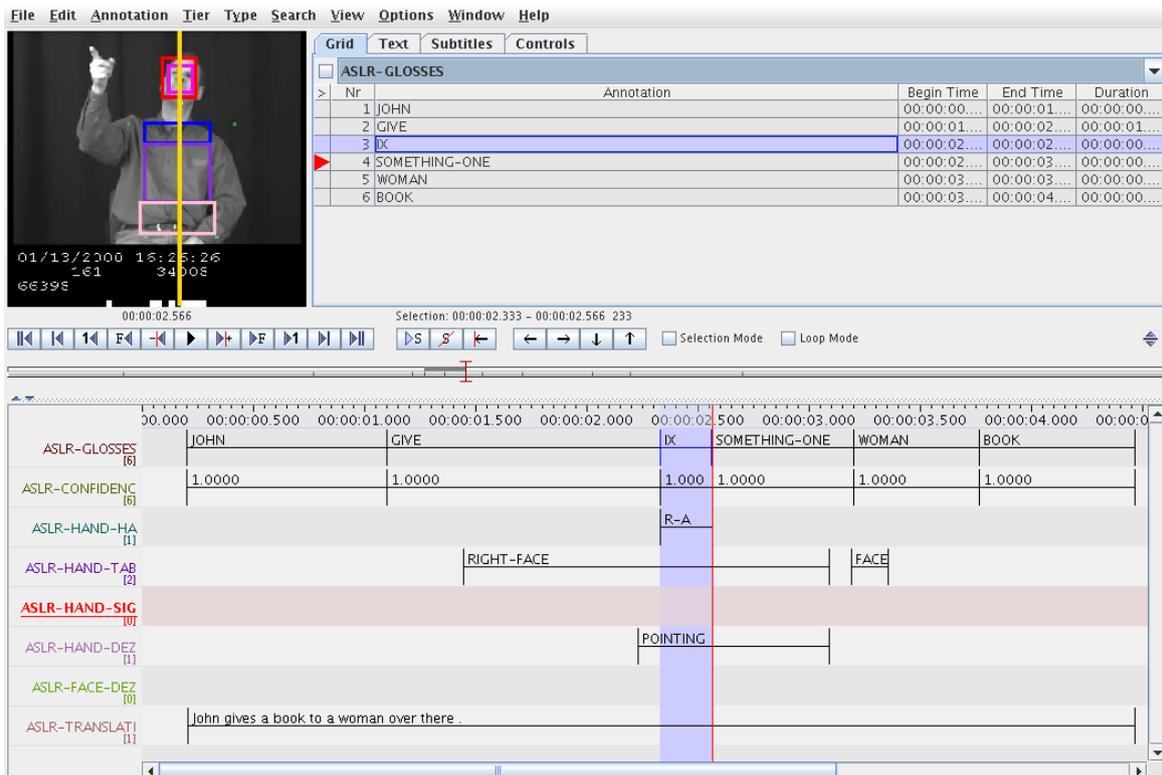


Figure 3: Screenshot of the ELAN tool with automatically generated annotations and video with overlaid features.

- P. Dreuw, C. Neidle, V. Athitsos, S. Sclaroff, and H. Ney. 2008. Benchmark databases for video-based automatic sign language recognition. In *International Conference on Language Resources and Evaluation*, Marrakech, Morocco, May. <http://www-i6.informatik.rwth-aachen.de/~dreuw/database.html>.
- John Fiscus. 2007. Nist speech recognition scoring toolkit (NIST-SCTK). <http://www.nist.gov/speech/tools/>.
- P. Jonathon Phillips, W. Todd Scruggs, Alice J. O'Toole, Patrick J. Flynn, Kevin W. Bowyer, Cathy L. Schott, and Matthew Sharpe. 2007. Frvt 2006 and ice 2006 large-scale results. Technical Report NISTIR 7408, NIST, March.
- A. M. Martinez, R. B. Wilbur, R. Shay, and A. C. Kak. 2002. Purdue RVL-SLLL ASL database for Automatic Recognition of American Sign Language. In *IEEE Int. Conf. on Multimodal Interfaces*, Pittsburg, PA, USA, October.
- Arne Mauser, Richard Zens, Evgeny Matusov, Saša Hasan, and Hermann Ney. 2006. The RWTH statistical machine translation system for the IWSLT 2006 evaluation. In *IWSLT*, pages 103–110, Kyoto, Japan, November. Best paper award.
- Carol Neidle. 2002 and 2007. Signstream™ Annotation: Conventions used for the American Sign Language Linguistic Research Project and addendum. Technical Report 11 and 13, American Sign Language Linguistic Research Project, Boston University.
- G. Quenot, D. Moraru, and L. Besacier. 2003. Clips at trecvid: Shot boundary detection and feature detection. In *TRECVID 2003 Workshop Notebook Papers*, pages 18–21, Gaithersburg, MD, USA.
- Daniel Stein, Philippe Dreuw, Hermann Ney, Sara Morrissey, and Andy Way. 2007. Hand in hand: Automatic sign language to speech translation. In *11th Conference on Theoretical and Methodological Issues in Machine Translation (TMI-07)*, pages 214–220, Skövde, Sweden, September.
- W. Stokoe, D. Casterline, and C. Croneberg. 1965. *A Dictionary of American Sign Language on Linguistic Principles*. Gallaudet College Press, Washington D.C., USA.
- A. Stolcke. 2002. SRILM - an extensible language modeling toolkit. In *ICSLP*, volume 2, pages 901–904, Denver, CO, September.
- C. Vogler and D. Metaxas. 2001. A framework for recognizing the simultaneous aspects of American Sign Language. *Computer Vision & Image Understanding*, 81(3):358–384, March.
- R.D. Yang, S. Sarkar, B. Loeding, and A. Karshmer. 2006. Efficient generation of large amount of training data for sign language recognition: A semi-automatic tool. In *International Conference on Computers Helping People with Special Needs (ICCHP 06)*.
- Morteza Zahedi, Philippe Dreuw, David Rybach, Thomas Deselaers, and Hermann Ney. 2006. Continuous sign language recognition - approaches from speech recognition and available data resources. In *Second Workshop on the Representation and Processing of Sign Languages: Lexicographic Matters and Didactic Scenarios*, pages 21–24, Genoa, Italy, May.

## Annotating Real-Space Depiction

Paul Dudis, Kristin Mulrooney, Clifton Langdon, Cecily Whitworth

Science of Learning Center on Visual Language and Visual Learning (VL2)

Dept. of Linguistics, Gallaudet University

800 Florida Avenue, NE

Washington D.C. 20002

E-mail: paul.dudis@gallaudet.edu, kristin.mulrooney@gallaudet.edu,

clifton.langdon@gmail.com, cecily.whitworth.gallaudet.edu

### Abstract

“Shifted referential space” (SRS) and “fixed referential space” (FRS) (Morgan 2005) are two major types of referential space known to signed language researchers (see Perniss 2007 for a discussion of alternative labels used in the literature). An example of SRS has the signer’s body representing an event participant. An example of FRS involves the use of “classifier predicates” to demonstrate spatial relationships of entities within a situation being described. A number of challenges in signed language text transcriptions identified in Morgan (2005) pertain to the use of SRS and FRS. As suggested in this poster presentation, a step towards resolving some of these challenges involves greater explicitness in the description of the conceptual make-up of SRS and FRS. Such explicitness is possible when more than just the signer’s body, hands, and space are considered in the analysis. Dudis (2007) identifies the following as components within Real-Space (Liddell 1995) that are used to depict events, settings and objects: the setting/empty physical space, the signer’s vantage point, the subject of conception (or, the self), temporal progression, and the body and its partitionable zones. We considered these components in a project designed to assist video coders to identify and annotate types of depiction in signed language texts. Our preliminary finding is that if we also consider the conceptual compression of space—which results in a diagrammatic space (Emmorey and Falgier 1999)—there are approximately fourteen types of depiction, excluding the more abstract ones, e.g. tokens (Liddell 1995).

Included in this poster presentation is a prototype of a flowchart to be used by video coders as part of depiction identification procedures. This flowchart is intended to reduce the effort of identifying depictions by creating binary (yes or no) decisions for each step of the flowchart. The research team is currently using ELAN (EUDICO Linguistic Annotator, [www.lat-mpi.eu/tools/elan/](http://www.lat-mpi.eu/tools/elan/)) to code the depictions focusing on the relationship of genre and depiction type by looking at the depictions’ length, frequency, and place of occurrence in 4 different genres: narrative of personal experience, academic, poetry, conversation. We also have been mindful that a good transcription system should be accessible in an electronic form and be searchable (Morgan 2005). In tiered transcription systems like ELAN the depiction annotation can simply be a tier of its own when it is not the emphasis of the research, or it can occupy several tiers when it is the forefront. In linear ASCII-style transcriptions the annotation can mark the type and beginning then end of the depiction. Our poster does not bring a complete bank of suggested annotation symbols, but rather the idea that greater explicitness as to the type of depiction in question may be beneficial to corpus work.

### 1. Introduction

This paper briefly describes a project aimed towards development of procedures to identify and annotate different ways users of any signed language create iconic representations. One main issue in the transcription of British Sign Language narratives identified by Morgan (2005) is the need for an effective way to demonstrate not only the interactions between what he calls Fixed Referential Space (FRS) and Shifted Referential Space (SRS), but also how linguistic items relate to them. We are reasonably certain that many researchers of other signed languages have similar concerns.

Our approach to this issue is based on Dudis’ (2007) investigation of dynamic iconic representations, or what he terms **depiction**. We first review how the recognition of additional elements within the signer’s conceptualization of her current environment as well as certain cognitive abilities leads to greater precision in describing the various types of depiction produced by signers. We then briefly describe our ongoing attempts to develop depiction identification procedures for purposes of coding and analysis.

### 2. Types of Depiction

To our knowledge, in their examination of depiction, most signed language researchers do not consider any elements within the signer’s conceptualization of the immediate environment other than the signer, the manual articulators, and the surrounding space. Dudis (2007) demonstrates that there are additional Real-Space elements (Liddell 1995) that need to be recognized so as to describe the different ways signers depict things, settings, and events with greater precision. In all there are approximately five Real-Space elements that typically take part in depiction: the setting (or space), the vantage point, temporality, the subject (or the self—note that this does not refer to the clausal subject), and the body. Cognitive abilities also play a role in depiction. The cognitive ability underlying all instances of depiction is conceptual blending (Fauconnier & Turner 2002); see Liddell (2003) for demonstrations on how the conceptual blending model is used to describe “life-sized” blends (surrogates in Liddell’s terms), depicting blends, and token blends. Depiction is the result of creating a network of mental spaces, one of which is Real Space. Another mental space in the network is one that has been built as

discourse proceeds, and it contains elements that correspond to Real-Space elements. The blending of these counterpart elements create the iconic representations that are of interest here, and the space in which they exist is called the blend.

Depictions of someone doing any type of activity involve the blending of several elements. The signer has two options here. First, a life-size blend could be created, one in which the Real-Space subject blends with the individual of the event being depicted. Since individuals exist in time and space, relevant counterpart elements also are blended. This type of depiction, which appears to be the SRS described by Morgan (2005), is represented by Figure 1. The box is the |setting|, the shaded figure is the |subject|, and the arrow represents |temporality|.

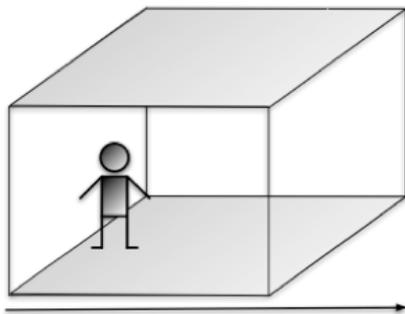


Figure 1

Often it is possible for the signer to choose to create a smaller-scaled depiction of the event. What contributes to this possibility is the cognitive ability to **compress** the setting of the depicted event onto a smaller portion of space, one that is in front of the signer. Since the space that takes part in the depiction does not include the space currently occupied by the signer, she (the Real-Space subject) is not part of the blend. This appears to be the FRS described by Morgan (2005). Figure 2 is a representation of this type of depiction. Since there is no |subject|, the signer is represented as a “regular” figure. The |setting| and |temporality| are represented by a smaller box and arrow. The time of the event being depicted can be compressed into a shorter span of “real time”, but so far we see no compelling reason to include this information in our annotations of depiction. Also, we borrow the terms “viewer” and “diagrammatic” from Emmorey and Falgier (1999) to describe the life-sized versus compressed representations.

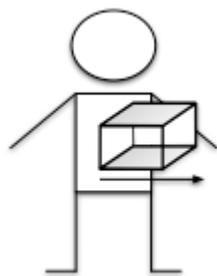


Figure 2

In the figure above we see that it is possible to select some but not all of the Real-Space elements that can take part in depiction. This appears to be what Fauconnier & Turner (2002) call **selective projection**. Dudis (2007) demonstrates that this cognitive ability contributes to the variety of depiction types that can be observed in everyday signed language discourse. As there is a dependency of sorts that certain Real-Space elements have on other elements, there appears to be a limited number of depiction types that signers can produce. For example, the subject must exist within a temporality and a setting, but as we have seen in Figure 2, it is possible to describe an event without creating a |subject| element.

Another cognitive ability that contributes to the variety of depiction types is **body partitioning** (Dudis 2007). The simultaneous activation of SRS and FRS depends on the ability to partition the manual articulators so that they can take part in the creation of representations distinct from the |subject|. We have observed that there are approximately four different types of depiction in which a |subject| is present (six if one wishes to distinguish between constructed dialogue and constructed action that does not involve partitioning; Winston (1991) and Metzger (1995) note that both appear to involve similar strategies). It is possible to depict dialogue and manual action with only the |subject| visible. It is also possible to depict action from the perspective of a patient, e.g. someone being punched, by partitioning off a manual articulator while keeping the |subject| activated; this type of depiction is represented in Figure 3. The manual articulators can also be partitioned off to produce simultaneous perspectives of the event being depicted. This type of depiction (Figure 4) has a participant of the depicted event represented using the Real-Space subject and one (or both) of the manual articulators. This allows the signer to depict, say, someone bumping into someone else by creating a viewer blend to depict specific features of only the patient while simultaneously creating a diagrammatic blend to depict the bump itself. A different type of depiction would be produced if the event is about an experiencer rather than a patient, e.g. someone seeing the bumping. Figure 5 represents this type of depiction. The thought balloon represents the psychological (as opposed to physical) experience one is having. Another example of this type of depiction is the expression of perceived motion (Valli and Lucas 2000). Morgan (2005) describes the possibility of creating “overlapped referential spaces” (p. 125), the co-activation of SRS and FRS. It seems clear that this involves the partitioning of the manual articulators.

Not all events involve an animate participant. It is possible to create a viewer blend to depict unobserved events such as a lightning hitting a tree in a forest. Because there is no animate participant to represent, no |subject| would be activated. Since this is a viewer blend, the location of the signer necessarily participates in the depiction. This location is the Real-Space vantage point. There are many (virtually infinite) locations within the setting of the event from which to view the event. One of these are selected and blended with the Real-Space vantage point, resulting in a blended |vantage point|. Figure 6 represents this type of depiction, with the dotted

figure representing the |vantage point|. We have already described above (Figure 2) another type of event depiction that does not have a |subject|. Because this involves a diagrammatic blend, the Real-Space vantage point is not integrated into the blend. However, this element is of course essential to the creation and development of the diagrammatic blend. After all, it is the limited portion of the space *in front of the signer* where the depiction takes place. We have also considered the ability to conceive of events apart from any specific setting in which they occur. However, as suggested by Langacker (1991), there is a dependency events have: they necessarily take place within a setting. While we were able to come up with expressions in which events are depicted without reference to specific settings, we have not determined whether it was useful to make a distinction between event depictions involving specific settings and those involving schematic settings. Also yet to be determined is the usefulness of identifying event depictions involving the cognitive ability of expansion, as opposed to compression. We can see this in the depiction of events occurring at, say, a subatomic level.

The rest of the types of depiction that we are currently concerned with here are setting depictions. They are non-temporal counterparts to the non-subject event depiction types just mentioned. A viewer blend can be created to depict where objects are located within a setting—say a light fixture in a kitchen. A diagrammatic blend can be created to depict the location of furniture within a room. Features of an object can be depicted apart from a specific setting. For example, the legs of an intricately carved wooden chair can be depicted in front of the signer rather than closer to the floor. Smaller objects can be expanded in size for more efficient depiction.

Classifier predicates (or what Liddell 2003 calls “depicting verbs”) are a staple of depictions of objects, settings, and events. A discussion of how they relate to the types of depiction just described is not possible here, but suffice it to say that we view them (or their components) as being types of depiction themselves. For example, a verb that depicts a punch being thrown could be (but not always) considered to be an instance of a depiction involving a |subject|.

### 3. Depiction Identification and Annotation Procedures

One of our project’s aims is to develop depiction identification and annotation procedures to assist video coders in their work. Among the introductory materials currently being developed, we are completing a flowchart of the types of depictions described in Section 2. The flowchart includes yes-no questions that eventually lead to coding instructions. For example, at one point in the flowchart the coder is asked whether there are two distinct visible entities that are life-sized (an example of this depiction is one that describes the event from the patient’s viewpoint). If the brief description fits the type of depiction observed, then the coder is shown an illustration similar to those in the above section and is instructed to use a particular code. If the description does not fit, then

the coder is instructed to move on to the next description. The flowchart has three major sections: depictions involving |subjects|, event depictions without |subject|, and setting depictions. In all there are between 8 to 14 types of depiction that we are currently interested at this stage of the project.

We use ELAN to annotate depiction observed in video texts. We currently are working with two tiers. One tier will be used to annotate instances of |subject| blends. Different types of |subject| blends will have their own code, and we are also determining a convenient way to identify blends that have been reactivated rather than created anew, as has been observed in narratives where an event is depicted from the viewpoints of multiple event participants. Another tier will be used to annotate instances of event depictions without a |subject| and of setting depictions. There are two reasons for having these two tiers. First, there are types of depictions that appear to be possible only when a |subject| is activated, e.g. those depicting dialogue and perception. The second reason is more well-known and has been documented in Morgan (2005) and elsewhere: signers often “move” between spaces. One of the things that might happen here, as described from a conceptual blending viewpoint, is that the depiction effectively becomes a setting depiction when the signer stops depicting the event to add information via linguistic items, e.g. nouns, that do not depict anything.

Future work will examine other types of depiction, including tokens, depictions that employ metaphor, and tokens, leading towards a more complete typology of depiction. While we begin with the analysis of depiction in simple narratives and related genres, we will eventually work with discourse in other settings. Testing depiction identification procedures in the coding of signed language discourse in academic settings, etc., are likely to reveal issues requiring the revision or refinement of these procedures. We also plan to ensure coder validity of the identification procedures. Ultimately, we hope that these procedures can be used to identify all types of depiction observed to occur any signed language discourse.

### 4. References

- Dudis, Paul. 2007. Types of Depiction in ASL. MS.
- Emmorey, Karen and Falgier, Brenda. 1999. Talking about Space with Space: Describing Environments in ASL. In E.A. Winston (ed.), *Story Telling and Conversations: Discourse in Deaf Communities*. Washington, D.C.: Gallaudet University Press, 3-26.
- Fauconnier, Gilles & Mark Turner. 2002. *The Way We Think*. NY: Basic Books.
- Langacker, Ronald. 1999. *Foundations of Cognitive Grammar, Vol II.: Descriptive Application*. Stanford, CA: Stanford University Press
- Liddell, Scott K. 2003. *Grammar, Gesture and Meaning in American Sign Language*. Cambridge: Cambridge University Press
- Liddell, Scott K. 1995. Real, surrogate, and token space: Grammatical consequences in ASL. In Karen Emmorey and Judy Reilly (eds.), *Language, Gesture, and Space*.

- Hillsdale, NJ: Lawrence Erlbaum Associates, 19-41.
- Metzger, Melanie. 1995. Constructed Dialogue and Constructed Action in American Sign Language. In Ceil Lucas (ed.) *Sociolinguistics in Deaf Communities*. Washington, D.C.: Gallaudet University Press
- Morgan, G. (2005). Transcription of child sign language: A focus on narrative *Journal of Sign Language and Linguistics*, 8, 119-130
- Valli, Clayton and Ceil Lucas. 1991. *Predicates of Perceived Motion in ASL*. In Fischer, Susan and Patricia Siple (eds.) *Theoretical Issues in Sign Language Research*. Chicago: University of Chiago Press.
- Winston, E. A. (1991). Spatial referencing and cohesion in an American Sign Language text. *Sign Language Studies*, 1991, 73, Winter, 73(winter), 397-410.

## Annotation and Maintenance of the Greek Sign Language Corpus (GSLC)

**Eleni Efthimiou, Stavroula - Evita Fotinea**

Institute for Language and Speech Processing (ILSP) / R.C. Athena

Artemidos 6 & Epidavrou,

GR 151 25 Athens

Greece

E-mail: eleni\_e@ilsp.gr, evita@ilsp.gr

### Abstract

This paper presents the design and development of a representative language corpus for the Greek Sign Language (GSL). Focus is put on the annotation methodology adopted to provide for linguistic information and annotated corpus maintenance and exploitation for the extraction of a linguistic model intended to support both sign language recognition and creation of educational content.

### 1. Introduction

The Greek Sign Language (GSL) has developed as a minority non-written language system -in a socio-linguistic environment similar to those holding for most other known sign languages- used as the mother language of the Greek deaf community.

Video recordings of GSL have been produced for various reasons but, the development of the Greek Sign Language Corpus (GSLC) is the first systematic attempt to create a re-usable electronic language corpus organised and annotated according to principles deriving from requirements put by specific application demands (Mikros, 2004). The GSLC is being developed in the framework of the national project DIANOEMA (GSRT, M3.3, id 35) that aims at optical analysis and recognition of both static and dynamic signs, incorporating a GSL linguistic model for controlling robot motion. Linguistic analysis is a sufficient component for the development of NLP tools that, in the case of sign languages, support deaf accessibility to IT content and services. To effectively support this kind of language intensive operations, linguistic analysis has to derive from safe language data -defined as data commonly accepted by a specific language community- and also provide for an amount of linguistic phenomena, which allow for an adequate description of the language structure. The GSLC annotation features have been, however, broadly defined to serve multipurpose exploitation of the annotated part of the corpus. Different instantiation of corpus reusability are provided by measurements and data retrieval, which serve various NLP applications along with creation of educational content.

### 2. Development and maintenance of GSLC

#### 2.1 Corpus development

A definition of corpus provided by Sinclair (1996) in the framework of the EAGLES (<http://www.ilc.cnr.it/EAGLES>) project, runs as follows:

“A corpus is a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language”. Furthermore, the definition of computer corpus in the same document crucially states that: “A computer corpus is a corpus which is encoded in a standardised and homogenous way for open-ended retrieval tasks...”.

Here we will use the term corpus as always referring to an electronic collection of pieces of language, also adopting the classification by Atkins et al. (1991), which differentiates corpus from a generic library of electronic texts as a well defined subset that is designed following specific requirements to serve specific purposes. Among the most prominent purposes for which oral language (written) electronic corpora are created, lies the demand for knowledge management either in the form of information retrieval or in the form of automatic categorisation and text dispatching according to thematic category. Electronic corpora differentiate as to intended use and the design requirements that they fulfil.

The design of GSLC content has been led by the demand to support sign language recognition as well as theoretical linguistic analysis. In this respect, its content organisation makes a distinction between three parts on the basis of the utterance categories to be covered.

The first part comprises a list of lemmata which are representative of the use of handshapes as a primary sign formation component. This part of the corpus is developed on the basis of measurements of handshape frequency of use in sign morpheme formation, but it has also taken into account the complete set of sign formation parameters. In this sense, in order to provide data for all sign articulation features of GSL, the corpus also includes characteristic lemmata with respect to all manual and non-manual features of the language.

The second part of GSLC is composed of sets of controlled utterances, which form paradigms capable to expose the mechanisms GSL uses to express specific core

grammar phenomena. The grammar coverage that corresponds to this part of the corpus is representative enough to allow for a formal description of the main structural-semantic mechanisms of the language.

The third part of GSLC contains free narration sequences, which are intended to provide data of spontaneous language production that may support theoretical linguistic analysis of the language and can also be used for machine learning purposes as regards sign recognition.

All parts of the corpus have been performed by native signers under controlled conditions that guarantee absence of language interference from the part of the spoken language of the signers' environment (DIANOEMA Project, 2006a; 2006b), whereas quality control mechanisms have been applied to ensure data integrity.

## 2.2 Content selection

The initial target of sign recognition imposed the demand for the collection of lists containing representative lemmata, capable to exhibit the articulation mechanisms of the language. These lists may provide a reliable test bed for initial recognition of single articulation units. Lemmata lists comprising the first part of the GSLC involve two categories, (i) commands related to robot motion control and (ii) simple and complex sign morphemes, representative of the basic vocabulary of GSL.

Morpheme selection was based on the minimum requirement of handshape frequency of occurrence, that imposed use of at least the 15 most frequent handshapes, which are responsible for the formation of a 77% of the whole amount of lemmata met in the environment of primary school education (unpublished measurement, V. Kourbetis: personal communication). Both categories contained simple and complex signs, taking into account the use of either one, or two hand formations. Except for handshapes, all other articulation parameters have been taken into account in lemma content design. These parameters include the sets of manual and non-manual features of sign formation and involve location, palm orientation, movement of the hand as well as facial expressions and head and body movement (Stokoe, 1978).

Internal organisation of lemmata lists includes categorisation according to motion commands, location indicators, number formation, finger spelling, temporal indicators, various word families, GSL specific complex sign roots and the standard signing predicate categories.

The video-corpus contains parts of free signing narration, as well as a considerable amount of elicited grouped signed phrases and sentence level utterances, reflecting those grammar phenomena of GSL that are representative for the structural organisation of the language. Theoretical linguistic analysis of such data allows for extraction of

safe assumptions as regards the rule system of the language and also provides a safe ground for the use of phrase level annotation symbols.

When structuring the phenomena list that are represented by controlled sentence groups in the video-corpus, a number of GSL specific linguistic parameters were taken into account, with the target to capture the main multi-layer articulatory mechanisms the language uses to produce phrase/sentence level linguistic messages, along with distribution within utterances of a significant number of semantic markers for the expression of quantity, quality and schema related characteristics. The two parts of the video-corpus (free narration and controlled sentences per grammar phenomenon) function complementarily as regards the target of rule extraction for annotation purposes and machine learning for sign recognition.

The phenomena for which GSLC provides extensive paradigms (Efthimiou, Fotinea & Sapountzaki, 2006) include the GSL tense system with emphasis on major temporal differentiations as regards present, past and future actions in combination with various aspectual parameters, multi-layer mechanisms of phrase enrichment for the expression of various adverbial values in phrase or sentence level, the use of classifiers, affirmation with all types of GSL predicates, formations of negation, WH- and Yes/No question formation, various control phenomena and referential index assignment.

In order to receive unbiased data, a strict procedural rule was to avoid any hint to natural signers as to preference in respect to sentence constituents ordering. In cases of deviation from neutral formations as when expressing emphasis, instructions to informants focused on the semantic dimension of the tested sentence constituent, rather than on possible structural arrangements of the relevant utterances. Furthermore, with the general aim to eliminate external destructions (such as environment language interference), the use of written Greek was excluded from communication with the natural signers.

## 2.3 Evaluation of the video-corpus

In order to ensure prosodic and expressive multiplicity, it has been decided to use at least 4 signers for the production of GSLC in all three parts of the corpus content. The selection of natural signers has been based on theoretical linguistics criteria related to mother language acquisition conditions (White, 1980; Mayberry, 1993). Signers chosen to participate in GSLC production should, hence, be deaf or bilingual hearing natural GSL signers, raised in an environment of deaf natural signers. This selection criterion strictly excludes the use of deaf signers that are not natural GSL signers, in order to ensure the highest degree of linguistic integrity of the data, and, at the same time, eliminate –if not completely make vanish of– the language interference effects from Greek to GSL throughout the development of the video-corpus.

Upon completion of the GSLC video recording, uninformed quality control procedures have been followed targeting at high degrees of acceptance of the video-recorded signing material. Each part of the video-corpus had to be evaluated by natural signers, on the basis of peer review, with respect to intelligibility of the linguistic message. In case a video segment was judged poorly, the segment had to be re-collected and re-evaluated, hence, ensuring that only highly judged video segments are included in the GSLC.

### 3. Corpus annotation

#### 3.1 Morpheme level annotation

. Technological limitations regarding annotation tools often impeded the use of data synchronised with video. The situation has slowly started to change as, at an experimental level, open tools have been started to develop to suit the needs of sign language annotation. Research projects, as the European ECHO (<http://www.nmis.isti.cnr.it/echo>) (2000-2004) and the American SignStream (<http://www.bu.edu/asllrp/SignStream/>) of the National Center for Sign Language and Gestures Resources (Boston University, 1999-2002) (Neidle, 2002) produced video-corpora that complied to a common set of requirements and conventions. Tools such as the iLex (Hanke, 2002) attempt to solve issues related to convention integrity of data, arising from the lack of a writing system which follows orthographic rules. In the same context, the Nijmegen Metadata Workshop 2003 (3. Crasborn, & Hanke, 2003) proposed a common set of metadata for use by sign language video-corpora.

The definition of annotation features assigned to a given signing string, reflects the extent of the desired description of grammatical characteristics allotted to the 3-dimensional representation of the linguistic message. Basic annotation fields of GSLC involve glosses for Greek and English, phrase and sentence boundaries, dominant and non-dominant hand information, eye-gaze, head and body movement and facial expression information, as well as grammar information such as tags on signs and grammar phenomenon description to facilitate data retrieval for linguistic analysis.

Starting from the need for theoretical linguistic analysis of minimal grammatically meaningful sign units, as well as the description of articulation synthesis of basic signs, the term sign morpheme has been adopted to indicate the level of grammatical analysis of all simple sign lemmata.

For the annotation of the video-corpus at the morpheme level, the basic phonological components of sign articulation, for both manual and non-manual features, have been marked on a set of representative simple morphemes and complex signs. For the representation of the phonological characteristics of the basic morphemes the HamNoSys (Hamburg Sign Language Notation

System, <http://www.sign-lang.uni-hamburg.de/projects/HamNoSys.html>) annotation system is used (Prillwitz et al., 1989)

The characteristics of sign articulation are (sometimes dramatically) modified when moving from lemma list signing to phrase construction, where prosodic parameters and various grammar/agreement markers (i.e. two-hands plural) impose rendering of lemma formation, subject to phrase articulation conditions. Hence, recognition systems have to be taught to correctly identify the semantics of lemmata incorporated in phrase formations. Furthermore, accurate morpheme level annotations serve sign synthesis systems that have to produce utterances with the highest possible level of naturalness.

#### 3.2 Sentence level annotation

Fully aligned with the phenomena list composing the controlled sentence groups of GSLC content, phrase level annotation focuses on coding the basic mechanisms of multi-layer articulation of the sign linguistic message and distribution of the most important semantic markers for the indication of qualitative, quantitative and schematic values. Both multi-layer articulation and semantic deixis are major characteristics of sign phrase articulation, whereas in the context of free narration, one major demand is the correct assignment of phrase boundaries.

Some of the most representative phrase level phenomena of GSL concern multi-layer articulation over one temporal unit that results in modification of the basic components of the sign phrase (Efthimiou, Fotinea & Sapountzaki, 2006). In the context of a nominal phrase, this is related to i.e. adjectival modification. The same holds for the articulation of predicative and nominal formations, which incorporate classifiers, or when providing tense indicators. A different type of phrasal annotation is adopted to indicate topicalisation of a phrase, irrespective of its grammatical category.

Sentence level annotation aims at providing for reliable extraction of sentence level structure rules, incorporating basic multi-layer prosodic articulation mechanisms, question formation and scope of quantification and negation.

For the safe use of GSLC, a subset of sentences, which are representative for all phenomena contained in the corpus, have been manually annotated. In free narration parts, sign utterance boundaries are manually marked according to generally accepted temporal criteria (segmentation boundary is set at the frame where the handshape changes from the last morpheme of the current signing string to the first morpheme of the next) and according to annotators' language feeling.

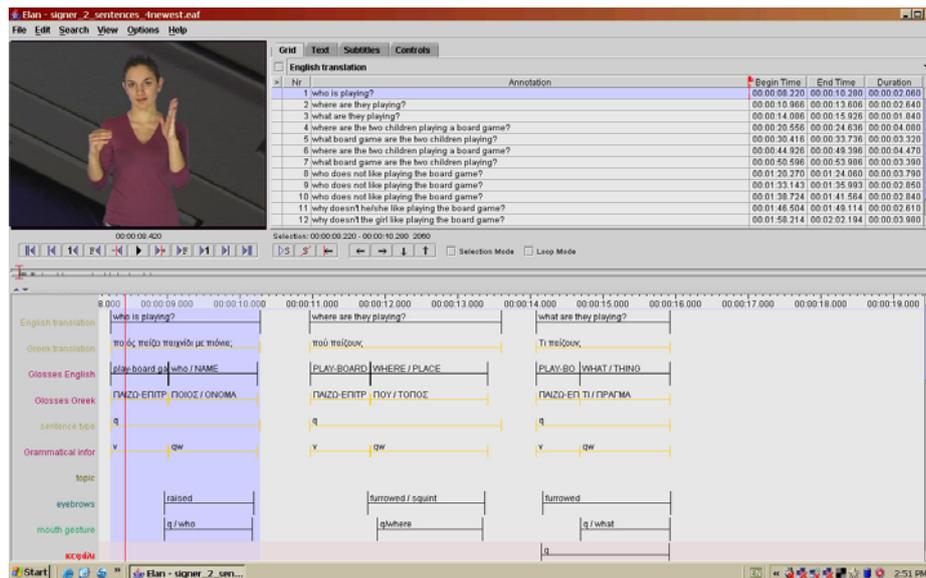


Figure 1: Annotation and retrieval of WH-question data in GSL.

The chosen annotation system is ELAN (Eudico Linguistic Annotator) the key characteristics of which are in a nutshell summarised next. ELAN (version 2.6) is an annotation tool that allows creation, editing, visualisation and retrieval of annotations for video and audio data, aiming at providing a sound technological basis for the annotation and exploitation of multi-media recordings. Figure 1 provides an instantiation of the GSLC annotation and retrieval procedure.

### 3.3 Evaluation of the annotated corpus

Assignment of annotations to GSLC involves two expert GSL annotators with expertise in sign language linguistics and sign language technological issues.

Annotation quality control is based on peer-review with annotation control on sample video-corpus parts, on a mutual basis by the expert annotators. Additionally, one external GSL expert annotator executes peer sample quality control on the whole annotated video-corpus. The parts of the annotated video-corpus for which conflicting evaluation reports are provided, are discussed among the three evaluators resulting in a commonly approved annotation string that is finally taken into account.

## 4. Exploitation of the annotated corpus

### 4.1 Extraction of measurements for sign recognition

In the context of DIANOEMA project, a linguistic model had to be extracted from GSLC, aiming to enhance recognition results as regards possible ambiguity or misclassified components. The linguistic model was the result of various measurements and of those parameters which formulate them as, for example, the total duration of annotated video with signing data, the set of annotation

tiers, the number of lemmata which have been assigned some feature, or the set of features been assigned.

The phenomena of interest were identified and various retrieval procedures were applied in the annotated corpus in order to collect a representative sample of their instantiations. Measurements of occurrences of the different instantiations of a phenomenon allowed for mapping conditions, which rule its different realizations. As a consequence, it was possible to evaluate most productive mechanisms of utterance and incorporate them to the linguistic model intended to perform smoothing of the recognition outcome.

The various retrieval operations performed on the total duration of the annotated corpus, took into account the whole set of annotation parameters (27 ELAN tiers) and assigned features. Files of occurrences of phenomena were created which often provided a demonstration of their realization, significantly deviating from commonly accepted options, the latter usually based on a limited set of data. Valuable use demonstrations were provided for phenomena such as the use of pronominal indices, negation, question and plural formation.

An example of how the linguistic model was constructed, is provided by the measurements output, which defined the options for plural formation in GSL. The vast majority of plural signs made use of classifiers to indicate plurality. The next most common option was to exploit location indices, where two-handed plural and repetition for plural formation (appreciated among the standard rule options) were left far beyond the top, followed only by the very rare occurrences of numeral and index based plural formations.

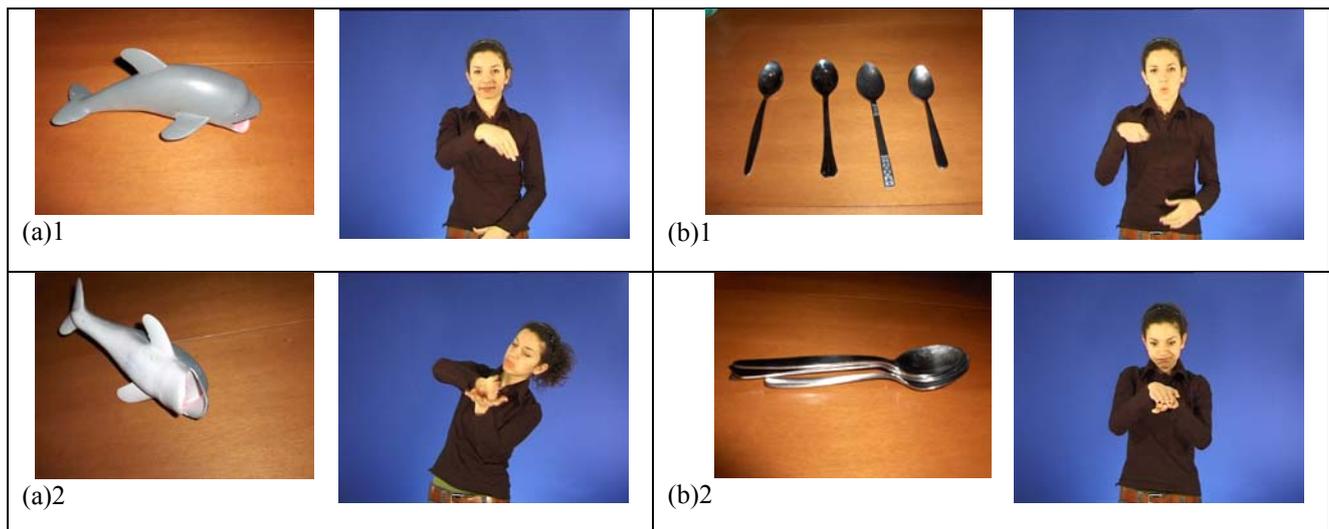


Figure 2: Icon driven classifier productions of GSL: (a) dolphin swimming (1), dolphin lying on flat surface (2); (b) spoons in a row (1), stacked spoons (2).

#### 4.2 Linguistic model for GSL classifiers

A specific part of the elicited corpus was devoted to the use of classifiers in GSL. In order to drive the informants to use a wide range of classifiers, different sets of stimuli were organised so as to cover the range of semantic properties assigned to base signs by use of appropriate classifiers. Elicitation focused on quantity, quality and spatial properties. The means to derive linguistic data were appropriate sets of icons, free discussion and story telling stimulated by film display.

The so derived data have been classified according to semantic indicator and are further elaborated in order to be incorporated in an educational environment as GSL grammar content. In Figure 2 it is demonstrated how icon driven classifier productions were derived. Example (a) demonstrates the use of flat B classifier to indicate the surface onto which a dolphin lies (2) opposite to the use of the sign for dolphin in the default case (1). Example (b) arranges spoons in a row repeatedly locating the handshape for spoon in the signing space (1), while in (2) a stack of spoons is indicated by a two hand formation of the flat B classifier.

#### 5. Concluding remarks

The current state-of-the-art on technological advances and the open scientific issues related to sign language technologies have brought about the significance of annotated corpora for decoding the various aspects of sign language articulation message.

An appropriately annotated sign language corpus may provide a re-usable source of linguistic data to be exploited in the environment of sign language technologies but also in diverse situations as incorporation of SLs in various Natural Language Processing (NLP) environments or the creation of

language teaching educational content. In this sense, an annotated corpus is essential to the development of sign recognition systems and also to the creation of adequate language resources such as lexical databases and electronic grammars needed in the context i.e. of Machine Translation. Language resources being equally crucial for the development of sign synthesis machines and conversion tools from spoken to sign language that often drive sign synthesis machines, underline the usability of a corpus which supports extraction of both reliable measurements and linguistic data.

GSLC design and implementation have equally focused on sign recognition support and on the extraction of a linguistic model for GSL. GSLC extensibility is intrinsically foreseen as regards both its content and adopted annotation features. This allows for corpus re-usability in linguistic research and sign language technology applications.

#### 6. Acknowledgements

This work has been partially funded by the European Union, in the framework of the Greek national project DIANOEMA (GSRT, M3.3, id 35).

#### 7. References

- Atkins, S., Clear, J. & Ostler, N. (1991). Corpus design criteria. *Literary and Linguistic Computing*, Vol. 7 pp.1--16.
- Bowden, R., Windridge, D., Kadir, T., Zisserman, A. & Brady, M. (2004). A Linguistic Feature Vector for the Visual Interpretation of Sign Language. In *Tomas Pajdla, Jiri Matas (Eds), Proc. 8th European Conference on Computer Vision, ECCV04. LNCS3022*, Springer-Verlag, Volume 1, pp. 391--401.
- Bellugi, U. & Fischer, S. (1972). A comparison of Sign language and spoken language: rate and grammatical mechanisms. *Cognition: International Journal of*

- Cognitive Psychology*, 1, pp. 173--200.
- Crasborn, O., Hanke, T. (2003). Metadata for sign language corpora. Available on-line at: [http://www.let.ru.nl/sign-lang/echo/docs/ECHO\\_Meta\\_data\\_SL.pdf](http://www.let.ru.nl/sign-lang/echo/docs/ECHO_Meta_data_SL.pdf).
- DIANOEMA Project (2006a), Composition features of the GSL Corpus, WP1-GSL Corpus, *Technical Report* (in Greek).
- DIANOEMA Project (2006b). Annotation Features of GSL, WP3- Language model / video annotations of GSL, *Technical Report* (in Greek).
- Efthimiou, E., Sapountzaki, G., Karpouzis, C. & Fotinea, S-E. 2004. Developing an e-Learning platform for the Greek Sign Language. *Lecture Notes in Computer Science* 3118: pp. 1107--1113. Springer.
- Efthimiou, E., Fotinea, S-E. & Sapountzaki, G. 2006. Processing linguistic data for GSL structure representation, In *Proceedings of the Workshop on the Representation and Processing of Sign Languages: Lexicographic matters and didactic scenarios, Satellite Workshop to LREC-2006 Conference*, May 28, pp. 49--54.
- ELAN annotator, Max Planck Institute for Psycholinguistics, available at: <http://www.mpi.nl/tools/elan.html>
- Fotinea, S-E., Efthimiou, E., Karpouzis, K. & Caridakis, G. 2005. Dynamic GSL synthesis to support access to e-content, In *Proceedings of the 3rd International Conference on Universal Access in Human-Computer Interaction (UAHCI 2005)*, 22-27 July 2005, Las Vegas, Nevada, USA.
- HamNoSys Sign Language Notation System: [www.sign-lang.uni-hamburg.de/projects/HamNoSys.html](http://www.sign-lang.uni-hamburg.de/projects/HamNoSys.html)
- 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*. Paris : ELRA pp. 923--926
- Karpouzis, K. Caridakis, G., Fotinea, S-E. & Efthimiou, E. 2007. Educational Resources and Implementation of a Greek Sign Language Synthesis Architecture, *Computers and Education*, Elsevier, Volume 49, Issue 1, August 2007, pp. 54--74, electronically available since Sept 05.
- Kraiss, K.-F. (Ed.), 2006. Advanced Man-Machine Interaction - Fundamentals and Implementation. *Series: Signals and Communication Technology*, Springer.
- Mayberry, R. (1993). First-Language acquisition after childhood differs from second language acquisition: The case of American Sign Language. *Journal of Speech and Hearing Research*, Vol. 36 , pp. 51--68
- Mikros, G. (2004). Electronic corpora and terminology. In *Katsoyannou, M., and Efthimiou, E., (eds) Terminology in Greek: Research and Implementation Issues*. Kastaniotis publications, Athens (in Greek).
- Neidle, C (2002). SignStream™ Annotation: Conventions used for the American Sign Language Linguistic Research Project. Boston, MA: *American Sign Language Linguistic Research Project Report No. 11*, Boston University.
- Prillwitz, S., Leven, R., Zienert, H., Hanke, T. and Henning, J. (1989). HamNoSys. Version 2.0. Hamburg Notation System for Sign Language. An Introductory Guide.
- Sinclair, J. (1996). Preliminary recommendations on corpus typology. EAGLES Document EAG--TCWG--CTYP/P, electronically available at: <http://www.ilc.cnr.it/EAGLES/corpusstyp/corpusstyp.html>.
- Stokoe, W. (1978). Sign Language Structure (Revised Ed.). *Silver Spring, MD: Linstok*
- White, L. (1980). Grammatical Theory and Language Acquisition. *Indiana University Linguistic Club*.

# **iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography**

**Thomas Hanke, Jakob Storz**

Institute of German Sign Language and Communication of the Deaf  
University of Hamburg  
Binderstraße 34, 20146 Hamburg, Germany  
E-mail: {thomas.hanke,jakob.storz}@sign-lang.uni-hamburg.de

## **Abstract**

This paper presents iLex, a software tool targeted at both corpus linguistics and lexicography. It is now a shared belief in the LR community that lexicographic work on any language should be based on a corpus. Conversely, lemmatisation of a sign language corpus requires a lexicon to be built up in parallel. We introduce the basic concepts for transcription work in iLex, especially the interplay between transcripts and the lexicon.

## **1. Background**

For empirical sign language research, the availability of Language Resources, their quality as well as the efficiency of software tools to create new resources is a pressing demand. The software solution iLex is our approach to meet these requirements at least to a certain extent: It is a database system to make existing resources available, and it is a tool to create new resources and to manage their quality.

Language resources for sign languages are special insofar as there is no established writing system for any sign language in the world. Notation systems can only partially fill this gap, and their most important drawback is the effort needed to describe signed utterances in enough detail that would allow the researcher to do without going back to the original data. In the early 1990ies, syncWRITER (Hanke & Prillwitz, 1995; Hanke, 2001) was our first attempt for a transcription tool that not only allowed the user to link digital video sequences to specific parts of the transcription, but also allowed the video to become the skeleton of the transcription. The drawback of that solution was that it was mainly targeted towards the presentation of the transcriptions in a graphically appealing way, but was not equally well equipped for any discourse-analytic or lexicographic purpose.

In the context of a series of special terminology dictionaries, we therefore developed an independent tool, GlossLexer (Hanke et al., 2001), concentrating on the development and production of sign language dictionaries, both in print and as multimedia hypertexts, derived from transcriptions of elicited sign language utterances. At the heart of this tool was a lexical database, growing with the transcriptions. This tool, however, was not suitable to adequately describe really complex signed utterances, as it reduced them to sequences of lexical entities as suitable only in a purely lexicographic approach.

iLex (short for “integrated lexicon”, cf. Hanke, 2002b)

now combines the two approaches: It is a transcription database for sign language in all its complexity combined with a lexical database. In iLex, transcriptions do not consist of sequences of glosses typed in and time-aligned to the video. Instead, transcriptions consist of tokens, i.e. exemplars of occurrences of types (signs) referencing their respective types. This has immediate relevance for the lemmatisation process. Due to the lack of a writing system, this is not a relatively straightforward process as for spoken languages with a written form featuring an orthography, but requires the transcriber’s full attention in type-token matching.

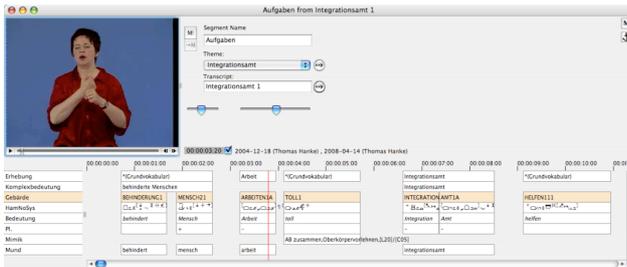
By providing tool support for this process, iLex enables larger and multi-person projects to create transcriptions with quality measures including intra-transcriber and inter-transcriber consistency.

For a research institute as a whole, the central multi-user database approach means that all data are available at well-defined places, avoiding data loss often occurring in a document-centric approach as researchers and students leave and enabling an effective data archiving strategy. Finally, combining data from several projects often is the key to achieve the “critical mass” for LR-specific research.

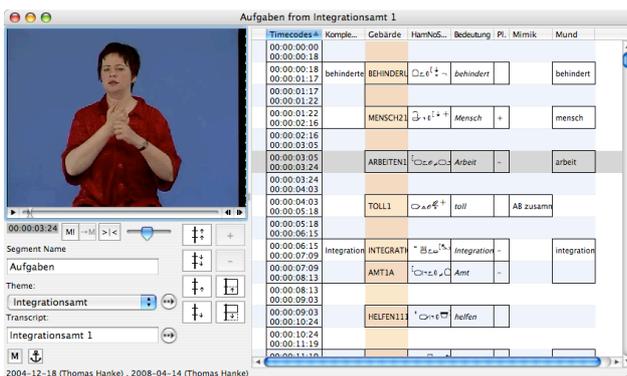
At the IDGS in Hamburg, iLex today is not only used in discourse analysis and lexicography, but a number of applied areas draw from the data collected and contribute themselves: The avatar projects ViSiCAST and eSIGN allow transcripts from the database to be played back by virtual signers (Hanke, 2002a; Hanke, 2004a); in computer-assisted language learning for sign languages, authoring tools can directly import iLex transcripts (Hanke, 2006).

## **2. Flow of Time**

iLex features a horizontal view of transcript data familiar to those using any other transcription environment: Time flows from left to right, and the length of a tag is proportional to its duration.



This view is complemented by a vertical view, where time flows from top to bottom. Each smallest interval of interest here occupies one row, irrespective of its length. A tag spans one or more such intervals. Unless it is partially overlapping with other tags, the tag is identical to one interval. The focus here is on interesting parts of the transcription, not on the flow of time. If the transcriber detects that two events are not fully cotemporal, but that one starts slightly after the other, for example, the time interval that the two tags have shared so far is split at the point of time where the second event really starts, and the second tag's starting point is moved down one line. This procedure ensures that slightly deviating interval boundaries are possible, but only as a result of a deliberate action by the user.



Which of these two views is used is determined by the current task, but also the user's preference. In any case, switching to the other view sheds new light on the transcription and thereby helps to spot errors.

### 3. A Data Model for Transcripts

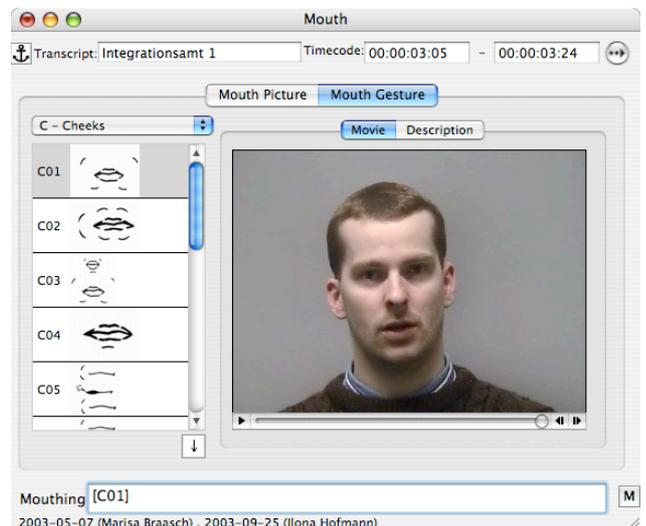
Despite the fact that iLex is the only transcription tool used in sign language research with a database instead of a document-centric approach, the data model for transcripts is more or less shared with other tools<sup>1</sup>: Transcripts are linked to a video<sup>2</sup> and have any number of tiers; a tier contains tags that are time-aligned to the video. Tier-to-tier relations define restrictions on the alignment of tags with respect to tags in superordinate tiers. However, iLex goes beyond this by introducing

<sup>1</sup> As the other systems, iLex's data model can be considered an implementation of the annotation model developed by Bird and Liberman (2001).

<sup>2</sup> iLex transcripts can link to only one "movie". This is no restriction, as iLex works well with movies containing more than one video track. At any point of time, the user can choose to hide tracks s/he is not currently interested in, e.g. close-up views that will only be used in mouthing or facial movements analysis.

different kinds of tiers. The most important kinds are:

- *Token tiers* contain tokens as tags, i.e. they describe individual occurrences of signs and as such are the most important part of a transcription. iLex allows double-handed and two-handed tokens, or partially overlapping one-handed tokens, but always ensures that the tokens at any point of time do not describe more than two hands per informant.
- In elicitation settings, *answer tiers* group tokens that are signed in response to a specific elicitation, describing the elicitation by referring to a picture, movie segment or text.
- Tags in *phrase structure tiers* group tokens into constituents or multi-sign expressions.
- Tags in *text tiers* simply have text labels. This is the kind of tags found in most other transcription environments. iLex allows the user to assign vocabularies to tiers, so that tags can be chosen from pre-defined lists of values. User-defined vocabularies can be open or closed, but iLex also offers a number of built-in vocabularies with special editors, e.g. in order to tag mouth gestures.

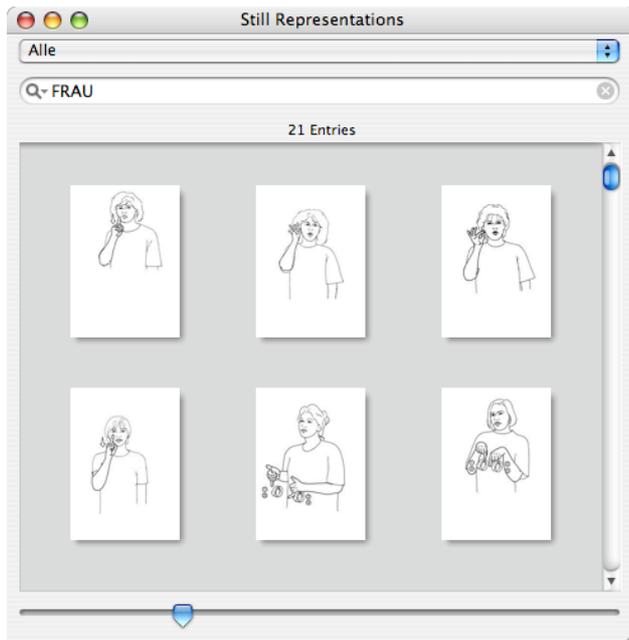


- Tags in *numerical data tiers* can be linked to horizontal and vertical coordinates in the movie frame. Thus, the user can enter data for these tags by clicking into the movie frame, e.g. to track the position of the eye or to measure distances. Tags could also be automatically created by external image processing routines indicating e.g. a likelihood for certain types of events, as a first step to semi-automatic annotation.
- Tags in *value (computed) tiers* are automatically inserted by the system as the user enters data into other tiers. E.g. a tier can be set up to show the citation form of the types referenced by tokens in another tier, in our case by means of a HamNoSys notation (Hanke, 2004b).

As with most database entities in iLex, the user can easily add metadata to transcripts, tiers, and tags. These may be ad-hoc comments, markers for later review, judgements, or structured data as defined by the IMDI metadata set or its extension for sign language transcription (cf. Crasborn & Hanke, 2003).

#### 4. Lemmatisation

Type-token matching is at the heart of transcribing with iLex, and iLex supports the user in this task. The user can identify candidates for the type to be related to a token by (partial) glosses, (partial) form descriptions in HamNoSys or meaning attributions. The search can be narrowed down by browsing through the types found, comparing tokens already assigned to a type with the token in question. By using alternatives such as browsing tokens or stills, an active competence in HamNoSys (or another notation system used in iLex instead) is not necessary.



Once the right type has been identified, it can easily be dragged into the transcript to establish the token. This procedure avoids simple errors such as typos, and allows for easy repairs. If it is later decided that a type needs to be split into several as form variation seems not to be free, tokens can be reviewed and reassigned (i.e. dragged into the new type) as necessary.

In the token, iLex used to provide a text to describe how the actual instance of the sign deviated from the citation form. The latest version categorises modifications in order to further reduce inconsistent labelling in this part as well.

#### 5. Importing Data from other Transcription Systems

Importing transcripts from other sources, such as ELAN, syncWRITER or SignStream documents (cf. Crasborn et al., 2004; Neidle, 2001), is done by a simple menu command. The results of this import process, however, are transcripts with only text tiers, and a second step is necessary to convert the text tiers describing tokens (in most cases by means of glosses) to real token tiers. iLex supports this process by learning a source-specific mapping table from external glosses to types and modifications in iLex. As inconsistencies may occur in the imported data if lemmatisation was not done rigidly, the transcriber's attention is required. More than one name for a single type is easily dealt with in the mapping mechanism. Different types under the same gloss label,

on the other hand, require close inspection of each token assigned.

#### 6. Dictionary Production

In the case of our special terminology dictionaries (cf. König et al., this volume), all of the data needed to produce the dictionary are stored in the database as the results of the transcription process or later analysis steps. This allows automatic production of a dictionary within reasonable time. For that, we use Perl as a scripting language linking the database with Adobe Indesign for laying out the print product and an HTML template toolkit to produce web applications. By just changing the templates (or adding another set), we can completely change the appearance of the dictionary and reproduce print and online versions within hours. Currently, we are developing another set of templates to optimise HTML output for iPhone/iPod touch devices that promise to become an ideal delivery platform for our dictionaries.

#### 7. Collaborative Approach

Using a central database for all people working in a project or even several projects at one institution not only serves data sustainability, but also allows for cooperative work. First and foremost, each transcriber contributes to the pool of types as well as tokens for each type making type-token matching easier or at least better informed. Other data, such as project-specific data views or filters, are easily shared between users. The results of introspection can quickly be made available to other users by using a webcam. Integration of camera support into the program allows sharing signed samples without the need to care about technical aspects such as video compression; appropriate metadata for the new video material is automatically added to the database.

The newest version of iLex takes a first step in supporting Web 2.0 technologies for collaboration: All data can be referenced by URLs. By simply dragging data from an iLex window into a Wiki or Blog, the URL is inserted and anyone with access to the iLex database can view the data talked about in a discussion by simply clicking onto the URL.

The "disadvantage" of collaboration of course is the need to agree on certain transcription conventions. While many aspects of the transcription process can be individualised, other data, such as the types inventory, need to be accessed by all users, and therefore need to be understood by all users; extensions need to be made in a consistent manner. Experience shows that a couple of meetings with all transcribers are needed if a new project is set up to work with the pool, especially if the new project's target differ significantly from what the other projects do.

#### 8. Technical Background

The name iLex stands for the transcription database as well as the front-end application used to access it.

The database normally resides on a dedicated or virtual database server. As the SQL database engine, we have chosen PostgreSQL, an open-source database server system that can be installed on a wide variety of

platforms.<sup>3</sup> It is rock-solid and has well-defined security mechanisms built in, it is well supported by an active user community, and features a couple of implementation aspects that are advantageous in our context, such as server-side inclusion of scripting languages including Perl.

Movies, stills and illustrations are not stored in the database, but only references to them. They can either reside on the users' computer or on a central file server. With video archives becoming rather large over time, of course only the second solution is viable in the long run.<sup>4</sup> This hybrid storage concept also allows users to work from home: Access to the database is low-bandwidth and therefore can be secured with a virtual private network approach, whereas the user can locally access the video currently in work without a performance hit. Tokens from other videos not available on the local computer then come over the network, but usually are that short that even slower connections should be fine.

The front-end software is available free of charge for MacOS X as well as Windows XP (with a couple features only available for MacOS), with German and English as user interface languages. (Localisation to other languages is easily possible.) Upon request, source code for the front end is also available except for a couple of functions where we decided to use commercial plug-ins instead of implementing the services ourselves. For single-user applications, the server and the client can be installed on the same machine, even on a laptop. However, unless that machine has plenty of RAM, page swapping will reduce the processing speed compared to a standard server-client scenario.

## 9. References

- Bird, S and M. Liberman (2001). A formal framework for linguistic annotation. *Speech Communication* 33(1,2), pp. 131–162.
- Crasborn, O. and T. Hanke (2003). *Metadata for sign language corpora*. Available online at: [http://www.let.ru.nl/sign-lang/echo/docs/ECHO\\_Metadata\\_SL.pdf](http://www.let.ru.nl/sign-lang/echo/docs/ECHO_Metadata_SL.pdf).
- Hanke, T. (2001). Sign language transcription with syncWRITER. *Sign Language and Linguistics*. 4(1/2), pp. 275–283.
- Hanke, T. (2002a). HamNoSys in a sign language generation context. In R. Schulmeister and H. Reinitzer (eds.), *Progress in sign language research: in honor of Siegmund Prillwitz / Fortschritte in der Gebärdensprachforschung: Festschrift für Siegmund Prillwitz*. Seedorf: Signum, pp. 249–266.
- Hanke, T., (2002b). iLex - A tool for sign language lexicography and corpus analysis. In: M. González Rodríguez, Manuel and C. Paz Suarez Araujo (eds.): *Proceedings of the third International Conference on Language Resources and Evaluation, Las Palmas de Gran Canaria, Spain*. Paris: ELRA, pp. 923–926.
- Hanke, T. (2004a). *Lexical sign language resources – synergies between empirical work and automatic language generation*. Paper presented at LREC 2004, Lisbon, Portugal.
- Hanke, T. (2004b). HamNoSys - Representing sign language data in language resources and language processing contexts. In: O. Streiter and C. Vettori (eds.): *Proceedings of the Workshop on Representing and Processing of Sign Languages, LREC 2004, Lisbon, Portugal*, pp. 1–6.
- Hanke, T. (2006). Towards a corpus-based approach to sign language dictionaries. In: C. Vettori (ed.), *Proceedings of a Workshop on the representation and processing of sign languages: lexicographic matters and didactic scenarios, LREC 2006, Genova, Italy*, pp. 70–73.
- Hanke, T. and S. Prillwitz (1995). syncWRITER: Integrating video into the transcription and analysis of sign language. In: H. Bos and T. Schermer (eds.), *Sign language research 1994: Proceedings of the fourth European congress on sign language research, Munich, Germany*. Hamburg: Signum, pp. 303–312.
- Hanke, T., R. Konrad and A. Schwarz (2001). GlossLexer – A multimedia lexical database for sign language dictionary compilation. *Sign Language and Linguistics* 4(1/2), pp. 161–179.
- Neidle, C. (2001). SignStream™: A database tool for research on visual-gestural language. *Sign Language and Linguistics*. 4(1/2), pp. 203–214.

<sup>3</sup> At the IDGS, we currently use a dedicated four-cores Mac Pro with 6 GBytes of memory and a mirrored harddisk. At some times, as many as 20 persons access the server without any experiencing any performance reductions.

<sup>4</sup> At the IDGS, we use a dedicated MacOS X Server file server with a storage area network (current size: 8 TB). We have experimented with video streaming servers before, but found that users rarely view more than a couple of seconds of a movie at once. In this situation, the negotiation overhead associated with streaming costs more than the streaming itself saves.

# Sign language corpora and the problems with ELAN and the ECHO annotation conventions

**Annika Herrmann**

University of Frankfurt am Main

Varrentrappstr. 40-42

60486 Frankfurt

E-mail: [herrmann@lingua.uni-frankfurt.de](mailto:herrmann@lingua.uni-frankfurt.de)

## Abstract

Corpus projects require logistic, technical and personal expertise and most importantly a conventionalized annotation system. Independently of its size, each project should use similar technical methods and annotation conventions for comparative reasons. To further enhance a unified conventionalization of sign language annotation, this paper addresses problems with ELAN annotation and the ECHO transcription conventions, shows imprecise usage examples and focuses on possible solutions. While building a corpus for a cross-linguistic sign language project in Germany, Ireland, and the Netherlands, various issues arose that ask for clarification. An appropriate time span annotation of signs is discussed as well as the need for a clear distinction of separate tiers. I will give transcription proposals for pointing/indexical signs and so called poly-componential or classifier constructions. Annotation should be as a-theoretical as possible without losing descriptive accuracy. In addition, I argue for a meticulous annotation of the eye gaze tier, as this is necessary for an adequate prosodic analysis. Finally the paper will show the usefulness of an additional tier to specify non-manuals that are concerned with adverbial, attitudinal and expressive facial expressions. The paper contributes to the important process of conventionalizing linguistic sign language annotation and the coding of signed video data.

## 1. Introduction

Large corpus projects with sign language data have recently received special attention. Sign languages are particularly endangered languages, as the social and cultural situation with regard to language acquisition and medical issues is a complex matter. In addition, linguistic research on languages in the visual-gestural modality and also cross-linguistic studies of sign languages world-wide, can give remarkable insights in the nature of language and cognition in general. Therefore, the documentation and preservation of signed data, either natural or elicited, is of enormous importance. However, relatively small corpus projects that investigate specific research issues and rely on a definite set of data can also be an invaluable contribution to linguistic sign language research. All these projects have to transcribe the video data and break down the visual signing stream into units that are evaluable and therefore available for analysis. This should be done in a comparable way for all sign languages and all projects. Sign language annotation conventions have not yet been uniformly developed on an international level, let alone been conventionalized for a European community. In an attempt to unify annotation conventions for sign languages the paper contributes to an ongoing standardization process and builds upon the ECHO annotation conventions, which proved to be well selected and highly sophisticated. These conventions evolved from the ‘Case Study 4: sign languages’ project, which is part of ECHO (European Cultural Heritage Online)<sup>1</sup> and since then became more and more established.

This paper elaborates on possible solutions for technical sign annotation and specifically looks at problematic cases of certain sign language constructions that

challenge an a-theoretical and cross-linguistic annotation of video data. To guarantee a most effective usage of search tools across various corpora a number of regulations and standards should be maintained and followed consistently.

The paper intends to stipulate clearly how to annotate specific aspects of signing and how to clarify some vague and problematic cases, constructions and components. In chapter 2 I will give some short introductory remarks about the project, the participants and the technical methodology. The following section (chapter 3) summarizes some important aspects of the annotation tool that is used and lists examples from the ECHO annotation system. Section 4 provides the core part of the paper and discusses specific annotation problems in six different paragraphs. I will address issues like time span annotation, accuracy of tiers that deal with eye gaze or aperture and indexical signs. With regard to comprehensive conventions, I will also give suggestions how to cope with the so called classifier constructions and also argue for the inclusion of an additional tier for specific non-manuals. After some short supplementary remarks, a last section giving an outlook (chapter 5) will conclude the paper.

## 2. The project

The subject of the dissertation project that I am currently working on in Germany, Ireland and the Netherlands requires the elicitation of specific signed sentences, contexts and dialogues. Therefore, I decided to create an annotated sign language video corpus for my own studies to guarantee comparative analysis. The study investigates how speaker’s attitude and focus particles are realized in sign languages (cf. Herrmann, 2007). In this project, data from three European sign languages (DGS, ISL and

<sup>1</sup> See <http://echo.mpiwg-berlin.mpg.de> for more information about the ECHO project in general.

NGT)<sup>2</sup> and altogether 20 native signers yield a set of over 900 sentences and short dialogues. Two video camcorders are used to provide a torso perspective as well as a smaller frame view showing the face of the respective signer. This facilitates annotation and is particularly important for research with regard to non-manual facial expressions.

The metadata information about participants and the recording situation will be edited along the lines of the IMDI metadata set (cf. Crasborn & Hanke, 2004), but cannot claim to be complete. The ELAN tool (Eudico Linguistic Annotator<sup>3</sup> provides the most adequate annotation software for my purposes, especially because one of the main interests of the study lies in the use of non-manuals. This annotation tool from the MPI in Nijmegen<sup>4</sup> is widely used for sign language annotation, but is mostly distributed in Europe. See Neidle (2001) and references for information on a different, but similar sign language annotation tool from the ASLLRP group, namely *SignStream*. Hanke (2001) presents the interlinear editor *syncWRITER*, but also shows that this software is not well-suited for large scale corpus projects.

Besides working with ELAN, I try to ensure comparability by mainly adopting the ECHO annotation system for sign languages (cf. Nonhebel et al., 2004), of which I will give some examples in the following section. Researchers, of course, may add coding to their individual needs and focus on specific tiers or aspects. However, some even basic adaptations to the ECHO conventions are considered to be necessary, as the given definitions are less than sufficient and should be clarified.

### 3. ELAN and the ECHO system

ELAN is perfectly suitable for theoretically independent transcription and annotation of multi-media and multi-channel based data, especially sign languages. Up to four videos can be time aligned and played simultaneously. The data can be clicked through frame by frame and a self defined number of tiers can be organized to guarantee precise annotation. The ECHO group of the 'Case Study 4: sign languages' has collected and defined a set of abbreviations and conventions to annotate video data of different sign languages. They agreed on approximately 16 tiers, plus minus one or two, as it might be necessary to have more than one translation or gloss tier in cases the text, apart from English, should also be displayed in another language. It is proposed that the tiers have a certain hierarchy resulting in parent tiers and child tiers. However, it is not the most important point to precisely adopt the number of tiers or the hierarchy, but to follow the defined designations and their short forms. Abbreviations for descriptive vocabulary within the tiers mostly rely on initials of the respective words like 'b' for (*eye*) *blink*, 'r' for *raised (eyebrows)*, etc. These abbreviations can be fed into an ELAN dictionary that can always be retrieved and used for new files. It is possible to constantly adjust and fine-tune the entries of the dictionary, save the template and use it again.

<sup>2</sup> DGS (Deutsche Gebärdensprache = German Sign Language), ISL (Irish Sign Language) and NGT (Nederlandse Gebarentaal = Sign Language of the Netherlands)

<sup>3</sup> cf. Hellwig (2008) for the latest ELAN manual

<sup>4</sup> [www.mpi.nl/lat](http://www.mpi.nl/lat)

## 4. Problematic cases and possible solutions

In the following sections I will provide examples that show some problematic cases and also annotation trials that were incorrect or misleading. I will present suggestions and show how these cases can be avoided or should be dealt with. First, I argue for a continuous annotation of the signing stream (4.1). In a second paragraph (4.2), I will contemplate a continuous annotation of the eye gaze tier, its combination with the eye aperture layer and how this information can be usefully searched for analysis. A third section (4.3) discusses some approximation towards an at least minimally distinguished annotation of pointing signs. The fourth section (4.4) is dedicated to the most diversely discussed topic of classifiers and how they can be annotated without adopting a specific theoretical framework. In a fifth paragraph (4.5), I will argue for the integration of an additional tier for certain facial expressions that cannot be segmented or described adequately by the available tiers. The last section (4.6) adds some final remarks on abbreviations that lack distinctness.

### 4.1 Time span

Assuming Sandler's (2006) Hand Tier model, signs consist of an onset or starting point (L), movement (M) and an endpoint location (L). A preparation phase precedes the sign and a relaxation phase follows it. As the syllable structure, however, is not always LML, it is often hard to define the start and endpoint of a sign. Where exactly does a movement end in case of an LM syllable? So, how are the on- and offsets of signs determined? Shall we annotate the separate signs or a signing stream integrating the transition periods?

Signing consists of a cohesive articulation stream with a certain prosodic structure. Even though the on- and offsets of signs can be defined more precisely than for words, the sign syllable not always has clear boundaries. Therefore, I argue that signing should be annotated as a continuous process that is interrupted when there is a hold or a significant pause. The transition from one sign to the other is often clearly visible through hand shape change, which seems to be the more adequate marker for the annotation domain. Figure 1 shows the continuous annotation of the glosses in the hand or gloss tier.

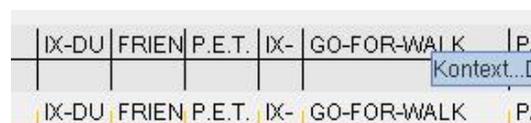


Figure 1: time span annotation ELAN

The only problem left is the fact that sign duration will not be precisely analyzable. However, this issue cannot entirely be solved by the vague separate sign annotation either, as sign boundaries are difficult to grasp. With regard to the rhythmic structure, holds, for example, are marked by (-h) and, of course, pauses or clear interruptions of the signing stream have to be indicated by a gap in the annotation line. The rest of the utterance, however, should be annotated continuously.

## 4.2 Accurate eye gaze aligned with eye blinks

Similar to the section above, I will also discuss the advantages of an accurate annotation of the tiers that are concerned with eye gaze and eye aperture. It seems only logical that the eye gaze tier should not exhibit any breaks except for eye blinks or closed eyes. The signer definitely has to look somewhere, whether it is linguistically significant or not. In addition, it is important to note that while a person closes the eyes or blinks, the eye gaze annotation should be interrupted, as it is physically impossible to blink and simultaneously look. Compare the following annotation examples, where the first tier shows eye aperture and the second tier below marks eye gaze.

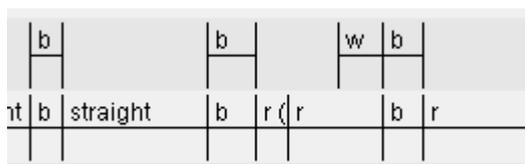


Figure 2: accurate eye gaze annotation

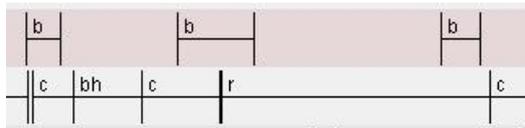


Figure 3: inconsistent eye gaze annotation

The ‘Signs if Ireland’ corpus project, conducted by the Centre for Deaf Studies in Dublin<sup>5</sup>, has annotated these tiers in a similar way, using ‘//’ for blinks and slightly different eye gaze abbreviations. Copying the blink domains would have been more accurate and also less difficult, but the method is basically the same.

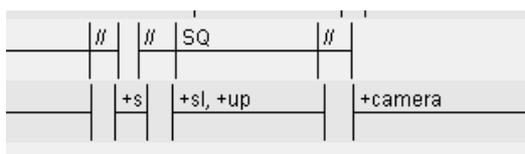


Figure 4: ISL annotation of eye gaze tier

The roles of eye gaze and eye blinks in sign language have not been studied extensively, but a few studies have focused on possible functions and occurrences of certain constructions.<sup>6</sup> If a lot of data is annotated like suggested, reliable assumptions can be made concerning incidences or spreading domains of eye gaze (e.g. their function for agreement or role shift).

In addition, eye blinks should be included in the eye gaze tier, although they are also supposed to be annotated in the eye aperture tier. They can easily be copied to the gaze tier which then also avoids a gaze annotation that co-occurs

<sup>5</sup> [www.tcd.ie/slscs/cds/research/featuredresearch\\_signcorpus.php](http://www.tcd.ie/slscs/cds/research/featuredresearch_signcorpus.php) and also cf. Leeson & Nolan this workshop

<sup>6</sup>See Thompson et al. (2006) for studies of eye gaze in relation to verb agreement or indexicals and Wilbur (1994) as well as Nespor and Sandler (1999) for eye blinks and prosodic issues.

with a blink in the eye aperture tier (see Figures 2 and 3 above). The continuous annotation of the eye gaze tier including blinks is also useful to exactly determine whether an eye gaze change occurs with or without an eye blink and the other way round. The duration and timing of blinks may also be important and should be accurate. Of course nobody can be forced to annotate every small detail. However, if it is decided to incorporate those tiers in the annotation, I argue for the above described way, though being time consuming, the precise annotation of both tiers can be especially relevant for prosodic analysis (cf. Wilbur, 1994, 1999; Nespor and Sandler, 1999) and all the according interfaces that exist.

## 4.3 Pointing signs

The question underlying this section is: How should pointing (signs) be transcribed? As the debate about the status of indexical signs is not clearly sorted out yet, we cannot adopt an annotation that distinguishes pronouns, articles, demonstratives or locatives etc. as it would favor a certain analysis and theory. For any kind of pointing, ECHO suggests the coding IND for index or indexical, and even though I use the widely accepted abbreviation IX, there is no further difference with regard to the underlying definition. However, for the standardized annotation I would like to offer a more detailed distinction of those pointing usages without taking a theoretical framework. No matter if researchers analyze indexicals as a grammatical system or as gestural pointing (Liddell, 2000, 2003), whether they argue for a three part pronominal system (Berenz, 2002; Alibašić Ciciliani & Wilbur, 2006), a first and non-first person distinction (Meier, 1990; Engberg-Pedersen, 1993) or a spatial deictic referents system (McBurney, 2002, 2005), it is still possible to specify the description in some more detail. At least the following distinctions ought to be made:

IX-1	for the index finger pointing to the signer’s chest
IX	for any other pointing by the index-finger
IX-dual (incl.)	pointing by the use of two extended fingers, if the signer is included
IX-dual (excl.)	pointing by the use of two extended fingers, if the signer is excluded
IX-(thumb)	pointing performed by extended thumb

Table 1: index/pointing (IX)

This differentiation would facilitate scouring the corpus for specific indexicals. If researchers are interested in any indexical, they can search for IX, but if they wish to look at index finger based pointing only, they can leave out the thumb examples. They can decide whether dual pointing may be relevant and so they do not have to go through every listed IX-example.

It is up to the annotator whether to add more information that can be attached to IX. Personally I prefer to indicate clear cases of locative pointing by the letter –a and use –pl for ‘plural’ pointing, marking a certain movement of the index-finger rather than pointing to just one location. However, this cannot be demanded of a general annotation convention, even though it does not make a difference with regard to the use of the search tool.

#### 4.4 Classifier signs are poly-componential

Sign languages can depict motion, location and information about the shape of objects and referents within the signing space and exhibit constructions that simultaneously represent nominal features within the verb. This has led Supalla (1986) to compare the constructions to classification systems found in many spoken languages. The handshapes represent the units that are analyzed as classifiers. However, with respect to signed languages, the notion ‘classifier construction’ has been challenged by authors, who claim that the link to spoken language classifier systems is weaker than expected and they suggest different terms and analysis (cf. Schembri, 2003, 2005; Liddell, 2003; Engberg-Pedersen, 1993; Edmondson, 2000). Classifiers are rather called complex predicates, poly-morphemic verbs, reference marker etc., and their status is being debated. Aronoff et al. (2003) and also Sandler & Lillo-Martin (2006), however, still accept the category ‘classifier construction’ in the sense of a definition given by Senft (2000) that the components should be morphemes that classify nouns according to semantic criteria. They argue that the differences and peculiarities of those constructions in sign languages are not enough to ask for a new terminology. Spoken language classifier systems, they say, are not always very similar to each other, too. Many researchers still use the traditional term and work on a precise distinction of various classifier categories.<sup>7</sup> This debate shows that an annotation of the so called ‘classifiers’ is a delicate issue.<sup>8</sup> As the annotation of signed video material should be most detailed and at the same time as much a-theoretical as possible, annotators cannot use specific notions like Handle-, Class/Entity-, or SASS-Classifier etc. However, it is clear that the constructions under discussion have to be marked as such, be it (cl-), traditionally for classifiers in general (as the BSL group of the ECHO data set has chosen), or be it (p-) for poly-componential (like in the NGT data)<sup>9</sup>. This, I do not intend to dictate. However, in the following I will adopt the (cl-) abbreviation just to decide for one option throughout the paper.

First of all it has to be clarified whether these constructions should be transcribed as a modified verb construction or by a paraphrase. I find it much more attractive to have a sign that is glossed in small capitals and then give the additional information that the construction reveals. Compare the following DGS examples where the additional information (*info*) is not yet specified.

a) EMMA LENA FLOWER GIVE-cl: <i>info</i>
b) EMMA LENA FLOWER (cl-) <i>give-info</i>

Table 2: annotation for cl-constructions

The a) example marks the action as the basic part of the construction and then adds the meaning of the modifications. Of course, in b) the verb appears as well,

<sup>7</sup> See Benedicto and Brentari (2003) and (2004) for an overview of different classifier analysis and their own approach.

<sup>8</sup> See Morgan & Woll (2007) for perspectives on classifiers with regard to acquisition, use in discourse, and impairment studies.

<sup>9</sup> cf. the NGT and BSL data (Crasborn et al., 2004 and Woll et al., 2004) from the ECHO project for sign languages

but in many cases the paraphrasing method leads to a far too detailed and often superfluous description of what is performed by the signer. The important thing is that the expressions and words following the categorization do not contain information that cannot be derived by examining the construction in isolation. The verb GIVE changes according to the object that is given, but the give-construction alone cannot mean *give-a-flower*. The noun has to be introduced into the discourse, so the construction itself can only mean *give-a-small-thin-object*. Therefore it should not be transcribed GIVE-cl:flower, but rather GIVE-cl:small-thin-object or something like GIVE-cl:flower-shape-object. In cases where a construction represents a certain class of objects or specific entities that are conventionalized, this must, of course, be indicated differently (WALK-cl:person, STAND-cl:tree<sup>10</sup>). The unclear definitions have led annotators to even transcribe a regular verb BLEAT as (*p-*) *bleating-sheep*, while sheep was already introduced. Annotations like (*-p*) *walk* or (*-p*) *stick in hand* do not seem very convincing, as they lack specification and information about what is done with the stick for example.<sup>11</sup> Temporal information like the ing-form should not be included in the sign language hand tier glossing either. These vague examples could be avoided when it is considered to first annotate the verbal root and then attach the additional information that the construction conveys. This is also desirable, because in cases where both hands represent different entities or objects (e.g. *The bird sits on a tree.*), the hands (right: RH, left: LH) can be glossed independently.

RH	SIT-ON-cl:bird
LH	STAND-cl:tree

Table 3: independent RH and LH annotation

This is much more descriptive than (*cl-*) *a-bird-sits-on-a-tree* or similar paraphrases. However, if the ‘verb plus modification’ annotation is not accepted to be convincing or adequate for general conventions, annotators nevertheless have to consider the different highly important points indicated in this section. Repeating a previously introduced noun in the (cl-) paraphrase, using a noun for information about the shape of objects, calling regular verbs (cl-) constructions etc. is not how systematic annotation should look like.

#### 4.5 Additional tier for ‘looks’

While annotating the data that I have elicited, I came across many cases where a certain relevant facial expression could not be described by entries or the sum of entries within the available tiers.

Especially when working in the area of semantics and pragmatics as well as prosodic phenomena, it seems necessary to have a separate tier, where non-manual adverbials, specific facial expressions, looks, and contoured or tense signing can be annotated. How should the non-manual realization of certain attitudes, expressive meaning, information structure etc. be annotated?

<sup>10</sup> STAND could also be glossed as BE-LOCATED

<sup>11</sup> Examples of annotations from the NGT data set: cf. Crasborn et al. (2004)

Sometimes even adverbial information is found in the GLOSS tier, which should only be used for manual signs or gestures. Examples like WALK-PURPOSEFUL are not desirable. Therefore it is useful, at least for studies focusing on non-manuals, to incorporate an additional tier that leaves space for such expressions that are difficult to describe but are nevertheless relevant. In the present study I have not included such an additional tier in the annotations yet, but used the notes tier for these instances so far. However, this is not very satisfying as overlaps occurred and the information discussed above does not belong to the category of notes. Just to give a few suggestions, the tier could be named *other NMFs*, *looks* or *extra facial expressions* for example.

#### 4.6 Some additional remarks

Finally I would like to further indicate something trivial, which I nevertheless find very helpful and worth considering. Even though it is possible to specifically search tier by tier, identical abbreviations for different expressions or annotations should be avoided. In the ECHO conventions ‘s’, for example, stands for (*head*) *shake* in the head tier and for *squint* in the eye aperture tier. This inadequacy can simply be solved by adding an ‘h’ to the abbreviations in the head tier, so it becomes ‘hs’ for *headshake*, ‘hn’ for *head nod* and ‘ht’ for *head tilt*, which seems to be used by many sign language researchers already. Further specifications like ‘ht-f’ for *head tilt forward* or ‘ht-b’ for a *backward head tilt* are optional and do not influence the searching process. On the long run, however, they could easily be included in the conventions as well.

### 5. Outlook

All these problems and cases of vague definitions and inaccurate usage came into view during the process of finding an appropriate annotation for my corpus and made me decide for certain options, for comparable and independent abbreviations, etc. The workshop and the examples in this paper show that even though many people are currently working on the annotation of sign language data, coding is far away from being conventionalized. Even within the ECHO project the groups worked with varying annotation short forms and slightly different opinions on how to annotate certain aspects of signing. However, a uniform annotation system is essential for various above mentioned reasons: for comparative analysis of different sign languages, simplified handling of search tool functions, comprehensive data exchange etc. It can also be helpful for future research with regard to machine translation and avatar usage for example (cf. among others Morrissey & Way, 2005; Stein et al., 2007).

The ECHO conventions show, that it is possible and eligible to agree on basic notions, and the effort currently undertaken to improve and extend those agreements is well justified. Some vague definitions and false usages have been disclosed, but the ECHO system is highly sophisticated and builds the fundament for all discussed examples. The suggestions I presented shall contribute to the ongoing development of adequate conventions. The paper supports a unified approach and promotes solutions that might be seen as still open to discussion.

Wide-ranging collaborations and comparable cross-linguistic data exchange on a basis of such unified annotation conventions may extremely improve linguistic discussions and the analysis of sign language data.

### 6. Acknowledgements

The corpus study is incorporated in a PhD project in Frankfurt am Main (Germany) that is part of a graduate college, which is granted by the DFG (German Research Foundation). I also would like to thank the MPI in Nijmegen for providing the tools and the know-how freely on the internet. Special thanks go to all the informants in Germany, Ireland and the Netherlands.

### 7. References

- Alibašić Ciciliani, Tamara, Wilbur, Ronnie (2006). Pronominal system in Croatian Sign Language. *Sign Language & Linguistics* 9, pp. 95--132.
- Aronoff, Mark, Meir, Irit, Padden, Carol & Wendy Sandler (2003). Classifier Constructions and Morphology in Two Sign Languages. In Karen Emmorey (ed), *Perspectives on Classifier Constructions in Sign Language*. Mahwah, NJ: Erlbaum, pp. 3--34.
- Benedicto, Elena, Brentari, Diane (2003). *Choosing an Analysis for Verbal Classifiers in Sign Language. A comparative Evaluation*, Ms., Purdue University.
- Benedicto, Elena, Brentari, Diane (2004). Where did all the arguments go? Argument-Changing Properties of Classifiers in ASL. *Natural Language and Linguistic Theory* 22.4, pp. 743--810.
- Berenz, Norine (2002). Insights into person deixis. *Sign Language & Linguistics* 5 (2), pp. 203--227.
- Crasborn, Onno, Hanke, Thomas (2004). *Metadata for sign language corpora. Background document for an ECHO workshop. May 8 + 9, 2003*, Radboud University Nijmegen.
- Crasborn, Onno, van der Kooij, Els, Nonhebel, Annika & Wim Emmerik (2004). *ECHO data set for Sign Language of the Netherlands (NGT)*. Department of Linguistics, Radboud University Nijmegen.
- Edmondson, William (2000). *Rethinking Classifiers. The Morpho-phonemics of Proforms*. Ms., University of Birmingham, UK.
- Engberg-Pedersen, Elisabeth (1993). *Space in Danish Sign Language*. Hamburg: Signum.
- Hanke, Thomas (2001). Sign language transcription with syncWRITER. *Sign Language and Linguistics* 4:1/2, pp. 275--283.
- Hellwig, Birgit (2008). *ELAN Linguistic Annotator. Version 3.4. Manual*.
- Herrmann, Annika (2007). Modal Meaning in German Sign Language and Irish Sign Language. In Perniss, Pamela, Pfau, Roland & Markus Steinbach (eds.), *Visible Variation. Comparative Studies on Sign Language Structure*. Berlin: Mouton de Gruyter, pp. 245--278.
- Liddell, Scott (2000). Indicating Verbs and pronouns: pointing away from agreement. In Karen Emmorey & Harlan Lane (eds.), *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima*. Mahwah, NJ: Erlbaum, pp. 303--320.

- Liddell, Scott (2003). Sources of meaning in ASL classifier predicates. In Karen Emmorey (ed) *Perspectives on Classifier Constructions in Sign Language*. Mahwah, NJ: Erlbaum, pp. 199--220.
- McBurney, Susan Lloyd (2002). Pronominal reference in signed and spoken language: are grammatical categories modality-dependent? In Richard P. Meier et al. (eds.), *Modality and Structure in Signed and Spoken Languages*. Cambridge: CUP, pp. 329--369.
- McBurney, Susan Lloyd (2005). Referential morphology in signed languages. Dissertation Abstract. In Anne Baker and Bencie Woll (eds.), *Language Acquisition: Special issue of Sign Language & Linguistics 8:1/2*, pp. 211--215.
- Meier, Richard P. (1990). Person deixis in American Sign Language. In Susan D. Fischer & Patricia Siple (eds.), *Theoretical Issues in Sign Language Research*. Chicago: University of Chicago Press, pp. 175--190.
- Morrissey, Sara, Way, Andy (2005). An Example-based Approach to Translating Sign Language. In *Proceedings of the Workshop in Example-Based Machine Translation (MT X-05)*, Phuket, Thailand, pp. 109--116.
- Morgan, Gary, Woll, Bencie (2007). Understanding sign language classifiers through a polycomponential approach. *Lingua* 117, pp. 1159--1168.
- Neidle, Carol (2001). SignStream. A database tool for research on visual-gestural language. *Sign Language and Linguistics 4:1/2*, pp. 203--214.
- Nespor, Marina, Sandler, Wendy (1999). Prosody in Israeli Sign Language. *Language and Speech* 42, pp. 143--176.
- Nonhebel, Annika, Crasborn, Onno & Els van der Kooij (2004). *Sign language transcription conventions for the ECHO project*. Version 9, 20 January 2004, Radboud University Nijmegen.
- Sandler, Wendy, Lillo-Martin, Diane (2006). *Sign Language and Linguistic Universals*. Cambridge: Cambridge University Press.
- Schembri, Adam (2003). Rethinking 'Classifiers' in Signed Languages. In Karen Emmorey (ed) *Perspectives on Classifier Constructions in Sign Language*. Mahwah, NJ: Erlbaum, pp. 3--34.
- Schembri, Adam, Jones, Caroline, & Denis Burnham (2005). Comparing action gestures and classifier verbs of motion: Evidence from Australian Sign Language, Taiwan Sign Language, and non-signers' gestures without speech. *Journal of Deaf Studies and Deaf Education* 10:3, pp. 272--290.
- Senft, Gunter (2000). *Systems of nominal classification: Language Culture and Cognition*. Cambridge: CUP.
- Stein, Daniel, Dreuw, Philippe, Ney, Hermann, Morrissey, Sara & Andy Way (2007). Hand in Hand: Automatic Sign Language to English Translation. In *Proceedings of Theoretical and Methodological Issues in Machine Translation (TMI-07)* Skovde, Sweden.
- Supalla, Ted (1986). The classifier system in American Sign Language. In C. Craig (ed.), *Noun classes and categorization: Typological studies in language*, Vol. 7 Amsterdam: Benjamins, pp. 181--214.
- Thompson, Robin, Emmorey, Karen, & Robert Kluender (2006). The relationship between eye gaze and agreement in American Sign Language: An eye-tracking study. *Natural Language and Linguistic Theory*, 24, pp. 571--604.
- Wilbur, Ronnie (1999). Stress in ASL: Empirical Evidence and Linguistic Issues. *Language and Speech* 42 (2.3), pp. 229--250.
- Wilbur, Ronnie (1994). Eyeblinks and ASL Phrase Structure. *Sign Language Studies* 84, pp. 221--240.
- Woll, Bencie, Sutton-Spence, Rachel, & Dafydd Waters (2004). *ECHO data set for British Sign Language (BSL)*. Department of Language and Communication Science, City University London.

## Building up Digital Video Resources for Sign Language Interpreter Training

**Jens Heßmann (1), Meike Vaupel (2)**

(1) University of Applied Sciences Magdeburg-Stendal

Fachbereich Sozial- und Gesundheitswesen

Breitscheidstr. 2

D-39114 Magdeburg

(2) University of Applied Sciences Zwickau

Fachbereich Gesundheits- und Pflegewissenschaften

Dr.-Friedrichs-Ring 45

D-08056 Zwickau

E-mail: [jens.hessmann@hs-magdeburg.de](mailto:jens.hessmann@hs-magdeburg.de), [Meike.Vaupel@fh-zwickau.de](mailto:Meike.Vaupel@fh-zwickau.de)

### Abstract

The development and implementation of new digital video facilities for Sign Language Interpreter Training calls for a more pragmatically oriented system of data classification than what is commonly used for linguistic purposes today. A corpus that addresses the needs of an interpreter training program should reflect the full spectrum of sign language and allow for comparative analyses and practical exercises in interpretation and translation. The universities of applied sciences in Magdeburg and Zwickau have installed the same type of digital video facility and are currently working on a classification system for archiving video resources for interpreter training and research. To adapt to the pragmatic aspect our starting point is translation theory, which is interdisciplinary in nature and bears potential to include both linguistic and translation oriented aspects. Since the official acknowledgement of German Sign Language an increasing number of interpreting and recently also translation tasks emerge, and with it an increasing number of varieties in textual representations. Besides research purposes, training institutions need to take this into consideration and adapt their data to a digital format that enables the students and teachers to have easy access to potentially all textual representations that they might encounter in reality.

### 1. New challenges to old practices in sign language interpreter (SLI) training

Sign language interpreter training has been offered at the universities of applied sciences in Magdeburg and Zwickau since 1997 and 2000, respectively. Both training programs are set in the institutional context of East German universities that experienced a major reorganization after the reunification of Germany. The training programs share an applied perspective in research and teaching as well as many of the features typical for small-scale academic ventures in a developing field. Thus, the provision of teaching materials and, more particularly, sign language video resources, adequate in content, format and technical quality, has been a constant concern. For want of better options, a hands-on approach was chosen for the last ten years, and both programs have amassed a heterogeneous collection of analogue and digital video films for teaching and research purposes. In most cases, the only way of accessing this material consists of picking the brains of those colleagues who may have worked with some video clip or exercise suitable for one's own didactic or research purposes.

As it happens, both Magdeburg and Zwickau have installed the same type of digital training facilities (henceforth 'video lab') towards the end of 2007. These video labs consist of individual workstations linked to a central video server that hosts all the resources in a unified digital format. Both institutions now face the major challenge of facilitating a process that will transform and

complement existing sign language materials so as to create an accessible library of video resources for research and training purposes. This presentation will report on our joint effort to undertake the first steps in this direction and focus especially on the criteria for annotating and archiving digital sign language resources.

### 2. Building up Sign Language Corpora: Specific demands of SLI Training

Building up a Sign Language Corpus, fundamental issues need to be raised such as *legal and ethical issues* or *issues regarding the administrative and technical prerequisites*. Up to now, questions of ownership and property rights have often been dealt with somewhat casually. Building up a digital library of video resources implies that such questions have been formally clarified. However, just what the conditions for using video materials gathered informally, passed on from one colleague to the next or published on the internet are, may be hard to decide. In order to create a legal basis for the desired cooperation and be able to access university funds, the two universities concerned will enter into formal agreements about the mutual use of video resources. This, in turn, demands that there are clearly defined ways of synchronizing, adding to and accessing the respective collections of resources. These fundamental topics are currently under scrutiny in both institutions. For the purpose of this workshop a third topic will be of specific interest, namely the *criteria for*

*annotating and archiving video resources.* While the process of digitizing and storing existing video materials can be dealt with somewhat mechanically, the development of systematic ways of annotating and organising sign language materials is crucial in order to make digital resources accessible. Clearly, this is an area where progress has been made in recent years, e.g. in the context of the ECHO project ('European Cultural Heritage Online')<sup>1</sup>. We will add to this discussion by considering the more specific demands of sign language interpreter training and research.

## 2.1 Demands on SLI

Sign language interpreting today is mostly performed as community interpreting which aims to provide or facilitate full access to intra-social public services in e.g. the legal, health care, educational, governmental, academic, religious, or social field. Interpreters must therefore be familiar with the form and content of a great variety of texts in their respective working languages. The working languages in our case are to date German as vocal language in written and spoken mode and German Sign Language. Interpreting can be either unilateral or bilateral and in both modalities multiple textual representations may occur. Until today, SLIs rarely specialize in just one field but are expected to be able to translate whatever written, spoken or signed text may occur in any given situation. It is due to the long history of oppression of sign languages that interpreters today are faced with a paradox. While a common definition of their interpreting task asks SLI to produce a target text that is presumed to have a similar meaning and/or effect as the source text (Pöchhacker, 2007), many spoken or written texts of vocal languages in the context of community interpreting have no such counterpart in sign language, for there has never been access to these areas. Following the definition of community interpreting, the sole access to these areas is often through interpreting, resulting in a target text that is based on little or no valid ground regarding its content and form. With increasing access of deaf professionals to the varying fields of community life a growing number of different sign language texts (one-time presentations and recorded) occur. Sign Language Interpreters and Translators are confronted with a very dynamic, fast-growing and changing language in use. In the case of an existing parallel text we face the problem that until today very few research has been done on register variation in sign language discourse (Hansen, 2007). We may be able to detect the overall function of the utterance but a classification of text functions and corresponding language registers must be considered as preliminary if there is one at all. We also must be aware that oral languages have less register variation than those with a long history of written codes (Biber, 1995). This leads to the notion of having skilled interpreters who not only possess exceptional textual skills but also know how to evaluate their skills and broaden their knowledge

autodidactically.

## 2.2 Demands on SLI Training

Acquisition and evaluation of textual skills are thus cornerstones of the SLI training. Training facilities should be able to provide their students with a great variety of different texts in both languages. While the students are exposed to an infinite number of vocal language texts in both the spoken and written mode in daily life, their access to sign language texts is limited in comparison. Some communicative events might not even be accessible for students at all, such as e.g. therapy sessions with a deaf therapist. Others might simply not be reachable, because they take place too far away. Magdeburg and Zwickau are both located in areas with a fairly small deaf community, which further limits exposure to sign language. Digital technology thus plays a crucial role in our training programs. It can and should never compensate for live encounters with the sign language community but can definitely add to it. It is vital to cover as many topics, constellations and situations as possible to prepare the students as thoroughly as possible for their ensuing professional life. With the video lab the material can be used for language/text and translation technique acquisition in class as well as for autodidactic purposes. Furthermore, it provides an option to compare and evaluate parallel texts in both languages as well as source and target text productions in regard to their adequacy in the respective interpretation or translation.

## 2.3 Demands on SLI Training Corpus

A corpus that addresses the needs of an interpreter training program should reflect the full spectrum of sign language in use and allow for comparative analyses and practical exercises in interpretation and translation.

Following the purposes mentioned above one can extract four major demands that reach beyond the needs of common linguistic corpora, namely:

- Extension and differentiation of sign language corpora to reflect the full spectrum of sign language use
- Creation of parallel corpora of spoken language texts to allow for comparative analysis and practical exercises
- Development of a system of classification that allows for following up systematic cross references not only within but between signed and spoken/written texts
- Collection of existing source-target text pairs, i.e. interpretation/translation of sign language and vocal language texts that may serve for analytical purposes as models, objects of critical reflection, etc.

It may seem odd to include vocal language texts in a sign language corpus but considering its purpose it seems mandatory to also work with parallel texts for comparative purposes. A carefully defined selection of spoken language texts in both oral and written forms that can be extracted from real interpreting/translation situations, can serve as models for comparison.

The corpus should be organized in a way that enables the SLI trainer to search for material according to the respective focus of the training, such as setting-oriented

<sup>1</sup> (cf. <http://www.let.ru.nl/sign-lang/echo/index.html>)

training (e.g. only health care texts), discourse type oriented trainings (e.g. only speeches), function oriented trainings (e.g. only instructive texts), phenomenon oriented trainings (e.g. constructed action), or for evaluation purposes (e.g. analyzing simultaneous interpretation). This calls for a modified approach for the classification of digital text material.

### 3. Digital Video Corpora as training resources: Towards a system of signed/spoken text classification

Over the years Magdeburg and Zwickau both have collected a great number of recorded sign language data that is used but not systematically archived for teaching. Most of the material was taped for teaching sign language or conducting sign language research: the number of explicit interpreting or translation material is comparably small. Archiving activities are limited to databases, which give only a very rough overview i.e. on topic (oftentimes not necessarily well suited), recording date if known, name of signer if known, length, and quality of the recording. These attempts neither fit the requirements for SLI training nor the requirements of the new video lab. What is required is a system of text classifications. In search of a theoretical underpinning of our attempt to systematize our material we found Pöchhacker's "Domains and Dimensions of interpreting theory" (2007) a useful model for a first careful approach. Since not enough research on sign language texts has been conducted, this model allows to translate an essentially text linguistic approach to the context of interpreting studies. According to Pöchhacker, interpreting studies differentiate between eight domains. Each can be characterized by a number of dimensions that form the interpreting event, which can be summarized in the following domain-dimension interplays:

1. *Medium* as either human or machine translation. Although there are just a few attempts to automate translation in the field of sign languages, this domain might gain a greater impact in future development.
2. *Setting* as differentiating between inter- and intra-social events, such as international conferences on the one hand and community interpreting in i.e. health care, court, education, etc. on the other.
3. *Mode* defining translation as simultaneous, short consecutive (without notes) and 'classical' consecutive (with notes), also giving information about the form of translation as interpreting or (sight) translation.
4. *Languages* considering the status and modality as in vocal vs. sign languages and conference language vs. migrant (minority) languages.
5. *Discourse* giving information about the type of text like speeches, debates or face-to-face talk.
6. *Participants* differentiating the status as equal representatives vs. individual with institutional representative, taking power constellations into

consideration.

7. *Interpreter* described as professionally trained, semi-professional (not certified or trained but working up to the same standards as professionals) or 'natural' bilingual individuals without training in special translation skills.
8. Accompanying *problems* such as simultaneity, memory, quality, stress, effect and role.

While the "interplay of the first seven dimensions serves to highlight some of the key factors in the various prototypical domains", the last dimension represents "a set of major research concerns to date" (Pöchhacker, 2007). According to this model an international conference prototypically is an interpreted event that is characterized by making use of a professional human interpreter in simultaneous working mode in a booth, most likely between typical spoken conference languages with equal representatives holding speeches. In contrast the typical interplay of intra-social dimensions, e.g. translating a doctor's appointment, would be characterized also by a human translator in the consecutive or simultaneous working mode, personally present in the situation who is oftentimes a semi-professional or 'natural' bilingual individual, interpreting between the official language of the country and a migrant/minority language for an individual that seeks help from a representative of a health care facility. Although patterns can be detected, the number of actual texts that are uttered in the respective situations is countless. Considering the underlying general goal of SLI training as stated in 2.2, purpose oriented metadata can be organized according to the domains/dimensions mentioned above, leading to a set of metadata different from those used in linguistic research today. It should enable the SLI trainer to search and pick material pragmatically, depending on the main focus of training. Bearing in mind that metadata should "allow the user to discover relevant material with a high precision and recall" (Wittenburg & Broeder, 2003), a more translation-oriented approach seems to be justified. Descriptions of the material should come up as "descriptions at a general level of the nature of the data that can be considered constant for a whole recording" (Hanke & Crasborn, 2003). From our present viewpoint - keeping in mind that we are at the very beginning - we consider the Pöchhacker model to meet these requirements in addition to general and technical information about the recording itself. Combining metadata as in use today with the Pöchhacker categories, we will have to extend these i.e. by adding information about the actors to the domain of *participants etc.* Work on this is still in progress and hopefully the discussion about our attempt will add to creating these categories. While most of the categories are specific to translation, the domain of *discourse* is the one where translation studies and linguistics obviously meet. As mentioned above, no sufficient research has been conducted that enables us to categorize sign language texts as we can for spoken language texts. Although even in spoken

languages there is a diversity of approaches of text classification (Adamzik, 2004), there are at least common labels that are used. The distinction between text external and internal factors described in Stede (2007), and for translations purposes by Nord (1995) reflects the problem. While external factors such as the function of a text, the situation, the degree of publicity can be notated in metadata, the internal features such as the structure of the text, syntactic patterns, typical lexical items in regard to the function must be part of a linguistic annotation. It seems practicable to not focus on text types as general categories of texts (e.g. speech, business letter) but to use this term according to Werlich (1975) and define function-oriented patterns of textual representations in regard to the contextual focus. Werlich defines five such patterns and labels them as *descriptive*, *narrative*, *expository*, *argumentative* and *instructive*. According to Biber, adapting the same labels and using them for a different language, bares the danger of denying or ignoring phenomena that are specific to this particular language (1995). This must be taken into consideration when dealing with labels developed for vocal languages and possibly apply them to signed languages. Furthermore the aspect of literacy/orality should be taken into consideration when constructing parallel texts, as “the context of primary orality means that the meaning of the exchange will be strikingly different from a similar exchange in the context of literacy” (Cronin, 2002). The potential in our approach might be to not only to be able to categorize and label but possibly also to gain insight into new patterns and forms of sign language communication. Metadata concerning external text factors in combination with linguistically annotated internal text factors will hopefully enable us in the long run to conduct combined searches such as looking for instances of constructed action (annotated data) in instructive texts (metadata) in educational settings (metadata).

#### 4. Next steps

Since both Magdeburg and Zwickau are under pressure to start storing their data in an organized compatible way, the first step (besides legal, ethical and administrative considerations) must be the implementation of a framework for metadata where future linguistic findings on sign language texts find their place and can easily be added. As pointed out, addressing the problem from the perspective of translation theory seems to be a useful approach, since the nature of translation study is interdisciplinary. We believe that there is potential for future research from a cross linguistic perspective: having stored context information about the communication event in which a text occurred or was translated and knowing more about register variation, parallel corpora can be drawn upon in SLI training. We are fully aware that we are talking about decades here, but we believe that in the long run it could lead to enhancements in translation as it enables deaf and hearing to perform a more

theoretically informed translation of spoken and/or written, respectably signed texts. Especially the growing market for sign language translations (e.g. translations of websites that are permanently accessible as movies on the site or sign language websites with subtitles and/or voice over) supports our attempt to systematize from a translation theory perspective.

#### 5. References

- Adamzik, K. (2004). *Textlinguistik. Eine Einführende Darstellung*. Tübingen: Niemeyer
- Biber, D. (1995). *Dimensions of register variations : a cross linguistic comparison*. Cambridge : Cambridge University Press
- Cronin, M. (2002). *The Empire talks back: Orality, Heteronymy and the Cultural Turn in Interpreting Studies*. In: Pöchhacker, F.; Schlesinger, M. (ed.). *The Interpreting Studies Reader*. London/New York: Routledge, S. 386-397.
- Hanke, T. ; Crasborn, O. (2003). *Metadata for sign language corpora*. Version : 15 Sep2008. <http://www.let.kun.nl/sign-lang/echo/events.html>
- Hansen, M. (2007). *Warum braucht die Deutsche Gebärdensprache kein Passiv? Verfahren der Markierung semantischer Rollen in der DGS*. Arbeiten zur Sprachanalyse, Bd. 48. Frankfurt a.M.: Peter Lang
- Nord, C. (1995). *Textanalyse und Übersetzen*. Heidelberg : Groos Edition Julius
- Pöchhacker, F. (2006). *Introducing Interpreting Studies*. New York: Routledge
- Stede, M.(2007). *Korpusgestützte Textanalyse*. Tübingen : Narr
- Werlich, E. (1975). *Typologie der Texte. Entwurf eines textlinguistischen Modells zur Grundlegung einer Textgrammatik*. Heidelberg: Quelle&Meyer
- Wittenburg, P. ; Broeder, D. (2003). *Metadata in ECHO*. Version : 10 Mar 2003. <http://www.mpi.nl/echo/tec-rep/wp2-tr08-2003v1.pdf>

# Semi-automatic Annotation of Sign Language Corpora

Marek Hruží, Pavel Campr, Miloš Železný

Department of Cybernetics, Faculty of Applied Sciences, University of West Bohemia  
Univerzitní 22, Pilsen, Czech Rep.  
mhruz@kky.zcu.cz, campr@kky.zcu.cz, zelezny@kky.zcu.cz

## Abstract

Automatic Sign Language Recognition is a problem that is being solved by many research institutes in the world. Up to now there is a deficiency of corpora with good properties such as high resolution and frame rate, several views of the scene, detailed annotation etc. In this paper we take a closer look at the annotation of available data.

## 1. Introduction

The first step of automatic sign language recognition is feature extraction. It has been shown which features are sufficient for a successful classification of a sign (Ong and Ranganath, 2005). It is the hand shape, orientation of the hand in space, trajectory of the hands and the non-manual component of the speech (facial expression, articulation). Usually the efficiency of the feature extracting algorithm is evaluated by the rate of recognition of the whole system. This approach can be confusing since the researcher cannot be always sure which part of the system is failing. However if the corpora were available with a detailed annotation of these features the evaluation could be more precise. A manual creation of the annotation data can be very time consuming. We propose a semi-automatic tool for annotating trajectory of head and hands and the shape of the hands.

## 2. Goal of the paper

The goal of this paper is to introduce a system for semi-automatic annotation of sign language corpora. There is some annotation software available (for example ELAN) but the possibilities of these programs are limited. Usually we are able to select a region in a video stream where a sign is performed and note some information about this sign. This process is inevitable for sign recognition and sign language understanding. However if we want to evaluate the feature extracting algorithm we need a lower level annotation of the features themselves in every frame. This annotation has several benefits. We can use the features from the annotation to build and test a recognition system. We can use the features to train models of movement and hand shape. And finally we can compare a set of automatically detected features with the features from annotation.

## 3. Annotation of features

There are many ways to describe the features needed for an automatic sign language recognition. We chose the following description:

- trajectory - a set of 2D points representing the mean of the contour of an object (or center of mass) for every frame
- hand shape and orientation - we use seven Hu moments (Hu, 1962)

- non manual component - a gray-scale image of the face

From this set of features we derived that the needed annotation of the image data is a countour of the hands and the head. Detecting the contour can be very time expensive for a human but there are many methods for extracting the contour automatically. Next step is to decide which object is represented by the contour. It is a very easy task for a human but again can be time consuming. That is why we developed a tracker for this purpose.

### 3.1. Tracking process

The tracker is based on a similarity of the scalar description of the objects. We describe the objects by:

- seven Hu moments of the contour
- a gray scale image (template)
- position
- velocity
- perimeter of the contour
- area of the bounding box
- area of the contour.

For every new frame all objects in the image are detected and filtered. Every tracker instance computes the similarity of the tracked object and the evaluated object.

$$S_{Hu} = \sum_{i=1}^7 \frac{1}{m_i^A} - \frac{1}{m_i^B} \quad (1)$$

where A denotes the first shape (tracked object in the last frame), B denotes the second shape (object in actual frame),

$$m_i^A = \text{sign}(h_i^A) \cdot \log(h_i^A) \quad (2)$$

$$m_i^B = \text{sign}(h_i^B) \cdot \log(h_i^B) \quad (3)$$

where  $h_i^A$  is the i-th Hu moment of the shape A and analogical for  $h_i^B$ .  $S_{Hu}$  then denotes the shape (contour) similarity. Next we present the similarity of the template. For

this purpose we have to compute the correlation between the template of the tracked object and the evaluated object.

$$R(x, y) = \frac{\sum_{x'} \sum_{y'} T'(x', y') \cdot I'(x + x', y + y')}{T' \otimes I'} \quad (4)$$

where

$$T' \otimes I' = \sqrt{\sum_{x'} \sum_{y'} T'(x', y')^2 \sum_{x'} \sum_{y'} I'(x + x', y + y')^2} \quad (5)$$

where

$$T'(x', y') = T(x', y') - \frac{1}{(w \cdot h) \cdot \sum_{x''} \sum_{y''} T(x'', y'')} \quad (6)$$

$$I' = I - \frac{1}{(w \cdot h) \cdot \sum_{x''} \sum_{y''} I(x + x'', y + y'')} \quad (7)$$

where  $I$  is the image we search in,  $T$  is the template that we search for,  $w$  and  $h$  are the width and height of the template respectively. Then

$$S_T = \max_{x, y} R(x, y) \quad (8)$$

is the template similarity. The other similarity functions are an absolute difference between the values in last frame and in the present frame.

$$S_P = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2} \quad (9)$$

is the similarity of position, where  $[x, y]^T$  is the center of the mass of the object.

$$S_V = \sqrt{(v_x^t - v_x^{t-1})^2 + (v_y^t - v_y^{t-1})^2} \quad (10)$$

is the similarity of velocity, where  $[v_x, v_y]^T$  is the velocity of the object. The velocity can be approximated as

$$\vec{v} = \begin{bmatrix} x_t - x_{t-1} \\ y_t - y_{t-1} \end{bmatrix} \quad (11)$$

thus the equation 10 becomes

$$S_V = \sqrt{(x_t - 2x_{t-1} + x_{t-2})^2 + (y_t - 2y_{t-1} + y_{t-2})^2} \quad (12)$$

$$S_{PC} = |p_t - p_{t-1}| \quad (13)$$

is the similarity of the perimeter of the object, where  $p$  is the perimeter of the object.

$$S_{ABB} = |abb_t - abb_{t-1}| \quad (14)$$

is the similarity of the area of the bounding box, where  $abb$  is the area of the bounding box of the object. A bounding box is a non-rotated rectangle that fits the whole object and has minimum area.

$$S_{AC} = |ac_t - ac_{t-1}| \quad (15)$$

is the similarity of the area of the object, where  $ac$  is the area of the object. Based on the values of the similarity functions the tracker has to determine the likelihood (or certainty) with which the object is the tracked object. The likelihood function can be built in many ways. We use a trained Gaussian Mixture Model (GMM) to determine the likelihood. Every similarity function responds to one dimension. There are seven similarity functions which means a 7D feature space and a 7D GMM. The training samples are collected during annotation with and untrained tracker. The untrained tracker doesn't give good results and that's why the user has to manually annotate almost every frame. This situation can be overcome by manually setting the tracker parameters. In this case the overall similarity function can be a linear combination of the partial similarity functions. That is

$$S = \vec{w}^T \begin{bmatrix} S_{Hu} \\ S_T \\ S_P \\ S_V \\ S_{PC} \\ S_{ABB} \\ S_{AC} \end{bmatrix} \quad (16)$$

where  $\vec{w}$  is the weighting vector. An expert can set the weights for better tracking performance. The weights can be then iteratively recomputed based on the data from annotation using a least squares method. After few iterations the data can be used to train the GMM.

As long as the tracker's certainty is above some threshold, the detected features are considered as ground truth. At this point all available data are collected from the object and saved as annotation. If the level of uncertainty is high, the user is asked to verify the tracking.

If a perfect tracker was available all the annotation could be created automatically. But the trackers usually fail when an occlusion of objects occurs. Because of this problem the system must be able to detect occlusions of objects and have the user verify the resulting tracking. In our system we assume that the bounding box of an overlapped object becomes relatively bigger in the first frame of occlusion and relatively smaller in the first frame after occlusion. We consider the area of the bounding box as a feature which determines the occlusion. In Figure 1 you can see the progress of the area of the bounding box of the right hand through the video stream of a sign *Brno*. Figure 2 is the difference of the area computed as

$$\Delta a = a_t - a_{t-1} \quad (17)$$

where  $a_t$  is the area of the bounding box in time (frame)  $t$ . Figure 3 shows the relative difference and thresholds.

$$\Delta a = \frac{a_t - a_{t-1}}{a_{t-1}} \quad (18)$$

The upper threshold set to 0.8 is used for the detection of first occlusion. The lower threshold set to -0.4 is used for

the detection of the first frame after occlusion. The experiments were done on database UWB-06-SLR-A (Campr et al., 2007).

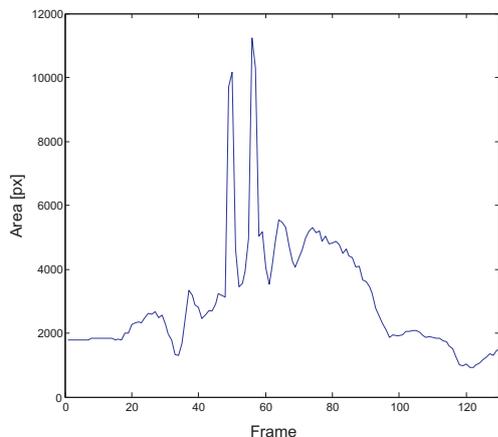


Figure 1: Area of the bounding box of the right hand in pixels.

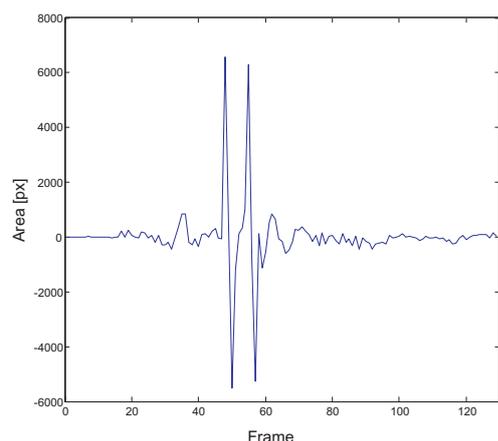


Figure 2: Difference of area of the bounding box of the right hand in pixels.

### 3.2. Annotation process

The annotation itself begins with loading the video file. In the first frame the trackers are initialized. There is one tracker for one object. In the case of sign language the objects are head, left and right hand. So there are three trackers in this scenario. The initialization process is as follows. The image is segmented using a skin color model. All the small objects and the very large objects are filtered out. Every tracker is created with a search window. If an object is found in this window, the tracker is initialized by this object. The result of the initialization is presented to a human. The human has to decide whether the trackers are initialized correctly and if not, he has to initialize them manually. The trackers are identified by a green contour of the tracked object, a blue bounding box of the object and a string with the class of the object (left hand, right hand, head). After

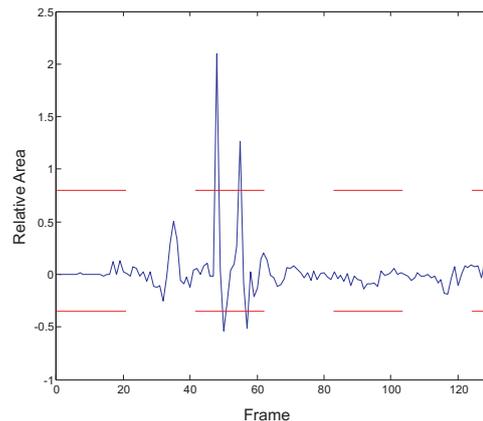


Figure 3: Relative difference of area of the bounding box of the right hand in pixels. The dashed lines are upper and lower thresholds for occlusion detection.



Figure 4: Selected frames (48, 49, 50) from the video stream of a sign Brno. Notice that in the frame 48 the relative difference of area is over the upper threshold and in frame 50 is below the lower threshold.

the initialization the above mentioned tracking process begins. The human operator can pause the video stream in any frame and with a key stroke he is able to view the stream frame-by-frame. If the area of the bounding box changes rapidly, the system pauses the stream automatically. Usually this is a sign where two or more objects collided with each other or were separated from each other. This state can create a confusion for the tracker and the user has to verify the correctness of the automatic annotation. If the annotation doesn't seem right, the user can modify it. In this case all the detected objects are presented to the user and he can annotate (assign a tracker to the object) the object. This way the user doesn't need to annotate every frame which means he saves a lot of time.

### 3.3. Verification process

After the annotation is done the user can verify it. The system loads the saved features of the video stream and presents them to the user. In every frame the system draws the detected contours and bounding boxes along with the string identifier into the image from the video stream. This way the user is able to tell whether the annotation was successful or not. Some additional information can be seen in the verification mode. It is a line connecting the center of mass of the object in the last frame and in the present frame. The length of the line is also written on the screen.

This may be helpful when an expert is setting the tracker parameters. Again, the user can pause the stream any time and view the video frame-by-frame.

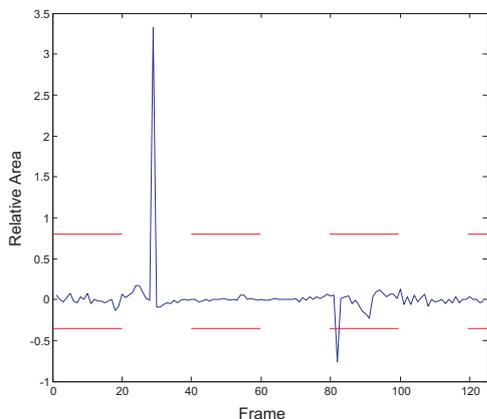


Figure 5: Relative difference of area of the bounding box of the right hand in pixels of the sign divka (girl).



Figure 6: Selected frames from the video stream of a sign divka. You can observe a frame just before occlusion (29), the first frame of occlusion (30) and the consequent frame (31).

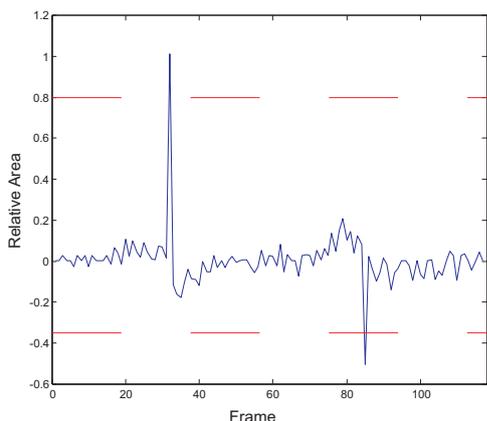


Figure 7: Relative difference of area of the bounding box of the right hand in pixels of the sign loucit se (farewell).

#### 4. Conclusion

We present a system for semi-automatic annotation of Sign Language Corpora. The system can help experts to an-



Figure 8: Selected frames from the video stream of a sign loucit se. You can observe the last frame of occlusion (83), the first frame after occlusion (84) and the consequent frame (85).

notate the sign language video streams without any major time consumption. The annotation is useful for feature extraction, as the features can be computed from the annotation data. This way a system of recognition can be developed independently from the feature extracting system. New algorithms for feature extraction can be compared with the baseline system, not only in the domain of recognition but also in the correctness of the extracted features. Up to now the annotation through tracker allows us to semi-automatically obtain the trajectory of head and hands and the shape of the hands. In the future we will extend the system to be able to determine the orientation of hands and combine it with a lip-reading system which we have available (Císař et al., 2007). The verification mode is a fast way to verify your annotation and it helps experts to set the tracker parameters manually.

#### 5. Acknowledgement

This research was supported by the Grant Agency of Academy of Sciences of the Czech Republic, project No. 1ET101470416 and by the Ministry of Education of the Czech Republic, project No. ME08106.

#### 6. References

- P. Campr, M. Hruží, and M. Železný. 2007. Design and recording of signed czech language corpus for automatic sign language recognition. *Proceedings of Interspeech 2007*, pages 678–681.
- P. Císař, M. Železný, J. Zelinka, and J. Trojanová. 2007. Development and testing of new combined visual speech parameterization. *Proceedings of the workshop on Audio-visual speech processing.*, pages 97–100.
- M. K. Hu. 1962. Visual pattern recognition by moment invariants. *IRE Trans. Information Theory*, 8:179–187.
- S.C.W. Ong and S. Ranganath. 2005. Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27:873–891.

# Corpus linguistics and signed languages: no lemmata, no corpus

**Trevor Johnston**

Department of Linguistics, Macquarie University  
Sydney, New South Wales, Australia  
E-mail: trevor.johnston@mq.edu.au

## Abstract

A fundamental problem in the creation of signed language corpora is lemmatisation. Lemmatisation—the classification or identification of related word forms under a single label or lemma (the equivalent of headwords or headsigns in a dictionary)—is central to the process of corpus creation. The reason is that signed language corpora—as with all modern linguistic corpora—need to be machine-readable and this means that sign annotations should not only be informed by linguistic theory but also that tags appended to these annotations should be used consistently and systematically. In addition, a corpus must also be well documented (i.e., with accurate and relevant metadata) and representative of the language community (i.e., of relevant registers and sociolinguistic). All this requires dedicated technology (e.g., ELAN), standards and protocols (e.g., IMDI metadata descriptors), and transparent and agreed grammatical tags (e.g., grammatical class labels). However, it also requires the identification of lemmata and this presupposes the unique identification of sign forms. In other words, a successful corpus project presupposes the availability of a reference dictionary or lexical database to facilitate lemma identification and consistency in lemmatisation. Without lemmatisation a collection of recordings with various related appended annotation files will not be able to be used as a true linguistic corpus as the counting, sorting, tagging, etc. of types and tokens is rendered virtually impossible. This presentation draws on the Australian experience of corpus creation to show how a dictionary in the form of a computerized lexical database needs to be created and integrated into any signed language corpus project. Plans for the creation of new signed language corpora will be seriously flawed if they do not take this into account.

## 1. Introduction

After a brief discussion of the nature and role of corpora in contemporary empirical linguistics, I describe the Auslan (Australian Sign Language) Corpus and the Auslan Lexical Database. I discuss what makes this a genuine linguistic corpus in the modern sense: lemmatisation (Kennedy, 1998). Lemmatisation of signs in the corpus is made possible by the existence of the Auslan Lexical Database. It is an indispensable aid to consistent sign identification through glossing. Lexical information found in the Auslan Lexical Database is being integrated into the annotations of the corpus of Auslan texts. I follow the discussion of the corpus and database by describing some of the annotation conventions observed in the Auslan Corpus that allow for the lemmatisation of lexical signs and, equally importantly, the conventions observed in the annotation non-lexical signs. Together both sets of practices and conventions ensure that the corpus becomes, and remains, machine-readable as it is enriched over time.

## 2. Corpora and empirical linguistics

Signed language corpora will vastly improve peer review of descriptions of signed languages and make possible, for the first time, a corpus-based approach to signed language analysis. Corpora are important for the testing of language hypotheses in all language research at all levels, from phonology through to discourse (Baker, 2006; McEnery *et al*, 2006; Sampson, 2004; Sinclair 1991). This is especially true of deaf signing communities which are also inevitably young minority language communities. Although introspection and observation can help develop hypotheses regarding language use and structure, because signed languages lack written forms and well developed community-wide standards, and have interrupted transmission and few native speakers, intuitions and researcher

observations may fail in the absence of clear native signer consensus of phonological or grammatical typicality, markedness or acceptability. The past reliance on the intuitions of very few informants and isolated textual examples (which have remained essentially inaccessible to peer review) has been problematic in the field. Research into signed languages has grown dramatically over the past three to four decades but progress in the field has been hindered by the resulting obstacles to data sharing and processing.

Moreover, as with all modern linguistic corpora, it should go without saying that signed language corpora should be representative, well-documented (i.e., with relevant metadata) and machine-readable (i.e., able to be annotated and tagged consistently and systematically) (McEnery & Wilson, 1996; Teubert & Cermáková, 2007). This requires dedicated technology (e.g., ELAN), standards and protocols (e.g., IMDI metadata descriptors), and transparent and agreed grammatical tags (e.g., grammatical class labels) (Crasborn *et al*, 2007). However, it also requires the identification of lemmata. Lemmatisation—the classification or identification of related forms under a single label or lemma (the equivalent of headwords or headsigns in a dictionary)—is absolutely fundamental to the process of corpus creation. A successful corpus project team should already have available a reference dictionary or lexical database to facilitate lemma identification and consistency in lemmatisation. Without lemmatisation a collection of recordings (digital or otherwise) with various related annotation files (appended and integrated into a single multimedia file or simply related to each other in a database) will not be able to be used as a true linguistic corpus as the counting, sorting, tagging, etc. of types and tokens is rendered virtually impossible.

Annotations began in 2005 and it is anticipated that it will take at least 10 years for a substantial number of these texts to be sufficiently richly annotated for extensive corpus-based research. However, given that corpus-based signed language studies are beginning from such a low base (essentially zero), a recent initial study of 50 annotated Auslan texts from this corpus is already one of the largest of its kind (Johnston *et al.*, 2007). A second corpus-based study on the co-occurrence of pointing signs with indicating verbs is being presented at this conference (de Beuzeville & Johnston, this volume).

### 3. The Auslan Corpus

The corpus brings together into one digital archive a representative sample of a signed language in which the video recordings themselves, along with appended metadata and annotation files, are openly accessible.<sup>1</sup> Importantly, the annotation files of the corpus are designed to facilitate expansion and enrichment over time by various researchers through repeated annotation parses of individual texts.

The Auslan Corpus is built from two sources: the Sociolinguistic Variation in Auslan Project (SVIAP)<sup>2</sup> and from the Endangered Languages Documentation Project (ELDP)<sup>3</sup>. Both datasets are based on language recording sessions conducted with native or near-native users of Auslan. The SVIAP corpus consists of films of 211 participants from the five major cities in Australia (Sydney, Melbourne, Brisbane, Adelaide and Perth). This yielded over 140 hours of unedited digital video footage of free conversation, structured interviews, and lexical sign elicitation tasks. The ELDP yielded approximately 300 hours of unedited footage taken from 100 participants from the same five cities. Each participant was involved in three hours of language-based activity that involved an interview, the production of narratives, responses to survey questions, free conversation, and other elicited linguistic responses to various stimuli such as a picture-book story, a filmed cartoon, and a filmed story told in Auslan. This footage has been edited down to around 150 hours of usable language production which, in turn, has been edited into approximately 1,700 separate digital movie texts for annotation. To date approximately 100 of these texts have been annotated using ELAN (EUDICO Linguistic Annotator) (Hellwig *et al.*, 2007). In total, the corpus consists of digital movies, ELAN annotation files

and IMDI metadata files (Johnston & Schembri, 2006).

### 4. The Auslan Lexical Database

The Auslan Lexical Database, consists of over 7,000 individual sign entries and was originally created as a FileMaker Pro database file (Johnston, 2001). Lexical signs in the form of short digital movie clips are headwords/lemmas of individual records/entries in the database. There are multiple fields coding information on the form, meaning and lexical status of each headsign. Form fields include one for phonological transcription using modified HamNoSys, several for dedicated feature fields coding for handshape, location, symmetry, etc.; and one field for morphological transcription which relates variants to stem forms. Meaning fields include several for definitions, semantic domains, and synonyms and antonyms. Lexical status fields include several for dialect, register, and stem/variant identification. The database lists a citation form of a lexical sign as a major stem entry, with common variant forms listed separately.

This database also now exists in two other forms: (i) an online, open access dictionary called *Auslan Signbank* (<http://www.auslan.org.au>) and (ii) a limited access researchers' reference database which also includes variant signs and newly identified signs. The database, in both its current forms, is being constantly corrected and augmented. Finally, signs in the database are organized and sequenced formationally, i.e., according to major phonological features of signs, such as handshape and location, so that scrolling through the database records displays formationally similar signs one after the other.

The Auslan Lexical Database is the source of information for a number of dictionaries of Auslan in three formats—print, CD-ROM, and internet (e.g., *Auslan Signbank*, mentioned above). By definition, the sign data is lemmatised. It serves as the reference point for the lemmatisation of the corpus annotations. However, since the identification of lexis in any language is always open-ended, it should be noted that corpus data is also used to test assumptions underlying the lemmatisation found in the Auslan Lexical Database itself. In other words, the source database and annotations are appropriately updated as required (as described below). This strategy is one possible solution to the 'database paradox' (van der Hulst *et al.*, 1998).

### 5. Lemmatisation in the Auslan Corpus

In order for a corpus of recordings of face-to-face language in either spoken or signed modalities to be machine readable, time-aligned annotations need to be appended to the source data using some form of multi-media annotation software. It is these appended annotations which are read by machine, not the source data itself. Strictly speaking, therefore, a written transcription of the text need not be created in order to do corpus-based research. However, just as with the Auslan Lexical Database, such a level of representation would be necessary in order to

<sup>1</sup> Open-accessibility will be implemented after an initial limited access period of three years from the time of the deposit of the corpus at SOAS in 2008.

<sup>2</sup> Australian Research Council research grant awarded to Adam Schembri and Trevor Johnston — #LP0346973 *Sociolinguistic Variation in Auslan: Theoretical and applied dimensions*.

<sup>3</sup> Hans Rausing Endangered Languages Documentation Program (School of Oriental and African Studies, University of London) language documentation project awarded to Trevor Johnston — #MDP0088.

carry out phonetic or phonological research of a corpus.

With respect to identified sign units, failure to integrate lexical information into the sign identifier, either as a transcription or a gloss-based annotation, immediately creates two problems: (1) the consistency and commensurability of data that is transcribed or glossed by multiple researchers or even the same researcher on different occasions; and (2) the effective unboundedness of the sign dataset. In other words, each sign articulation which may be distinctive would have its own distinctive transcription because each form would have its own representation, or its own distinctive gloss reflecting contextual meaning. The unique identification of sign types—lemmas—would thus not be achieved and one of the prime motivations for the creation of a linguistic corpus in the modern sense would be undermined from the very outset.

### 5.1 ID-gloss vs. GLOSS vs. translation

Lexical signs need to be identified using a gloss which is intended to uniquely identify a sign. In the Auslan Corpus project this is referred to as the *ID-gloss*. An ID-gloss is the (English) word that is used to label a sign all of the time within the corpus, regardless of what a particular sign may mean in a particular context or whether it has been systematically modified in that context. For example, if a person signs HOUSE (a sign iconically related to the shape of a roof) but actually means *home*, or performs a particularly large and exaggerated form of the sign HOUSE, implying *mansion*, (without that modified form itself being a recognized and distinctive lexeme of the language) then the ID-gloss *house* would still be used in both instances to identify the sign in the annotation.

A consistently applied label of this type means it is possible to search through many different ELAN annotation files and find all instances of a sign to see how and when it is used. Only if a sign always has the same ID-gloss can we search, using computers, for how that sign is used in different ways in the corpus.

The ID-gloss is thus not meant to be a translation of meaning. So if the signer produces SUCCESS but means ‘achieve something’, it is still annotated with the ID-gloss SUCCESS; and if a person signs IMPORTANT but means ‘main’ or ‘importance’, it is still labeled IMPORTANT.

This is crucial. Without consistency in using the ID-gloss it will be impossible to use the corpus productively and much of the time spent on annotation will be effectively wasted because the corpus will cease to be, or never become, machine readable in any meaningful sense. It will not actually be the type of corpus that linguists want to have access to, i.e., a machine readable set of annotated and linguistically tagged texts (which are also representative samples of a language). It will just be a collection of reference texts, a corpus in the ‘old fashioned’ sense.

With respect to distinguishing between glossing and

translation, meaning is assigned to the text through glossing only indirectly through the unavoidable fact that the ID-gloss, which is primarily intended to identify a sign, actually uses an English word (or words) that bears a relationship to meaning of the sign. In other words, the ID-gloss is not chosen arbitrarily or randomly. It is highly motivated. However, it is not intended as a translation because within the ELAN annotation files of the corpus, translations are made on their own dedicated tiers. In assigning an ID-gloss we are simply labeling a sign so that it can be uniquely and quickly identified for subsequent tagging with linguistic markers (e.g., for grammatical class, sign modification potential, presence or absence of constructed action, semantic roles, and so on) during a later annotation parse, or searched for with or without these tags being taken into consideration. Apart from the obvious motivation of the English word used to gloss a sign, no serious attempt is being made in the assigning of an ID-gloss to translate a sign.

### 5.2 Selecting the appropriate ID-gloss for a sign

Annotators refer to the dictionary of Auslan in one of two forms—Auslan Signbank ([www.auslan.org.au](http://www.auslan.org.au)) or the Auslan Lexical Database (a FileMaker file)—to view signs and their assigned annotation ID-gloss.

If a sign in the text being annotated appears to be a lexical sign and cannot be found in the dictionary, the annotator chooses a simple English word to gloss that sign as appears to be appropriate. If the annotator cannot avoid using a word that has already been used in the dictionary as an ID-gloss they append a distinguishing number after the gloss. Thus, if HOUSE already exists in the dictionary as the ID-gloss of a sign (and there is also no ID-gloss currently used that is HOUSE2) then the new ID-gloss would be HOUSE2. Similarly, if HOUSE2 already existed as an ID-gloss, HOUSE3 would be created. After an annotation parse has been completed and the ELAN annotation file is submitted back to the corpus managers, the dictionary is updated, if necessary. For example, if a new sign is recognized as a new unrecorded sign, a new dictionary entry will be created with its own distinct ID-gloss (which may or may not be the same as the ID-gloss suggested by the original annotator).

The only time an existing sign form will be assigned a different ID-gloss is when corpus data justifies the identification of a completely distinct and unrelated meaning for the sign form. In such cases the sign form receives its own distinctive the ID-gloss and the two signs are treated as homonyms.

### 5.3 Annotation conventions: ID-glosses

The consistent use of the same ID-gloss for the same sign is the single most important act in building a machine-readable sign language corpus. It is reinforced by the adherence to a relatively small set of annotation and glossing conventions that ensure that similar types of signs are glossed in similar ways. The following are just a

few indicative examples of these types of conventions.

**Negative incorporation** If a sign incorporates a negative as part of its meaning, the main verb gloss is given first followed by a gloss for the negative element. This makes it easier to search and sort signs by meaning and name (e.g., KNOW and KNOW-NOT will be next to each other if sorted alphabetically or both will be found if sub-string search routines are used). It also means all negative incorporation is expressed the same way, rather than sometimes with words like DON'T (e.g., if glossed as DON'T-KNOW rather than KNOW-NOT) or sometimes with an entirely different word form, such as WON'T for WILL-NOT.

**Variant forms** Sometimes a sign form is clearly recognizable as a minor variant of a more common or standard form, using a slightly different handshape, movement pattern or location. These minor variations are not normally reflected in any change to the ID-gloss. Generally speaking, one does not want there to be an unnecessary proliferation of ID-glosses through attempts to encode in the gloss itself information about formational variation. Many of the possible variant forms of many signs have already been recorded in Auslan Lexical Database and are well understood. Therefore, the ID-gloss assigned these variant forms is often the same as the citation or unmarked form. However, if phonetic or phonological analysis is the focus of the annotations being created, specific phonological tiers in the ELAN annotation templates can be used utilized for this purpose. On these tiers, transcription using dedicated fonts such as HamNoSys can be used to capture the actual form of the sign. Alternatively, if the variant form noted in the textual example is unrecorded in the Auslan Lexical Database and appears to be particularly noteworthy and is not part of some grammatical modification that will be recorded on other tiers of the annotation, a brief addition to the ID-gloss can encode this. In these cases, a letter code of the handshape change is added after a hyphen (e.g., SUGAR-K would signify the sign SUGAR made with a K handshape), or a word for the variant location or the variant movement is added (e.g., KNOW-cheek signifies KNOW made on the cheek). However, all such additions to any ID-gloss should be kept to an absolute minimum and should not be done in a way that would confound search and sorting routines.

**Numbers** If a signer uses a number to refer to anything it is annotated using wordS, not digits. For example, NINE-TEEN-EIGHTY-SEVEN rather than 1987, FOUR-TEEN-YEARS-OLD rather than 14-years-old.

**Points** All ID-glosses for points begin with the initials PT (for 'point'). This allows for all pointing signs in the corpus to be identified regardless of the grammatical function that may or may not be attributed to them by various annotators. Indeed, this glossing convention enables one to collect and compare all instances of points, facilitating their subsequent relabelling if textual evidence

justifies reanalysis. Further grammatical details are given whenever possible (e.g., PT:PRO signifies 'pointing sign functioning as a pronoun', PT:DEM signifies 'pointing sign functioning a demonstrative pronoun', and PT:POSS signifies 'pointing sign functioning a possessive pronoun'). Indeed, annotations may be even more detailed. For example, PT:PRO3pl signifies 'pointing sign as a third person plural pronoun'. If the handshape changes from what is normally expected, that information is included immediately after the pt, in parentheses. For example, PT(B):PRO1sg signifies 'first person singular made with a flat handshape'. However, in many cases, it will be difficult, or even impossible, for an annotator to be able to make a very detailed grammatically rich annotation with certainty. Provided the convention of ID-glosses for pointing signs beginning with PT is adhered to then decisions about the actual function of certain pointing signs can be deferred until more textual examples are collected.

**Sign names** Sign names are prefixed with *sn:* followed by the proper name in lower case. Thus a sign name for a person called Peter would be written as *sn:peter*. Additional information may be added, but is not required. For example, if the sign name is based on fingerspelling the relevant letter(s) and a hit regarding sign form can be added after the gloss, thus: *sn:peter(-P-shake)*. If the sign name is identical in form to a lexical sign the relevant sign may be identified after the name in brackets: *sn:peter(ROCK)*.

**Foreign borrowings** Lexical signs which are clearly recent or idiosyncratic borrowings from another signed language and which are generally not considered to be Auslan signs are given best gloss possible followed by the name of the signed language. For example, the borrowed sign COOL from ASL would be written as COOL(ASL)

## 5.4 Lexical vs. non-lexical signs

Lemmatization can only apply to lexical signs. However, many signed meaning units found in natural signed language texts are not lexical signs. As a number of signed language linguists have noted one needs to distinguish at two major types of meaning units—lexical signs and non-lexical signs (e.g., Johnston & Schembri, 1999; Sandler & Lillo-Martin, 2006). *Lexical sign* is reserved for a form whose meaning in context is more than the conventionalized and/or iconic value of its components (handshape, location, etc.) within the inventory of meaning units of a given signed language in a given context, and that meaning is consistent across contexts. It is essentially, equivalent to the commonsense notion of *word* (Sandler & Lillo-Martin, 2006). The term *non-lexical sign* is reserved for a form that has little or no conventionalized or language-specific meaning value beyond that of its components in a given context (e.g., depicting or 'classifier' signs).

### 5.4.1 Annotation conventions: non-lexical signs

As with ID-glosses, a relatively small set of annotation

and glossing conventions need to be adhered to in order to ensure that similar types of non-lexical signs are glossed in similar ways. Without such conventions, these categories of signs cannot be easily extracted from the corpus for analysis and comparison. The following are just a few indicative examples of these types of conventions.

**Depicting signs** These ‘do it yourself’ signs are not listed in signed language dictionaries because their meaning is too general or context specific to be given a meaningful entry description. In the Auslan corpus all such signs begin with *pm* (for “property marker”) as the handshape shows a property of the object.<sup>4</sup> Since handshape is a very salient feature of depicting signs it is included in the annotation gloss for these types of signs in the following format

—  
*pm(handshape):brief-description-of-meaning-of-sign.*

For example an upright index finger representing the displacement of a person would be annotated thus: *pm(1):person-walks*. One does not need to annotate full details of the form of the depicting sign in order to create a grammatically useful annotation because the form of the sign is visible in the video that is always attached to the ELAN annotation file. However, should such information be important, it belongs on separate tiers of the annotation file dedicated to encoding phonetic and phonological information about individual signs.

**List buoys** A list buoy is a hand which is held throughout a stretch of discourse, usually on one’s left (or weak) hand, and uses count handshapes to mark the movement to each of a sequentially related set of entities or ideas. The handshape can be held in space throughout the articulation of each item, or appear and reappear if two-handed signing demands it be removed in order to produce certain signs. The signer usually grabs or points to a relevant finger of the buoy for each item in the list. The buoy is prefixed with *buoy* (or simply the letter *b* for ‘buoy’) followed by a label of the handshape being used in brackets and, after a colon, a short description of what it stands for. So an index finger held up to indicate the first of a series of items would be annotated: *buoy(1):first-of-one* or *b(1):first-of-one*. As each finger is added for each item they are annotated accordingly in turn: *buoy(2):second-of-two* or *buoy(3):third-of-three*. If the handshape anticipates all of the members of a series by holding up two, three, four, or five extended fingers throughout, the range is stated: *buoy(8):three*. In this latter case especially, but it is also possible in the other instances, the dominant hand may simultaneously point at a specific finger of the buoy (or it may hold it). This is annotated on the dominant hand according to the finger identified and whether it is a pointing or holding action (e.g., *PT:buoy-third-of-five* or *HOLD:buoy-third-of-five*). If

<sup>4</sup> This terminology is borrowed from Slobin and Hoiting. However, any abbreviation, consistently applied, would be appropriate (e.g., *cl:* for ‘classifier sign’, or *d:* for ‘depicting sign’).

the dominant simply points to the entire buoy, it is annotated as *PT:buoy*. There is no need to repeat information about the buoy itself (handshape and/or number of entities) on the annotation for the dominant pointing hand because the annotation for the subordinate (weak) hand will have that information about the buoy already coded.

**Fingerspelling** Any time a signer uses fingerspelling, the word is prefixed with *fs:* for ‘fingerspelling’ followed by the word spelled, thus—*fs:word*. If not all the letters of a word are spelled, and it is clear what that word is, the omitted letters are put in brackets—*fs:wor(d)* not *fs:wor*. If the fingerspelling is for multiple words, a new annotation is begun for each word even if it is one continuous act of fingerspelling—*fs:mrs fs:smith* not *fs:mrssmith*. By following these conventions, it is easier for the number of fingerspellings to be counted and the types of words that are fingerspelled to be identified. If the form of a lexical sign is a single fingerspelled letter which could mean various things, the letter is followed by the word it stands for—*fs:m-month, fs:m-minute, fs:m-mile*.

#### 5.4.2 Annotation conventions: gesture

A gesture is neither a lexical sign nor a non-lexical sign. Gestures are quite common in naturalistic signing. As with depicting signs, when identifying or glossing a gesture one need not describe the form of the gesture on the sign identification (glossing) tier. The form of the gesture is visible in the associated movie or can be coded on separate dedicated phonetic or phonological tiers in the annotation file. One would thus write something like *g:how-stupid-of-me* not *g:hit-palm-on-forehead*.

## 6. Conclusion

No claim is being made here that the specific glossing conventions used in the Auslan Corpus should form the basis of a standard for all signed language corpora. Though consistency across signed language corpora in annotation protocols would facilitate cross linguistic comparisons and thus be extremely desirable, the most important considerations in the first instance are the principles of lemmatisation and consistent treatment (glossing) of various sign types.

However, there is no escaping the observation that any attempt to build a linguistic corpus, in the modern sense, of a signed language without reference to, or without the prior existence of, a relatively comprehensive lexical database of the language in question could well be plagued by difficulties. It would be extremely difficult, if not impossible, to control the proliferation of glosses referring to the same sign without a lexical database that is arranged by, or searchable on, formational or phonological criteria. This principle is fundamental to the entire enterprise of corpus creation in signed language linguistics. Without lexical resources of this type, plans to create signed language corpora are unlikely to produce anything resembling what is today commonly understood by a linguistic corpus.

Linguists need to be able to identify each sign form uniquely and this must be done by sorting sign forms phonologically. This is the role of the lexical database. Without this, one could not locate and compare sign forms in order to determine if a new unique gloss is required for a particular sign form rather than just the association of an additional sense to an existing one. Once again this is a piece of information to be added the lexical database, not included in the annotation at the ID-gloss level. To a computer using searching or sorting routines on a corpus, non-uniquely identifying glosses would be next to useless.

The lexical database and its representation in dictionaries in various forms, is thus an unavoidable prerequisite for creation of a viable corpus. However, it need not be exhaustive. After all, it is highly likely a corpus will actually reveal unrecorded lexical signs which need to be added to the reference lexical database.

## 7. References

- Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
- Crasborn, O., Mesch, J., Waters, D., Nonhebel, A., van der Kooij, E., Woll, B., et al. (2007). Sharing sign language data online: Experiences from the ECHO project. *International Journal of Corpus Linguistics*, 12(4), 535-562.
- Hellwig, B., van Uytvanck, D., & Hulsbosch, M. (2007). EUDICO Linguistic Annotator (ELAN). <http://www.lat-mpi.eu/tools/elan/>
- Johnston, T. (2001). The lexical database of Auslan (Australian Sign Language). *Sign Language & Linguistics*, 4(1/2), 145-169.
- Johnston, T., & Schembri, A. (1999). On defining lexeme in a sign language. *Sign Language & Linguistics*, 2(1), 115-185.
- Johnston, T., & Schembri, A. (2006). Issues in the creation of a digital archive of a signed language. In L. Barwick & N. Thieberger (Eds.), *Sustainable data from digital fieldwork: Proceedings of the conference held at the University of Sydney, 4-6 December 2006* (pp. 7-16). Sydney: Sydney University Press.
- Johnston, T., & Schembri, A. (2006). *The use of ELAN annotation software in the Auslan Archive/Corpus Project*. Paper presented at the Ethnographic Research Annotation Conference, University of Melbourne, Victoria, Australia (February 15-16).
- Johnston, T., de Beuzeville, L., Schembri, A., & Goswell, D. (2007). *On not missing the point: Indicating verbs in Auslan*. Paper presented at the 10th International Cognitive Linguistics Conference, Kraków, Poland (15-20 July).
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London and New York: Longman.
- McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, T., Xiao, R., & Tono, Y. (Eds.). (2006). *Corpus-Based Language Studies*. London and New York: Routledge.
- Sampson, G. (2004). *Corpus Linguistics: Readings in a Widening Discipline*. London: Continuum.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Teubert, W., & Cermáková, A. (2007). *Corpus Linguistics: A Short Introduction*. London: Continuum.
- van der Hulst, H., Crasborn, O., & van der Kooij, E. (1998, December). *How SignPhon addresses the database paradox*. Paper presented at the Second Intersign Workshop, Leiden, The Netherlands.

# Interactive HamNoSys Notation Editor for Signed Speech Annotation

Jakub Kanis, Zdeněk Krňoul

Department of Cybernetics, Faculty of Applied Sciences,  
University of West Bohemia, Pilsen, Czech Republic  
{jkanis, zdkrnoul, campr, mhruz, zelezny}@kky.zcu.cz

## Abstract

This paper discusses the practice with an annotation of signs of signed speech and the creation of a domain-specific lexicon. The domain-specific lexicon is primarily proposed for an automatic signed speech synthesizer. The symbolic notation system based on HamNoSys notation has been adopted as a perspective solution for this purpose. We have developed two interactive editors: SignEditor and SLAPE which allow to create and to expand the lexicon. The first one is intended for the direct insertion of notation symbols and the second one is for more intuitive notation through a graphical interface. The sign notations in both editors can be immediately converted into the avatar animation which is shown in the 3D space. It allows annotators who have no rich experiences with symbols organization to notate signs more precise. At present, our lexicon contains more than 300 signs. This initial lexicon is targeted to the domain of information systems for train connections. Further expansion will cover new areas where the automatic signed speech synthesizer can be used.

## 1. Introduction

The barrier in the communication between hearing impaired and hearing people should make everyday complications. The problem is that hearing people are usually not familiar with the signed speech while deaf people with the majority language. Our research aim concerns on everyday communication systems. To cope with this problem needs combination of many knowledges from different research areas, for example, the audiovisual and the signed speech recognition (Campr et al., 2007; Campr et al., 2008), the audiovisual speech (talking head) and the signed speech synthesis (Železný et al., 2006; Krňoul et al., 2008), and the bidirectional translation between the majority and the signed speech (Kanis et al., 2006; Kanis and Müller, 2007).

The goal of an automatic signed speech synthesizer is to create an avatar which uses the signed speech as a main communication form. In order to emulate the human behavior during the signing the avatar has to express manual components (hand position, hand shape) and non-manual components (face expression, lip articulation) of the performed signs. The task of the signed speech synthesis is implemented in several steps. The source utterance has to be first translated into the corresponding sequence of signs since the signed speech has different grammar than the spoken one. Then it is necessary to concatenate the relevant isolate signs to create the continuous signed speech utterance. The non-manual components should be supplemented by the talking head which is for example able to articulate the words from the utterance in the case of the Signed Czech (SC) or express the face gestures in the case of Czech Sign Language (CSE).

This paper describes experiences with a representation and a collection of the signs for an avatar animation. A lexicon of isolated signs in appropriate representation is the necessary part of the synthesis system. The everyday communication system intended to a certain domain involves that the lexicon includes the relevant signs only. A notation editor is one possibility how to create and administrate such a domain-specific lexicon of the relevant signs.

## 2. Synthesis System Background and Data Acquisition

The straightforward solution of the signed speech synthesis should be based on video records of a real signing human. A concatenation of these records has better quality and realism than the avatar animation. On the other hand, the choice of the avatar animation allows the possibility of low-bandwidth communication, arbitrary 3D position and lighting, and the possibility of a change of an appearance of the animation model. There are two ways how to automatically solve the problem of the signed speech synthesis. The first one is based on the record of the real human motions in the 3D space and is called *data driven synthesis*. The second one is based on a symbolic notation of signs and is called *synthesis from the symbolic notation*. Each solution has certain advantages and disadvantages (Elliott et al., 2000; Kennaway, 2001).

The recorded data in data driven synthesis are processed and directly applied to the animation process. The advantages are the obtaining of the full 3D trajectories and the realistic motions of the animation model but the low accuracy and extensibility of recorded signs considered as the disadvantages. In addition, we need a special and expensive equipment to obtain the proper data. For example, the arm motions are recorded by some specialized motion capture systems<sup>1</sup>. The various shapes of hands have to be simultaneously measured by two data gloves<sup>2</sup> and for the acquisition of the face gestures we have to use the other motion capture system fixed on speakers head. The advantages of the synthesis from the symbolic notation are accuracy of the generated trajectories, easy editing of symbols, and easy extensibility by new notation features. A lexicon for the synthesis can be composed from different sources and created at different times. The disadvantages are complicated conversion of symbols to the articulation trajectories and the animation which looks robotic.

The reason for the choice of the symbolic notation is to provide the decomposition of signs to the smallest units,

<sup>1</sup>Vicon, Qualisys, BTS

<sup>2</sup>CyberGlove

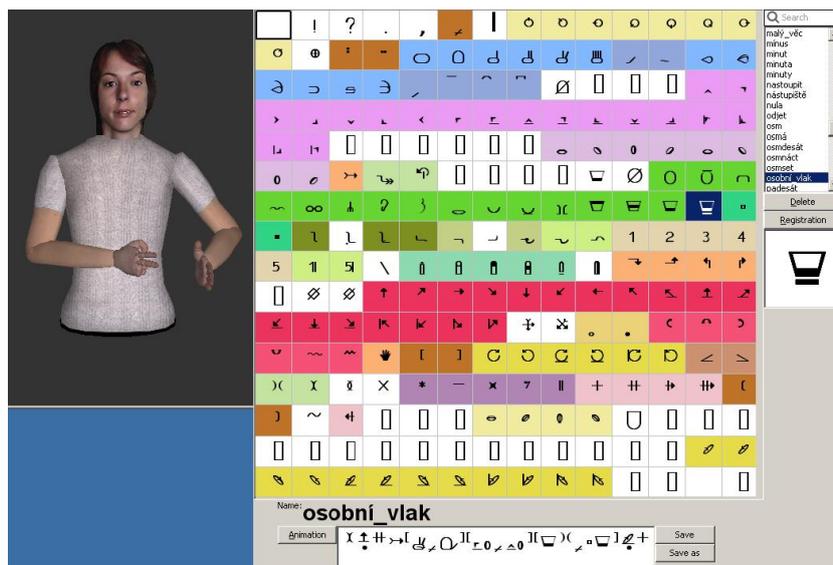


Figure 1: The screen shot of SignEditor. The notation example of sign “passenger train”.

components of the sign. This decomposition is essential for another linguistic research of a sign language. We have to mention that there is no universal sign language and the sign languages are not derived from spoken ones. For example, CSE has the specific morphology, phonetics and grammar. The basic item of CSE is the sign as in other sign languages. The sign mostly matches one word or concept in spoken language but this do not hold true in any case. The main difference between spoken Czech and CSE is that the CSE is visual-spatial language. It means that CSE is not perceived by ears but eyes, is based on shapes and motions in space. For example, the hand shapes are combined with the finger orientations in particular relationships between the dominant and the non dominant hand. In the case of the 3D trajectories acquired by the motion capture system this decomposition of sign can not be easily made. The question is how to transform these trajectories and representation for the “same” sign which is signed in different place of the sign space.

Hence, we have designed the rule based synthesizer (Krňoul et al., 2008) which uses the lexicon based on the symbolic notation. Two sign editors for administration of the lexicon are presented (Section 3.). The first editor is intended for the direct insertion of notation symbols and the second one is for more intuitive notation trough a graphical interface. Both editors share a feedback given by the avatar animation as support for the created or edited signs.

### 3. Notation System and Editors

We consider the following assumption for the notation system. Each sign is composed from two components: the manual and non-manual. The non-manual component expresses the gesture of face, the motion and position of head and other parts of upper body. The manual component is expressed by shapes, motions and positions of hands. The signs are realized in a sign space. The sign space is approximately specified by the top of head, elbows sideways raise, and horizontal line below stomach.

We can found several notations for a general purpose and also for gestures of various sign languages (Stokoe et al., 1976; Liddell and Johnson, 1989; Macurová, 1996; Rosenberg, 1995). The majority of notations comes from the linguistic research of various sign languages where they substitute the written form. We have made the analysis of these notations with primal interest in the manual component and with the respect to the notation ambiguity for an automatic computer processing. The Hamburg notation system (HamNoSys) (Hanke and Schmalig, 1989) was chosen. HamNoSys version 3.0 was preferred for a low degree of ambiguity, good meaning of symbols, description of arm movements and hand shapes. However, we consider that a converter from one notation system to other should be developed in our future work.

#### 3.1. SignEditor

This editor is intended for a direct notation of the signs in HamNoSys symbols. The main component of the editor is a table of all defined HamNoSys symbols (it is just only a character map of the HamNoSys font<sup>3</sup>). The symbols are divided into color groups which associate the symbols with a similar function. For example, the symbols for hand shapes are blue, the symbols for location are green, etc. The user can choose the particular symbols by double clicking on the picture of the symbol in the table. The selected symbols are directly entered into the edit line below the symbol table (the standard edit commands can be used in this line). The created sign can be named, saved, and processed to get its spatial form (the feedback avatar animation). The editor allows browsing the created lexicon and searching for the particular signs too. In Figure 1 is the screen shot of the SignEditor. There is the feedback animation on the left, the symbol table in the center, and the browsing window on the right.

<sup>3</sup>Available at <http://www.sign-lang.uni-hamburg.de/Software/HamNoSys/HamNo33.zip>

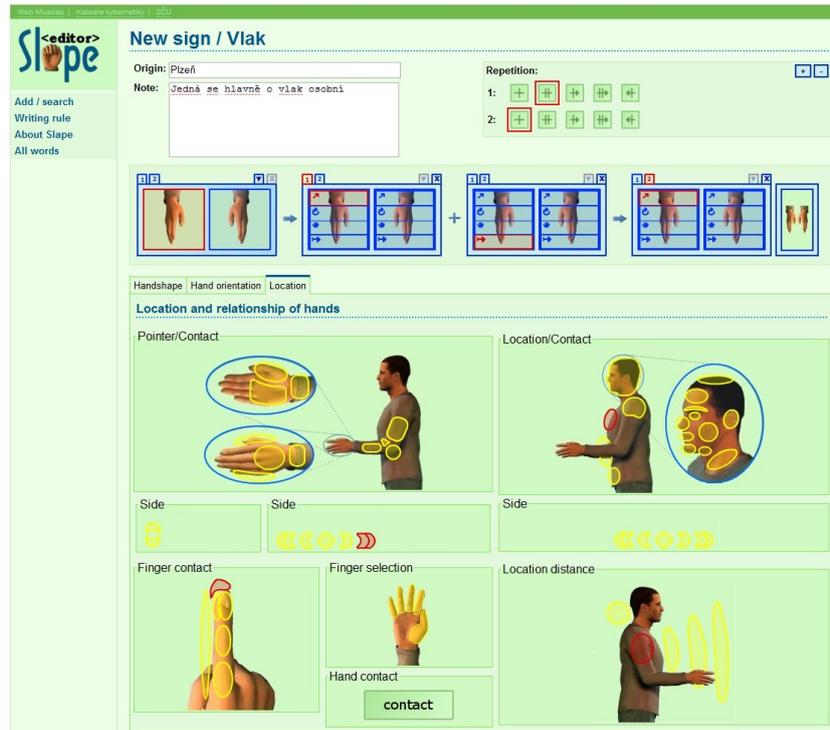


Figure 2: The screen shot of SLAPE editor. The notation example of sign “passenger train”.

### 3.2. SLAPE

The direct notation requires the full familiarity with the given notation system. Therefore we have developed the editor SLAPE (Sign Language Portable Editor) to make the notation available for all users (including hearing impaired). The main role of SLAPE is expansion of the sign lexicon just by the hearing impaired users. SLAPE enables to notate new signs in a simple graphic way and edit already saved signs. The notation process requires only fundamental familiarization with the symbolic notation. The sign notation consists from the selection of relevant graphical icons. These icons represent the particular sign components. The selection is repeated until the whole sign is not completed. All selections are converted to the representation in the predefined notation system. Primarily, we have implemented the conversion to the HamNoSys.

The principle of the notation by SLAPE editor is based on the items which are arranged to an arbitrary length sequence. Each item consists of two panels for the dominant and non dominant hand. Users can use one or both panels to select icons for particular hand shape, orientation, location, movement or select icons for a hand symmetry. The items are successively filled according to passage of notated sign. The connection determines the time relationship of the neighboring items. By clicking on connection, user can determine which items will be performed sequential or simultaneous. The items can share additional properties, for example the repetition or movement modalities. The screen shot of editor is depicted in Figure 2.

The SLAPE is implemented as a client-server web application. The server is implemented in Java. It executes user's requires and provides a storing of the signs in the database. The Hibernate tool is used to implement the ob-

ject relation mapping on the persistent layer for the communication with the database. The JBoss Seam framework is used as base structure to integrate Java Server Faces and Facelets tools. The client is implemented in the HTML and JavaScript code and runs in an arbitrary web browser. The Flash technique is applied for the design of notation forms and icons. The client provides good portability on various operation systems and platforms.

## 4. Feedback Animation

The usage of HamNoSys notation without any feedback makes the possibility of the structural mistakes. Therefore, our editors are supplemented with a feedback module to provide the correctness of the notation and immediately visualization of the created sign. The feedback module can be divided to a module for a rendering of the animation model and to a module for a forming of the animation trajectories (a trajectory generator).

### 4.1. Rendering of Animation Model

Our animation algorithm employs a 3D geometric animation model of the avatar in compliance with the H-Anim standard<sup>4</sup>. The animation model covers 38 joints and body segments. Each segment is represented as textured triangular surface with the normal vector per vertex. The segments are connected by joints of an avatar skeleton. One joint per segment is sufficient for this purpose. A controlling of the skeleton is carried out through the rotation of joints around three axes ( $x, y, z$ ). The rotations of the shoulder, elbow, and wrist joints are not directly controlled but they are completed from 3D positions of the wrist joints. The inverse

<sup>4</sup>Available at [www.h-anim.org](http://www.h-anim.org).

		<i>hamsym.dat</i>
<b>Location</b>	Pointer segment of body Index of pointer segment Location segment Array of location index Index to array of location Distance from location segment	HAMSYM $\cdot$ finger 180.0 90.0 0.0
<b>Action</b>	Relative change (x,y,z) Type of the motion Size of amplitude Amplitude gain Turning of motion amplitude Angles for circle sector	HAMSYM $\curvearrowright$ locsegname hanim_15 idxloc 121 398 21 123 whichidloc 2 distance 0.4
<b>Orientation</b>	Orientation of wrist	HAMSYM $\sim$ typemov zigzag turn 0.0 amplit 1.0
<b>Handshape</b>	Handshape vector Finger flexion Mask for finger selection Thumb shape	HAMSYM $\cdot$ change 0.2 0.0 0.0

Figure 3: Left panel: The list of all items. Right panel: An example of the items stored in the definition file.

kinematics<sup>5</sup> is employed to perform the analytic computation in the real time.

The further control of the animation model is performed by a local deformation of the triangular surfaces. The local deformations are for the detail animation of an avatar pose. It is primarily used for the animation of the avatar's face and the tongue. The triangular surfaces are deformed according to the designed animation schema in our synthesis approach (Krňoul et al., 2008). The deformation of the given triangular surface is defined by a transformation individually given for each vertex. These transformations are derived from influence zones defined on the triangular surface by spline functions which are constructed from several 3D points. The rendering of the animation model is implemented in C++ and OpenGL language. The animation model is shown in Figure 1 on the left.

#### 4.2. Trajectory Generator

The HamNoSys is detailed enough. However, it is difficult to define some rules and actions for all symbol combinations to cover the entire notation variability. We made a few restrictions in order to preserve maximum degree of freedom. In this assumption, the annotation of the sign have a good meaning for the user familiar with HamNoSys as well as signs are obvious enough for the transformation to the avatar animation.

The trajectory generator automatically carries out the syntactic analysis of the symbolic string on the input and creates a tree structure. For structurally correct symbolic string, we have one parse tree where each no leaf node is determined by one parsing rule. Each node of the tree is described by two key frames to distinguish the dominant and non dominant hand. The structure of the key frame is composed from a set of items specially designed for this purpose (Figure 3 on the left). These items are filled in each leaf node from the symbol descriptors stored in the definition file (Figure 3 on the right). Currently, the definition file covers 138 HamNoSys symbols. The generator uses 374 parse rules to perform syntactic analysis of the input string. In addition, the 39 rule actions were added in manner that one rule action is connected with each parse rule. The number of used symbols, parse rules, and actions is in Figure 4.

Block	HamNoSys Symbols		Parser	
	#Base	#Auxiliary	#Rules	#Actions
Symmetry	4	0	8	8
Handshape	12	11	35	6
Finger and palm orientation	26	3	34	5
Location	30	17	85	14
Action	40	42	204	15
Link of blocks	0	0	8	1

Figure 4: The statistic of symbols, rules, and actions used by the HamNoSys parser.

The processing of the parse tree is carried out by several tree walks whilst the size of the tree is reduced. The initial tree walks put together the items of the key frames according to the type of the rule actions. The reduced tree is processed by the next tree walks to transform the key frames to the trajectories accordance with the timing of the particular nodes. Finally, we obtain the final trajectories for both hands in the root node of the tree. The final step is transforming the trajectories into the avatar animation.

The acceptance of signs defined as a string of HamNoSys symbols by the parser causes some limitations. The order of the HamNoSys general notation structure defined as block sequence of a symmetry operator, starting point configuration, and actions is completely preserved. For the block of the starting point configuration, the hand shape and a finger orientation is without any restriction as well as the block of symmetry operators in all eight variants. The variants of hand location for separate pose of dominant or non dominant hand agree to HamNoSys body location symbols table. The only limitation is in the notation of two handed locations. The location symbols for finger, hand and arm are involved in relation to two handed location where the notation of the precise contact is extended. We have implemented two precise variants of the relationship of the dominant hand:

- Relationship between the dominant hand and body: We have to select one symbol to determine pointer location of the dominant hand, further symbol to determine the target body location and finally symbols to define the type of notated relationship.
- Relationship between the dominant hand and non dominant hand: We have to select one symbol to determine pointer location of the dominant hand, one symbol to determine the dominant hand target location, further the symbol for the type of the relationship and finally the symbol for the hand location.

The example of annotation is showed in Figure 5. In contrast to HamNoSys definition, the type of the relationship should be one of the list: behind the body, in contact with body, near to the body or the farthest distance.

Such relationship of hands can be used for arbitrary hand segment and location in correspondence with the notation obvious for synthesis process. The block of starting point configuration is followed by block of actions or as well block of movements. The HamNoSys definition of movements on absolute and relative is preserved too. The rela-

<sup>5</sup>Available at [cg.cis.upenn.edu/hms/software/ikan/ikan.html](http://cg.cis.upenn.edu/hms/software/ikan/ikan.html).



ant is not enough in this case. There could be more possibilities how to annotate more precise this location. These experiences partially agrees with comments by inspectors who check signs in the lexicon. The most frequent inspector's comments are related to incorrect notation using symbols for:

- hand shape including configuration of thumb
- finger or palm direction
- location given by thorax symbol
- number of repetition
- speed of movements

Nevertheless, the incorrect notation is caused by a collision of dominant hand with body. The some collided signs can be corrected the notation sequence for the precise contact. However, some collisions are caused by moves of hands in the proximity of body and can not be thus corrected in the same way.

The several limitations of notated signs are also caused by missing features in current implementation of the trajectory generator. The missing rule action for movement modality, for example the notation of fast or slow movements, are very important and have to be implemented for the following lexicon creation. The important feature, which should be included too, is the symbol sequence contains “between” symbol. Thus several hand shapes and locations should be repair and represent more precisely. The notation variant with symbols for contact in action block is not yet implemented. This variant will be implemented by a solving of the body segment collisions in more general way to avoid all possible collisions occurred in the synthesis process.

## 6. Conclusion

We have discussed the experiences with the domain-specific lexicon for the automatic signed speech synthesis. Two editors for notation of signs in HamNoSys symbols were introduced. The first one is SignEditor which is intended for the direct insertion of the notation symbols. The second one is SLAPE which is designed for more intuitive notation trough the graphical interface and for the good portability. Both editors share a feedback given by the avatar animation as support for the created or edited signs. The SignEditor was used to create our lexicon for the railway station domain. Nowadays, the lexicon contains more than 300 signs of CSE.

## 7. Acknowledgments

This research was supported by the Grant Agency of Academy of Sciences of the Czech Republic, project No. 1ET101470416 and by the Ministry of Education of the Czech Republic, project No. MŠMT LC536.

## 8. References

Pavel Campr, Marek Hruží, and Miloš Železný. 2007. Design and recording of czech sign language corpus for automatic sign language recognition. In *Proceedings of the Interspeech 2007*. Antwerp, Belgium.

- Pavel Campr, Marek Hruží, Jana Trojanová, and Miloš Železný. 2008. Collection and preprocessing of czech sign language corpus for sign language recognition. In *LREC 2008*. Workshop proceedings: Construction and Exploitation of Sign Language Corpora, ELRA.
- R. Elliott, J.R.W. Glauert, J.R. Kennaway, and I. Marshall. 2000. The development of language support for the visicast project. In *4th Int. ACM SIGCAPH Conference on Assistive Technologies (ASSETS 2000)*. Washington DC.
- Tomáš Gura and Václav Ptáček, 1997. *Dictionary of Czech Sign Language (in Czech)*. ČUN, Praha, Olomouc, Czech Rep.
- Thomas Hanke and Constanze Schmalig, 1989. *HamNoSys, Version 3.0*. University of Hamburg, <http://www.sign-lang.uni-hamburg.de/Projects/HamNoSys.html>.
- Jakub Kanis and Luděk Müller. 2007. Automatic czech - sign speech translation. *Lecture Notes in Artificial Intelligence 4188*.
- Jakub Kanis, Jiří Zahradil, Filip Jurčíček, and Luděk Müller. 2006. Czech-sign speech corpus for semantic based machine translation. *Lecture Notes in Artificial Intelligence 4188*.
- J.R. Kennaway. 2001. Synthetic animation of deaf signing gestures. In Ipke Wachsmuth and Timo Sowa, editors, *Lecture Notes in Artificial Intelligence*, pages 146–157. London.
- Zdeněk Krňoul, Jakub Kanis, Miloš Železný, and Luděk Müller. 2008. Czech text-to-sign speech synthesizer. *Machine Learning for Multimodal Interaction, Series Lecture Notes in Computer Science*, 4892:180–191.
- Jiří Langer, Václav Ptáček, and Karel Dvořák, 2004. *Vocabulary of Czech Sign Language (in Czech)*. Univerzita Palackého v Olomouci, Olomouc, Czech Rep.
- S K Liddell and R E Johnson. 1989. American sign language: the phonological base. *Sign Language Studies*, 64:195–277.
- Alena Macurová. 1996. Why and how to notate signs of czech sign language (notes to discussion) (in czech). *Speciální pedagogika*, 6:5–19.
- Miloš Potměšil. 2004. *Dictionary of Czech Sign Language O-Ž (in Czech)*. Fortuna, 1 edition.
- Miloš Potměšil. 2005. *Dictionary of Czech Sign Language A-N (in Czech)*. Fortuna, 1 edition.
- Amy Rosenberg. 1995. Writing signed languages in support of adopting an asl writing system. Master's thesis, University of Virginia.
- William C. Stokoe, Carl Croneberg, and Dorothy Casterline. 1976. *A Dictionary of American Sign Language on Linguistic Principles*. Silver Spring, 2 edition.
- Miloš Železný, Zdeněk Krňoul, Petr Císar, and Jindřich Matoušek. 2006. Design, implementation and evaluation of the czech realistic audio-visual speech synthesis. *Signal Processing, Special section: Multimodal human-computer interfaces*, 86:3657–3673.

## Corpus-based Sign Dictionaries of Technical Terms – Dictionary Projects at the IDGS in Hamburg

**Lutz König, Susanne König, Reiner Konrad, Gabriele Langer**

Institute of German Sign Language and Communication of the Deaf (IDGS), University of Hamburg  
Binderstr. 34, 20146 Hamburg, Germany

{Lutz.Koenig,Susanne.Koenig,Reiner.Konrad,Gabriele.Langer}@sign-lang.uni-hamburg.de

### Abstract

In this paper we give an overview on the six corpus-based sign language dictionaries of technical terms produced by the lexicographical team at the IDGS in Hamburg. We shortly introduce the different work steps. Then we focus on those work steps, which deal with or rely on corpus data. The consistent token-type matching and annotating accomplished during the transcription process allows for comparing the transcribed answers and for evaluating them quantitatively and qualitatively. Based on this analysis appropriate signed translations of technical terms are selected. In the dictionaries all single signs included in the selected answers are listed and described as they would be in a general sign language dictionary. During the process of transcription, selection and analysis assumptions and practical decisions have to be made. We discuss some of the assumptions and decisions that have proven valuable over time, as well as some open questions.

### 1. Projects

At the Institute of German Sign Language (IDGS), six dictionary projects in such diverse technical fields as computer technology, psychology, joinery, home economics, social work, and health and nursing care have been carried out. A seventh project on landscaping and horticulture is in progress.

Six of the seven dictionaries are based on corpus data collected from deaf experts in the respective fields.

Elicitation methods, such as interviews and picture prompts, corpus design as well as annotation, transcription, sign analysis and dictionary production have been continually developed and refined over the years. Many procedures rely heavily on the use of a relational database (iLex; see Hanke & Storz, this volume).

The following table provides an overview on the the six projects and their elicited corpus data:

	Psychology	Joinery	Home Economics	Social Work	Health and Nursing Care	Landscaping and Horticulture
Timeframe	1993-1995	1996-1998	1998-2000	2001-2003	2004-2007	2006-2009
Number of technical terms	900	800	700	450	1000	710
Number of signed translations included in the dictionary	1270	2800	1560	940	2330	
Stimuli:						
• written terms	900	800	700	450	1000	710
• pictures	0	550	340	0	190	410
informants (filmed)	5	16	17	15	18	11
informants (transcribed)	5	10	11	10	10	
Hours of filmed material						
• interview	2	3,5	2	5	5	3,5
• conversation	12	19	15	9	8,5	5,5
• elicitation	7	32,5	37,5	40,5	93,5	37
answers (total)	3600	13500	12500	9600	43200	21100
answers (transcribed)		8900	9800	6800	15200	
Number of transcribed tokens (single signs)		18700	26350	15800	29500	
Number of						
• types		1370	1750	1766	1450	
• productive signs		2800	2850	50	2300	

Table 1: Figures of technical sign dictionaries and corpus data.

Each project is completed within a timeframe of about 2,5 years which allows for a coverage of 500 to 1000 technical terms. In order to provide DGS equivalents to

technical terms a corpus-based and descriptive approach has been chosen. Nearly all technical content has been produced in cooperation with experts from educational or

academic institutions of the respective field. These experts compile a list of technical terms, write the definitions for these terms and produce appropriate illustrations. All lexicographic work concerning eliciting, transcribing, analysing, presenting signed translations and single signs, and producing the actual dictionary is carried out by the lexicographical team at the IDGS. From 1996, the core team has consisted of four to six deaf and three hearing colleagues. Most team members have been working in the dictionary projects since 1996. This has facilitated a continuity of experience and know-how as

well as a continuous improvement of methods and procedures.

## 2. Work steps

The following table outlines the main work steps in our empirical approach, following a roughly chronological order. All tasks concerning technical and terminological information (e.g. definition, illustration, subject categories, synonyms) which are executed by experts in the field working in vocational or academic institutions are not listed in table 2. Also, the production steps are left out.

Work step	Tasks and procedures	Progression and results
<b>(1) Data collection</b>		
(1a) Preparation	<ul style="list-style-type: none"> <li>• Searching for deaf informants (fluent DGS signers trained and working in the field)</li> <li>• Testing equipment and studio setting</li> <li>• Elicitation material</li> </ul>	Word list (ordered by subject categories) with context information, combined with illustrations
(1b) Data collection	<ul style="list-style-type: none"> <li>• Interview (standardised)</li> <li>• Interview (pre-structured)</li> <li>• Elicitation (written terms and pictures as stimuli)</li> </ul>	<ul style="list-style-type: none"> <li>• social and linguistic background (meta-data)</li> <li>• conversational data</li> <li>• spontaneous responses</li> </ul>
<b>(2) Definition of the corpus</b>		
(2a) Documentation and segmentation	<ul style="list-style-type: none"> <li>• Formatting digitised material</li> <li>• Linking films to database (iLex)</li> <li>• Conversational data: Segmentation and description of content; Tagging (linking to terms)</li> <li>• Elicited data: Segmenting in subject categories; Tagging (linking to terms)</li> </ul>	QuickTime <sup>®</sup> movies  Content: search by written German; direct access to DGS equivalents via terms Direct access to all answers via terms
(2b) Review of data	Conversational data: <ul style="list-style-type: none"> <li>• Qualitative evaluation of informants' DGS competence</li> </ul> Elicited data: <ul style="list-style-type: none"> <li>• Tagging repetitions, wrong, and odd answers for exclusion</li> <li>• Annotating informants' and transcribers' judgement of the answer</li> </ul>	Priority list for transcription  Defined corpus for transcription (including conversational data) Documentation of appropriateness of answers for selection process
<b>(3) Transcription and annotation</b>		
(3a) Token-type matching	<ul style="list-style-type: none"> <li>• Identification of lexemes, variants and modifications</li> <li>• Identification of productive signs and others (e.g. numbers, indexes, manual alphabet etc.)</li> </ul>	Direct access to tokens via types and vice versa
(3b) Annotation	<ul style="list-style-type: none"> <li>• Form (HamNoSys) for types (citation form) and tokens (variation, realisation in context, deviation)</li> <li>• Mouthing</li> <li>• Meaning (types and tokens)</li> </ul>	Search by HamNoSys  Search by mouthing Search by meaning
<b>(4) Selection of signed translations</b>		
(4a) Selection	Selection of answers and DGS equivalents as translations of the corresponding technical term	DGS translations of technical terms
(4b) Filling gaps	New combinations of signs and mouthings or coining new signs	

Work step	Tasks and procedures	Progression and results
(5) Analysis of conventional and productive signs used in the selected DGS translations		
(5a) Conventional signs	<ul style="list-style-type: none"> <li>• Empirical status</li> <li>• Sign form (citation form, variants, modifications)</li> <li>• Meaning</li> <li>• Iconic value and visualisation technique</li> <li>• Use of signing space</li> <li>• Similar and related signs (synonymous, homonymous signs)</li> <li>• Comments (e.g. dialect, variation of form)</li> </ul>	Lexical analysis and description of lexemes and productive signs
(5b) Productive signs	<ul style="list-style-type: none"> <li>• Iconic value and visualisation technique</li> <li>• Similar and related signs</li> </ul>	
(5c) Quality control	<ul style="list-style-type: none"> <li>• Consistent token-type matching</li> <li>• Consistent description of types and productive signs</li> </ul>	

Table 2: Work steps focussing corpus-related tasks.

### 3. Corpus-related tasks

Data collection, reviewing and annotating are time-consuming procedures. Due to the timeframe of 2-3 years for each dictionary project, annotation and transcription is restricted to the elicitations and conversational data of about 10 informants. This means that the corpus represents a relatively small section of all existing or possible translations of technical terms in DGS discourse. Nevertheless there are some striking arguments in favour of a corpus-based approach:

- The selection process can be based on the frequency of elicited answers.
- The transcribed data show the variety of signs, sign-mouthing and sign-sign combinations. This provides a solid basis for assumptions on sign formation and sign structure, and for decisions in the lexicographic process.
- All decisions can be traced back to the original data which allows for revision of transcription and lexical analysis.

From the joinery project on the corpus data contained suitable translations for almost every technical term so that newly coined signs make up for less than 1,5% of all given translations in each dictionary.

Over the years, elicitation techniques, documentation, segmentation, annotation and transcription have been developed and refined. Corpus-related tasks are the definition of the corpus (reviewing of data, see table 2: step 2b) and transcription (token-type matching and annotation, see table 2: step 3). Also, the selection of elicited answers (see table 2: step 4) as well as the lexical analysis and description of conventional and productive signs (see table 2: step 5) require transcribed and annotated data and are thus strictly corpus-related. In the following, we describe the tasks and procedures of these work steps in more detail.

#### 3.1 Review of data

The pre-structured interviews are segmented in question (interviewer) and response tags (interviewee) and the

contents are translated or summarised in written German. Further, sequences that correspond to technical terms are tagged so that spontaneous conversational DGS equivalents and elicited answers can be transcribed and easily compared to each other (see table 2: step 2a).

The signers' DGS competence and signing styles are evaluated on ground of the conversational data. Deaf colleagues check the following aspects of signing: general impression (naturalness and fluency of signing, comprehension), context and text structure, grammar, lexicon, mouthing, facial expression, reference to technical terms. The evaluation results in a priority list determining which informants will be transcribed first.

All elicited answers of each informant are linked to the corresponding technical term and reviewed by a second deaf colleague. First, the appropriateness is assigned to the given response. Valid answers are selected for transcription. Wrong or odd answers with regard to content and form (e.g. slips of the hand) are excluded from transcription. Also repeated, identical answers are marked and excluded from transcription. Second, the informants' judgements of their answers are documented and the answers are evaluated by the transcriber. For example: the informant shows that he is doubtful or feels uncomfortable with his signing, he wants to correct the answer, or he does not have a valid translation but wants to make a proposal. Even if the answer is spontaneous, the transcriber can judge the response as due to the elicitation setting and not likely to occur in natural signing, as a proposal or as an atypical DGS construction. Annotating informants' and transcribers' judgements provides important clues for evaluating the tagged and transcribed answers during the selection process.

#### 3.2 Transcription and annotation

During the transcription process, conventional signs (lexemes), their variants and modifications, as well as productive signs and other instances of signing, such as indexes, numbers or fingerspelling are identified, classified and annotated.

Tokens are compared to each other and to already existing types. Similar tokens with regard to form and



forms can be identified. Instances of sign modification, orientation and location in the signing space are related to the basic form by means of glossing conventions or by annotating the deviation in form of the token using HamNoSys.

### 3.4.3 Meaning

Conventionalised meanings of signs are frequently used and widespread across signers. Even out of context, the sign's form is associated with a certain meaning. Many conventional signs are combined with a mouthed word that corresponds to the intended meaning. Due to the elicitation method of using written stimuli, mouthings may occur more often than in natural DGS discourse. The informant may also be tempted to produce spontaneous sign-mouthing combinations which we consider as productive uses of conventional signs. As a third effect many responses to German compounds are sign strings that follow the sequence of the compound parts. Referring to words by mouthing and by combining signs to sign strings, are common strategies in DGS to express specific meanings, especially those of technical terms. The problem, however, is to determine the well-formedness and the degree of conventionalisation of these constructions. As long as there is no reference corpus of natural DGS discourse, decisions are primarily based on native signers' intuition.

Many signs are polysemous, i.e. one sign is used to express different meanings. This phenomenon is reinforced by the combination with different mouthings. In general, these meanings can all be related to the underlying image of an iconic sign.

In addition, the interplay of mouthing and iconicity is one reason for a high degree of lexical variation (synonymy). In different sign forms different aspects of the extralinguistic referent can be visualised. In DGS there are, for example, at least four conventional signs for the meaning 'garden'. Two signs visualise raking (one with bent fingers representing the rake by substitutive technique, the other representing the hands handling a rake by manipulative technique), another digging (flat hand representing the blade of a spade; substitutive technique) and a fourth sticking seeds or cuttings into the ground (hands representing the hands holding small objects; manipulative technique).

### 3.4.4 Iconic value and visualisation technique

The iconic value of a sign cannot be directly and exclusively determined on the basis of corpus data. However, some evidence can be drawn from the corpus. Especially modifications, which exploit the iconic value of the basic sign, can be helpful. Possible modifications can also provide clues to the visualisation technique employed. Formational elements of a sign need to be set in relation to its conventional meanings and to be compared to productive sign use, existing variants of the sign and related signs with similar forms and underlying images. For these checks corpus data serve as a reference. In addition, signers' popular explanations may also

provide valuable hints to describe the underlying image of a sign or its parts.

The iconic value is a valid criterion to distinguish sign homonyms. Signs with the same form and different underlying images are considered to be homonyms.

Iconic signs can be classified according to the visualisation technique involved (Langer 2005, König, Konrad & Langer, in preparation). Most of the signs can be analysed by three different techniques:

- Substitutive technique: The hand stands for the object or a significant part of the object e.g. the flat hand, palm down, represents a vehicle moving along a path.
- Manipulative technique: The hand stands for the hand of a person touching, holding or using an object, e.g. the fist represents the hand holding something.
- Sketching technique: The hand or part of the hand – e.g. the fingertip – works like a drawing tool, just like a pencil or a brush, tracing the shape of an object into three-dimensional space, e.g. the index finger is used to draw a circle.

### 3.4.5 Use of space

In order to enable the user to use a given sign in context, we provide information on how the sign can be modified by exploiting the signing space. Signs are divided into four categories:

- Invariable signs: The form of such a sign is fixed and cannot vary without becoming incomprehensible or changing the meaning of the sign. Most of these signs are body-anchored.
- Variable signs: The forms of most signs can vary by being orientated or located in space.
- Variable body location: This sub-group of variable signs can be modified by changing their body locations to express a more specific meaning. For example, in the citation form of the DGS sign for 'blood' or 'to bleed' the dominant hand starts near the palm of the non-dominant hand. A change of location, starting at the shoulder, means 'blood or bleeding at the shoulder'.
- Variable body and space location: Some signs can change location in space or on the body to express different meanings, e.g. with the open 5-hand one can mark a specific area in space or on the body.

### 3.4.6 Similar and related signs

Cross-references to homonymous signs and signs with similar forms are given in the sign entries. These references help to understand how forms, meanings and underlying images are interconnected and to become aware of similar signs that are not to be confused with the given sign. For example, as mentioned above some signs can be analysed as a lexicalised modification of a basic form.

In iLex cross-references to types of same and similar form can be used as another way of access to sign forms when searching a certain type by sign form during the

transcription process.

### 3.4.7 Comments

For some signs, additional information is given concerning specific use or aspects of the sign form. Dialectal variants can be identified by analysing the distribution of sign uses with regard to affiliations of informants to dialectal regions. Due to the relatively small sample sizes of our corpora, definitive dialectal surveys cannot be conducted. However, in some cases there is good evidence in the corpus for marking these signs as regional dialects.

Further, in some cases it is hard to decide whether the citation form is one- or two-handed, with or without repetition or if a circular movement is executed clockwise or anti-clockwise in the standard form. If corpus data suggests that these forms co-exist in free variation, a comment is added to the sign entry.

## 4. Assumptions and practical decisions

The central problems of analysing signed equivalents of technical terms are the identification of sign lexemes and the token-type matching, which require practical decisions based on theoretical assumptions. There are two central phenomena the lexical analysis of DGS signs has to cope with: iconicity and mouthing.

### 4.1 Iconicity

Many signs are iconic. The iconic value of a sign can be helpful to determine and differentiate sign lexemes. The underlying image of an iconic sign may in many cases be interpreted “literally” as a picture or displayed action. This “visual” interpretation of the sign often reveals one of its conventional core meanings. Different meanings of the same sign are related to each other in some way. They either can be related to the underlying image or they are derived from each other by metonymic or metaphoric processes.

As a consequence, a sign form with different meanings, which are related to each other by the underlying image of the sign, is considered one polysemous lexeme. In the sign entry the meanings are listed as shown in figure 1 (above).

For many iconic signs the underlying image can be re-activated and changed to produce a modified form. This intentional change of form often results in a more specific meaning. Similar sign forms that can be related to a conventional sign on basis of a change of the underlying image with a predictable meaning change are considered modifications of the respective conventional sign. Modifications are dependent sign forms (word forms) of a basic sign. In the dictionary modifications are listed in separate, but related entries. In the electronic version of the dictionary modifications and basic sign are cross-linked.

Signs, which differ slightly in form but are used to express the same meaning and share the same underlying image and visualisation technique, are interpreted as variant forms of each other. For example the sign

POWDERIB (see figure 1) has a variant form with a different handshape (thumb touches all other fingers).

### 4.2 Mouthing

Mouthings are not considered to be part of the sign lexeme. They refer to words of the spoken language, a different language system with other symbols (cf. Ebbinghaus & Hessmann 2001). Mouthings copy or specify the meaning of a sign and therefore have to be taken into account for determining the meaning of a sign. As a consequence, a sign covers different meanings when it is accompanied by different mouthings, i.e. it is polysemous. This is especially true for semantically underspecified iconic signs that allow for a wide range of different but related meanings.

A distinction is to be made with regard to the frequency of mouthing-sign combinations. Some mouthings are frequently used with a conventional sign. Other mouthings are added spontaneously to convey a specific meaning in a given context. In general, these mouthings are in accordance with the underlying image of the sign or with its core meaning.

We call stable mouthing-sign combinations conventional uses of a sign. Meanings consistently associated with a conventional sign are listed in the sign entry. We call spontaneous mouthing-sign combinations productive uses of a sign. These meanings are not considered conventional meanings and are therefore not listed as meanings in the sign entry.

## 5. Open questions

Even though most assumptions and decisions have proven valuable over time, some questions remain to be answered. One major problem is how to determine the degree of conventionalisation of signs. With regard to sign strings that follow the structure of a German compound, this is an even more complicated question. As of yet, we have no means to determine whether these are instances of lexicalised forms or just ad-hoc combinations of signs with the primary function of providing an adequate context for the mouthed word to facilitate lip-reading. Are there other criteria than frequency (statistical methods) to identify lexicalised sign combinations as true sequential compounds? Are there constraints for the combination of signs such as signs from different regions or signs whose underlying images do not fit the intended meaning? We expect that a larger corpus of natural DGS data can give clues to answer these questions.

Since in many vocational fields there are no or few deaf experts working together and communicating in sign language, it is hard to imagine, how decisions about conventionalised “technical signs” as equivalents to established technical terms in spoken language can be based on empirical data of natural signing. One lesson we learned from the evaluation of elicited and conversational data is that there are few conventionalised technical signs and a large variety of DGS equivalents. We decided to show these differences and to give insight into general sign formation processes.

## 6. References

- Arbeitsgruppe Fachgebärdenlexika (Ed.) (1996). *Fachgebärdenlexikon Psychologie*. Hamburg: Signum. URL: <http://www.sign-lang.uni-hamburg.de/plex> (last accessed March 2008).
- Arbeitsgruppe Fachgebärdenlexika (Ed.) (1998). *Fachgebärdenlexikon Tischler/Schreiner*. Hamburg: Signum. URL: <http://www.sign-lang.uni-hamburg.de/tlex> (last accessed March 2008).
- Ebbinghaus, H., Hessmann, J. (2001). Sign language as multidimensional communication: Why manual signs, mouthings, and mouth gestures are three different things. In P. Boyes Braem, R. Sutton-Spence (Eds.), *The hands are the head of the mouth: the mouth as articulator in sign language*. Hamburg: Signum, pp. 133--151.
- König, S., Konrad, R., Langer, G. (in preparation). 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.
- Konrad, R., Hanke, T., Schwarz, A., Prillwitz, S., Bentele, S. (2000). *Fachgebärdenlexikon Hauswirtschaft*. Hamburg: Signum. URL: <http://www.sign-lang.uni-hamburg.de/hlex> (last accessed March 2008).
- Konrad, R., Schwarz, A., König, S., Langer, G., Hanke, T., Prillwitz, S. (2003). *Fachgebärdenlexikon Sozialarbeit/Sozialpädagogik*. Hamburg: Signum. URL: <http://www.sign-lang.uni-hamburg.de/slex> (last accessed March 2008).
- Konrad, R., Langer, G., König, S., Schwarz, A., Hanke, T., Prillwitz, S. (Ed.) (2007). *Fachgebärdenlexikon Gesundheit und Pflege*. Seedorf: Signum. URL: <http://www.sign-lang.uni-hamburg.de/glex> (last accessed March 2008).
- Langer, G., Bentele, S., Konrad, R. (2002). Entdecke die Möglichkeiten. Zum Verhältnis von Mundbild und Gebärde in Bezug auf die Bedeutung in der DGS. *Das Zeichen* 59, pp. 84--97.
- Langer, G. (2005). Bilderzeugungstechniken in der Deutschen Gebärdensprache. *Das Zeichen* 70, pp. 254--270.

# Content-Based Video Analysis and Access for Finnish Sign Language – A Multidisciplinary Research Project

Markus Koskela<sup>α</sup>, Jorma Laaksonen<sup>β</sup>, Tommi Jantunen<sup>γ</sup>, Ritva Takkinen<sup>δ</sup>, Päivi Raino<sup>ε</sup>,  
Antti Raike<sup>ζ</sup>

<sup>α,β</sup> Helsinki University of Technology, Dept. of Information and Comp. Science, P.O. Box 5400, FI-02015 TKK

<sup>γ,δ</sup> University of Jyväskylä, Department of Languages, P.O. Box 35 (F), FI-40014 University of Jyväskylä

<sup>ε</sup> Finnish Association of the Deaf, Sign Language Unit, P.O. Box 57, FI-04001 Helsinki

<sup>ζ</sup> University of Art and Design, Media Lab, Hämeentie 135C, FI-00560 Helsinki

E-mail: markus.koskela@tkk.fi, jorma.laaksonen@tkk.fi, tommi.jantunen@campus.jyu.fi,  
ritva.takkinen@campus.jyu.fi, paivi.raino@kl-deaf.fi, antti.raike@taik.fi

## Abstract

This paper outlines a multidisciplinary research project in which computer vision techniques for the recognition and analysis of gestures and facial expressions from video are developed and applied to the processing of sign language in general and Finnish Sign Language in particular. This is a collaborative project between four project partners: Helsinki University of Technology, University of Jyväskylä, University of Art and Design, and the Finnish Association of the Deaf. The project has several objectives of which the following four are in the focus of this paper: (i) to adapt the existing PicSOM framework developed by the Helsinki University of Technology regarding content-based analysis of multimedia data to content-based analysis of sign language videos containing continuous signing; (ii) to develop a computer system which can identify sign and gesture boundaries and indicate, from the video, the sequences that correspond to signs and gestures; (iii) to apply the studied and developed methods and computer system for automatic and semi-automatic indexing of sign language corpora; and (iv) to conduct a feasibility study for the implementation of mobile video access to sign language dictionaries and corpora. Methods for reaching the objectives are presented in the paper.

## 1. Introduction

This paper presents four key objectives of a research project that aims to develop computer vision techniques for the recognition and analysis of gestures and facial expressions from video in order to apply them to the processing of sign language, and especially Finnish Sign Language (FinSL). The project is a collaborative effort of four project partners, all representing the leading Finnish research units in their own fields: Helsinki University of Technology, University of Jyväskylä, University of Art and Design, and the Finnish Association of the Deaf. The composition of the consortium reflects the fact that the visual analysis and computerized study of sign language is a multidisciplinary challenge that calls for expertise in a large variety of scientific fields.

## 2. Objectives of the Project

### 2.1 Methods for Content-Based Processing and Analysis of Signed Videos

The first objective of the project is to develop novel methods for a content-based processing and analysis of sign language videos, recorded using a single camera. The PicSOM<sup>1</sup> retrieval system framework (Laaksonen et al., 2002), developed by the Helsinki University of Technology regarding content-based analysis of multimedia data, will be adapted to continuous signing, to facilitate the automatic and semi-automatic analysis of sign language videos. The framework has been

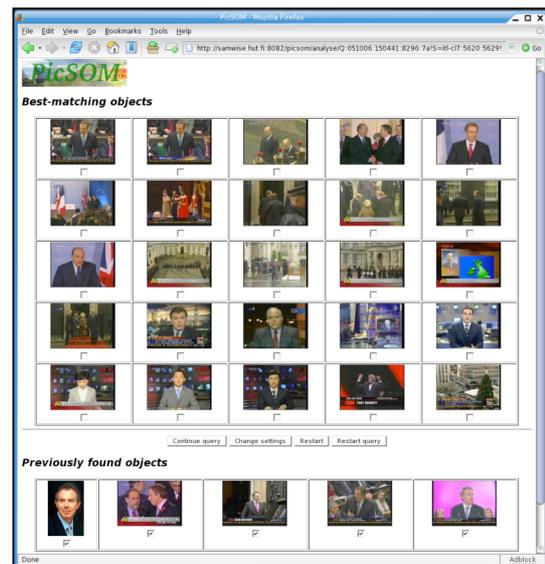


Figure 1: The user interface of PicSOM during an interactive retrieval task "Find shots of Tony Blair" from a database of recorded broadcast news.

previously applied to content-based retrieval and analysis in various application domains, including large photograph collections, broadcast news videos, multispectral and polarimetric radar satellite images, industrial computer vision, and face recognition. Figure 1 shows an example of the PicSOM user interface during interactive retrieval from a database of recorded broadcast news programs (Koskela et al., 2005).

<sup>1</sup> <http://www.cis.hut.fi/picsom/>

The PicSOM system is based on indexing any type of multimedia using parallel Self-Organizing Maps (SOMs) (Kohonen, 2001) as the standard indexing method. The Self-Organizing Map is a powerful tool for exploring huge amounts of high-dimensional data. It defines an elastic, topology-preserving grid of points that is fitted to the input space. It is often used for clustering or visualization, usually on a two-dimensional regular grid. The distribution of the data vectors over the map forms a two-dimensional discrete probability density. Even from the same data, qualitatively different distributions can be obtained by using different feature extraction methods.

During the training phase in PicSOM, the SOMs are trained with separate data sets, obtained from the multimodal object data with different automatic feature extraction techniques. The different SOMs and their underlying feature extraction schemes then impose different similarity functions on the images, videos, texts and other media objects. In the PicSOM approach, the system is able to discover the parallel SOMs that provide the most valuable information, e.g., for retrieving



Figure 2: A PicSOM analysis of a signed sequence KNOW MATTER CLEAR 'Well of course, it is obvious!' (*Suvi's* article 3, example video 6) using the standard MPEG-7 *Edge Histogram* feature.

relevant objects in each particular query. Recently, the system has also been applied to other ways of analyzing video material, i.e. shot boundary detection and video summarization (Laaksonen et al., 2007).

The existing general-purpose video feature extraction methods will provide a starting point for the analysis of recorded sign-language videos in this project. At a later stage, more specific features for the domain of sign-language videos will be developed. Figure 2 shows an example of an analysis of a signed sequence with a SOM trained using a standard MPEG-7 *Edge Histogram* image feature. The sequence is from *Suvi*, the online dictionary of FinSL.<sup>2</sup>

## 2.2 Automatic Segmentation of Continuous Sign Language Videos

The second objective of the project is to develop a computer system which can both (i) automatically indicate meaningful signs and other gesture-like sequences from a video signal which contains natural sign language data, and (ii) disregard parts of the signal that do not count as such sequences. In other words, the goal is to develop an automatized mechanism that can identify sign and gesture boundaries and indicate, from the video, the sequences that correspond to signs and gestures.

An automatic segmentation of recorded continuous sign language data is an important first step in the automatic processing of sign language videos and online applications. Traditionally, the segmentation of sign language data has been done manually by using specific video annotation programs such as ELAN<sup>3</sup> (e.g. Crasborn et al., 2007) or SignStream<sup>4</sup> (Neidle, 2001). However, identifying signs and gestures from the video this way is extremely time consuming, a preliminary segmentation of one hour of data requiring two weeks of active working time from one person at the minimum. Automating or even semi-automating this preliminary and mechanical step in the data-handling phase would facilitate the workflow considerably.

So far there have been no real attempts to identify *only* sign and gesture-like *forms* from the stream of natural signed language video. Projects dealing with sign recognition (e.g. Ong & Rangarath, 2005) have all included the semantic recognition of signs' content as one of their goals. Also, most of the research done until now has dealt only with the recognition of isolated signs from data produced specially for research purposes. In this project the semantics of signs are not directly dealt with; the objective being data and signer independent identification of signs/gestures and their boundaries.

<sup>2</sup> <http://suvi.viittomat.net/>

<sup>3</sup> <http://www.lat-mpi.eu/tools/elan/>

<sup>4</sup> <http://www.bu.edu/asllrp/SignStream/>

Linguistically, the automatic identification of signs and gestures and their boundaries will be grounded as far as possible on prosodic information. For example, linguistic boundaries in sign languages are typically indicated by changes in the facial prosody, i.e. by the changes in the posture and movement of the mouth, eyes, eyebrows, and head (e.g. Wilbur, 2000). For the automatic detection of these changes, we shall apply our existing face detection algorithm (cf. Figure 3), which is capable of detecting the eyes, nose, and mouth separately (Yang & Laaksonen, 2005).



Figure 3: An example of face detection from a recorded sign language video. The detected eyes, nose, and mouth are also shown with separate bounding boxes.

In addition to still image features extracted from single video frames, an essential feature in the analysis of recorded continuous-signing sign language is that of motion. For tracking local motion in the video stream, we apply a standard algorithm based on detecting distinctive pixel neighborhoods and then minimizing the sum of squared intensity differences in small image windows between two successive video frames (Tomasi & Kanade, 1991). An example of detected local motion is illustrated in Figure 4. The tracked points that remain stationary are not shown.

We assume that the parts of the signal where there is significantly less or no local motion correspond to the significant junctures such as the beginning and ending points of lexematic signs. However, the exact relation between motion and sign boundaries is an open research question that is essential to this objective and will be studied extensively within the research project. It can be assumed that a combination of a hand detector, still image feature extraction, and motion analysis are needed for a successful detection of sign and gesture boundaries. The PicSOM system inherently supports such fusion of different features extracted from different modalities.

During the project, the analysis of motion tracked interest points will be further developed to test the general assumption in the current signed syllable



Figure 4: An example of tracked point features marking the local movement in the sign JOYSTICK excerpted from the phrase 'The boy is really interested in playing computer games' (*Suvi's* article 1038, example video 3).

research, according to which sign internal phonological movements function as syllables' sonority peaks, that is, as the most salient parts of the signed signal (e.g. Jantunen, 2007; Jantunen & Takkinen, in press). We hypothesize that if the sonority assumption is correct, the motion tracked interest points should cumulate relatively more to the parts of the signal within the signs, not to the parts outside them.

### 2.3 Testing Methods for Indexing Existing Sign Language Material

The third objective is linked to generating an example-based corpus for FinSL. There exist increasing amounts of recorded video data of the language, but almost no means for utilizing it efficiently due to missing indexing and lack of methods for content-based access. The studied methods could facilitate a leap forward in founding the corpus. The tool for automatic processing, created and tested in this project, will be further applied to segmenting and indexing the pre-existing FinSL data in order to prepare an open-access visual corpus for linguistic research. Lacking content-based indexing and retrieval tools, the digitized data found in video magazines and online publications in FinSL covering the last 25 years has up to now been scarcely utilized within the FinSL research. It should be emphasized, however, that the functionality provided by the PicSOM system can be already used as such to analyze and index the visual content of signed video material and to construct a nonverbal index of recurrent images and video clips.

### 2.4 Implementation of Mobile Video Access to Sign Language Dictionaries and Corpora

The fourth objective is a feasibility study for the implementation of mobile video access to sign language dictionaries and corpora. Currently an existing dictionary can be searched by giving a rough description of the location, motion and hand form of the sign. The

automatic content-based analysis methods could be applied to online mobile phone videos, thus enabling sign language access to dictionaries and corpora.

In this application, it will be more essential than in the previous ones that the speed and robustness of the implementation can be optimized. We do not expect that the quality of mobile sign language videos could be good enough for accurate classification. However, we believe that by combining the automatic video analysis methods with novel interaction and interface techniques, we can take a substantial step towards a mobile sign language dictionary.

### 3. Conclusion

In this paper we have outlined the key objectives of the research project that aims to develop computer vision techniques for recognition, indexing, and analysis of sign language data. We believe that the PicSOM system developed by the Helsinki University of Technology provides an excellent basis for this task. As the project proceeds, we will explore more methods and apply the PicSOM system to the massive video data that will be the foundation of the new FinSL corpus.

### References

- Crasborn, O., Mesch, J., Waters, D., Nonhebel, A., van der Kooij, E., Woll, B., Bergman, B. (2007). Sharing Sign Language Data Online. Experiences from the ECHO Project. *International Journal of Corpus Linguistics* 12(4), pp. 537–564.
- Jantunen, T. (2007). Tavu suomalaisessa viittomakielessä. [The Syllable in Finnish Sign Language]. *Puhe ja kieli* 27(3), pp. 109–126.
- Jantunen, T., Takkinen, R. (in press). Syllable Structure in SL Phonology. In D. Brentari (Ed.), *Sign Languages*. Cambridge, UK: Cambridge University Press.
- Kohonen, T. (2001). *Self-Organizing Maps*. Third edn. Springer-Verlag.
- Koskela, M., Laaksonen, J., Sjöberg M., Muurinen, H. (2005). PicSOM Experiments in TRECVID 2005. Online Proceedings of the TRECVID 2005 Workshop. Gaithersburg, MD, USA. November 2005.
- Laaksonen, J., Koskela, M., Oja, E. (2002). PicSOM – Self-Organizing Image Retrieval with MPEG-7 Content Descriptions. *IEEE Transactions on Neural Networks*, 13(4), pp. 841–853.
- Laaksonen, J., Koskela, M., Sjöberg M., Viitaniemi, V., Muurinen, H. (2007) Video Summarization with SOMs. Proceedings of 6th International Workshop on Self-Organizing Maps (WSOM 2007). Bielefeld, Germany. September 2007.
- Neidle, C. (2001). SignStream™. A Database Tool for Research on Visual-Gestural Language. *Sign Language & Linguistics* 4(1/2), pp. 203–214.
- Ong, S. C. W., Ranganath S. (2005). Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(6), pp. 873–891.
- Suvi = Suvi – Suomalaisen viittomakielen verkkosanakirja* [Online Dictionary of Finnish Sign Language]. [Helsinki]: Kuurojen Liitto ry [The Finnish Association of the Deaf], 2003. Online publication: <http://suvi.viittomat.net>.
- Tomasi, C., Kanade, T. (1991). Detection and Tracking of Point Features. Carnegie-Mellon University Technical Report CMU-CS-91-132. April 1991.
- Wilbur, R. B. (2000). Phonological and Prosodic Layering of Nonmanuals in American Sign Language. In K. Emmorey, H. Lane (eds.), *The Signs of Language Revisited. An Anthology to Honor Ursula Bellugi and Edward Klima*, pp. 215–244. Mahwah, NJ, London: Lawrence Erlbaum Associates.
- Yang, R., Laaksonen, J. (2005). Partial Relevance in Interactive Facial Image Retrieval. Proceedings of 3rd International Conference in Pattern Recognition (ICAPR 2005). Bath, UK. August 2005.

# The Klagenfurt lexicon database for sign languages as a web application: LedaSila, a free sign language database for international use

**Klaudia Krammer, Elisabeth Bergmeister, Silke Bornholdt, Franz Dotter, Christian Hausch, Marlene Hilzensauer, Anita Pirker, Andrea Skant, Natalie Unterberger**

Centre for Sign Language and Deaf Communication  
Universitätsstraße 65-67, A-9020 Klagenfurt  
E-mail: first.name.family.name@uni-klu.ac.at

## Abstract

The Klagenfurt on-line database for sign languages "LedaSila" (Lexical Database for Sign Languages, <http://ledasila.uni-klu.ac.at/>) is designed in such a way that it is possible to present all the information which can be found in any good monolingual or bilingual (printed) dictionary. It offers the possibility to enter semantic, pragmatic as well as morphosyntactic information. Furthermore, a detailed analysis of the non-manual and manual parameters of a sign is possible. For single signs and for the afore-mentioned areas, sign language examples in form of videos can be up-loaded. The videos are not restricted to a single format, although the Apple QuickTime® video format (.mov) is recommended, as it provides the best quality in relation to size.

LedaSila offers the possibility to search for any information already contained in it (including single signs or formational parameters, naturally), to document a sign language, or analyse it linguistically. The search function is accessible to all internet users. In case someone wants to use the database for sign language documentation and/or analyses which necessitate the insertion of any data, an authorisation from the Centre for Sign Language and Deaf Communication in Klagenfurt is required.

When using LedaSila for documentation and/or analysis of a sign, a user does not have to follow a specific order when entering the data. Another advantage of this database is that the user is free to decide whether to enter data only in one field (e.g. semantics or region) or to do a full analysis of the sign. A special feature of LedaSila is the possibility to add new categories and values at any time. This is especially important for an analysis tool which is designed to be used internationally. This feature ensures that all categories and values needed for a specific sign language are available.

LedaSila can be used free of charge for non-commercial deaf and scientific issues. The only requirements for using this database are a fast Internet connection and the Internet Explorer 5.5 or higher. The database is hosted on a server of the University of Klagenfurt. All information (including videos) is stored directly on the web server. This means that using LedaSila comes with zero administration. The server at the University of Klagenfurt is operated by the central IT department ensuring data backups and other administrative server supervision. The international sign language linguistic community is invited to take advantage of this easily manageable database.

## 1. Introduction

A description of the original database on the basis of which the current version was developed can be found in Krammer et al. (2001). From the point of view of the Centre for Sign Language and Deaf Communication (Zentrum für Gebärdensprache und Hörbehindertenkommunikation, ZGH), it was imperative to provide a sign language database which offered free access. Many institutions which do sign language research operate on a relatively low budget which prevents them from purchasing the necessary software, thus making their work difficult. With the Klagenfurt database "LedaSila" (<http://ledasila.uni-klu.ac.at/>), there is now finally a free tool available for all sign language analysts as well as other interested parties.

## 2. Technical aspects

### 2.1 Programming

LedaSila is implemented as a pure web application which can be accessed via <http://ledasila.uni-klu.ac.at>. This means that no installation is necessary on the client server; the application can be fully operated on the web browser. The website was programmed in ASP (Active Server

Pages). This is a Microsoft technology which enables the programming of dynamic web pages. Server related applications which are needed for automatic notifications have been programmed in Microsoft .NET 2003 (Visual Basic).

The system is hosted on a Windows 2003 server of the University of Klagenfurt. All information (including videos) is stored directly on the web server.

### 2.2 Client requirements

#### 2.2.1. Hardware

For users who will administrate LedaSila – and will thus use the web framework – a computer with a central memory of 256 MB is recommended. The website and additional applications have been developed for a resolution of 1024 x 768 pixels.

As for the Internet connection, it is recommended that users who will administrate the application have a broadband connection.

#### 2.2.2. Software

As many people still use Microsoft's Internet Explorer for browsing web pages and this browser is also standard at the University of Klagenfurt, the design has been optimised for Internet Explorer 5.5 and higher. This

browser is recommended for an optimal use of the search function – for the entering of data it is even a requirement. LedaSila uses the Apple QuickTime® video format (.mov) as this seems to provide the best quality compared to size. Although other video formats would be possible, QuickTime videos can be displayed directly within the web page. For other video types, the corresponding player needs to be installed on the accessing clients.

### 2.3 Multi-language feature

LedaSila is a multi-language application. At this moment, the languages German and English are available and can be switched any time. When LedaSila is used in projects in other countries, new languages can be added rather easily. On request for the inclusion of a new application language, the respective project team obtains a list of phrases and words used in LedaSila for translation. The translated phrases are then included as a new language into LedaSila by ZGH. Although the complete translation could be performed within LedaSila using its National Language Support (NLS) module, it is more convenient to perform the translation in a separate editor beforehand. Minor modifications to the translation can be directly performed within LedaSila.

It is also possible to translate only the categories and values and leave labels and application messages in English. This dramatically diminishes the necessary efforts for including a new language into the system and might be sufficient for the use in academic environments, in which English user interfaces should not be a show-stopper.

### 2.4 Openness of the set of categories

The descriptive values of a sign are not hard coded, and can thus be expanded with new items at any time. New descriptive values can be directly entered in the application by using the web interface, without changing the program. At the moment, new categories and values can only be added or removed by the administrator. The administrator function is performed by a collaborator of the ZGH. If a project leader of another project wants to enter a new category or a new value, this has to be arranged with and agreed on with the administrator. All new categories and values have to be translated into all languages available in the database. This arrangement and the central administration of this function guarantee that there will be no overlapping of entries.

## 3. Practical Use of LedaSila

In brief, the database consists of two areas which will be described here: a general search function which does not require any registration (i.e. accessible to all Internet users) and a "restricted analysis area" which requires a log-in. Before choosing one of these functions, a user can select the application languages (at the moment German and English) by clicking on the respective flag icon down left (cf. Figure 1).



Figure 1: Welcome screen of LedaSila

The help texts in the database are currently available in German and English. The ZGH is working on a translation into Austrian Sign Language (Österreichische Gebärdensprache, ÖGS) and into International Sign (IS). The written help texts can be called up by clicking on the button "?" (cf. Figure 1). On the written help page, there will then appear two more buttons, labelled "ÖGS" and "IS".

### 3.1 Search function

The users may search for signs in the database in three different ways: simple search, advanced search or via the number of the sign.

#### 3.1.1. Simple search

Clicking on the button "Search" will lead to the "Simple search" page (cf. Figure 2<sup>1</sup>). In this search mask, the users may choose between the following options:

- Semantics
- Region
- Type of sign
- Place of articulation
- Hand shape
- Word field.

For the option "Semantics", the users will get an empty text field. If they type in e.g. "evening", all signs with "evening" in their semantic entries will be displayed. For some other criteria, the values may be selected from a list. As especially the categories "Place of articulation" and "Hand shape" have a large number of values, graphics were implemented in order to facilitate the choice. By clicking on the green arrow to the right of the selection field, a window containing the graphics will open. The users may click on the chosen value directly in this window; the value will then be automatically transferred into the selection field.

<sup>1</sup> Figures 2, 3, and 4 are clippings from the full screen (cf. Figure 1).

Figure 2: Simple search

After the selection of categories and values, the search will be started by clicking on the button "search". The results will be displayed as a list (cf. Figure 3).

Sign #	Semantics	Semantic field	Region	Links
# 03582	guten Abend		Wien	6 Links
# 05329	der Abend	Day time/Time	Oberösterreich, Steiermark, Wien	4 Links
# 10140	der Abend	Day time/Time	Wien	4 Links

Figure 3: Search results

When you click on the number of a sign (to the left of "Semantics"), the analysis of the sign will be displayed. For each sign entered into the database, there are two videos available: one with high (H) resolution and one with low (L) resolution. A click on the respective symbol will start the video in a new window. If a sign has variants, these will be shown in the results via links (cf. Figure 3). For example, the ÖGS-sign "guten Abend" has six variants. The users may call up the variants by clicking on the indicated total number (e.g. *6 Links*) of the variants. If the search results do not fit onto a single page, the users may navigate between the different pages via the usual "previous page" / "next page" function. Using the option "Save result list as text file" the list will be saved on a local drive.

### 3.1.2. Advanced search

If a user should need more search criteria than are offered by the simple search, they have to click on the button "advanced" which leads to the "advanced search". On this page, search criteria may be selected and combined in a user-defined way, by clicking on "add filter" (cf. Figure 4). Search criteria may be deleted at all times by using the function "delete filter". For each category, the users may choose which values should correspond (=) or not correspond (<>) to it. Additionally, it is possible to combine the search criteria with "and" or "or", according to the rule "and" comes before "or".

Figure 4: Advanced search

Similar to the simple search, the advanced search is started by clicking on "search". The results are again displayed as a list.

### 3.1.3. Entering the sign number

Each sign is automatically assigned an unambiguous number. This number allows direct access to the sign. For this, the number has to be entered into the input field "Sign #" (cf. Figure 5), followed by the enter key. The sign will be immediately displayed in a detailed view. Given that the user has the necessary authorisation, it can also be edited. This kind of search is of special advantage for analysts, because this ensures quick access to a sign.

## 3.2 Input area

In order to be able to work in the input area, the users will need an authorisation assigned by the ZGH. Together with this authorisation, the users will be given a "role" and the respective rights. In special cases, it is also possible to assign rights which are not originally defined in the role of a user. The following roles are defined in the user concept:

- An "administrator" has full administrative rights for all functions of the database. (This role is taken on by a collaborator of ZGH.) An administrator may e.g. manage users and authorisations, carry out translations, edit the sign analyses of other users and manage projects.
- A "project leader" serves as "local" user administrator. They may assign rights to other project users or edit signs which have been entered in the course of the project. They are not able to edit categories, values or specifications. Such changes can only be done after consulting the administrator.
- An "analyst" is able to enter, analyse and edit signs. They can only modify signs that have been entered by them, though.

### 3.2.1. General guidelines

After entering their user names and passwords, the users can access the input area for a sign analysis. Depending on the role of a user, there are different editing options available. For an "analyst", for example, the area "administration" will not be displayed, but the button "New Sign" will appear (cf. Figure 5).

#### 3.2.1.1. Elements of navigation in the input windows

There are two different kinds of input windows: the general presentation of a sign (cf. Figure 5), which is opened by clicking on the button "New Sign" and the

specific analysis masks. The hyperlinks which lead to the specific analysis masks are arranged in the general presentation of a sign (cf. Figure 5). By clicking on a link (e.g. Edit Semantics), the respective analysis mask is opened. In the specific analysis masks, the buttons "OK" and "Cancel" are always available in the right lower corner. By clicking on the button "OK", the entered data will be confirmed and saved; the user automatically comes back to the general presentation of the sign. All entered data can be deleted by clicking on the button "Cancel".

For a quick and consistent entering of data, there are selection windows for specific input fields. When there is a selection window, this is indicated by a small grey field with three dots right to the input field (cf. Figure 10). By clicking on this field, the window will open.

Graphic representations of the entries are always available when there is a green button with a white arrow beside the input field (cf. Figure 11). The selection window can be opened by clicking on this arrow.

#### 3.2.1.2. Entering a new sign

There are two ways to enter a new sign: either new data may be entered, or an already analysed sign may be copied with the link "Copy sign", thereby getting a new sign number. The user may then edit the data for a new sign (using the copy) without changing the entries of the original sign.

#### 3.2.1.3. Entering the data

When entering the data, the user does not have to keep to a certain order. There is also no obligation to fill in a fixed number of entries. Both a single entry and a full analysis are possible. This flexibility allows a quick input of signs into the database where they can be called up again via their number and edited. Through the possibility of a minimal entry, the database can be used either as a simple word list or for a complex scientific description.

#### 3.2.1.4. Adding new categories and values

It is a special feature of this database that new categories and values may be added at any time. This is especially important for a database intended for international sign analysis. It is possible that the existing categories/values are sufficient for the analysis of Austrian Sign Language (ÖGS), but not for a detailed analysis of other sign languages. If a new category or value is needed, the administrator must be contacted by the analyst (cf. also 2.4.).

### 3.2.2. Sign analysis

The general presentation of a sign is opened by the button "New Sign". This mask is divided into several areas where data may be entered (cf. Figure 5). The following description adheres to this division.

Figure 5: General presentation of a sign

#### 3.2.2.1. Video upload

Videos may be uploaded for the areas "Sign", "Semantics", "Pragmatics" and "Morphosyntax". A click on "Edit videos" will open a new window (cf. Figure 6<sup>2</sup>), in which the video file can be selected by using the button "Search". The correct assignment to the respective category (e.g. "Sign") has to be done in the area "Category". It is also necessary to choose the video quality (high or low); an explanatory note accompanying the video is optional.

Figure 6: Working area "Edit videos"

After all the data have been entered, the video may be uploaded by clicking on the button "Start upload". Videos are directly uploaded to the web server. Subsequent access to uploaded videos is only available via the application. LedaSila takes care about where to store videos and therefore helps to avoid inconsistencies which might occur when linking videos to different signs. A video may be deleted via the hyperlink "Delete video" at any time.

#### 3.2.2.2. Linking of signs

If a sign has one or more variants, these can be linked to each other, so that their semantic relationship can be displayed. Links can be created by using the hyperlink "Links". By this, a new window will be opened in which the respective sign number may be entered in the field "#"; the new link will be created by clicking on the button "Add".

<sup>2</sup>Figures 5 - 12 are clippings from the full view.

Links are always bi-directional, this means if sign A is linked to sign B, sign B is automatically linked to sign A.

### 3.2.2.3. Sign information

In order to be able to enter something into this area, the analyst has to click on "Edit sign information", thereby opening the respective selection window. Here, the user may enter general information about a sign: type of sign (one handed, symmetric, asymmetric, and combined), person (name of the signer) and region in which the sign is used. Where necessary, a comment can be added.

The names of signers which have already been entered into the database can be called up as selection list. If the user wants to add a new name, it can be typed in into the field "Signer". The new name is automatically added to the selection list.

Since it is possible that a sign is used in several regions, more than one region may be selected. In case a region is not yet in the list, it may be entered like a new signer name.

### 3.2.2.4. Semantic information

In the area "Semantics", the meaning of a sign is entered. If a direct translation into the respective spoken language is impossible, the meaning can be described in the field "Paraphrase". In addition, information about "Connotative meaning", "Etymology" and "Language change" may be given. For the sign example, input fields are available for the glosses and the translation (cf. Figure 7).

Label	Input Field
Semantics	<input type="text"/>
Paraphrase	<input type="text"/>
Connotative meaning	<input type="text"/>
Etymology	<input type="text"/>
Language change	<input type="text"/>
Example 1	<input type="text"/>
Translation 1	<input type="text"/>

Figure 7: Possible entries in "Semantics"

Also in this area, not all information has to be entered at once. It is useful to enter first the respective meanings of the sign into the field "Semantics". This is important because the search function of the database is designed in a way that the entries in this field are searched for matches with the search word. All other related entries may be filled in later at any time.

### 3.2.2.5. Pragmatic information

The input field for pragmatic entries is similarly designed as the one for semantic information. It only differs in the labels for the entries. Data can be entered for "Usage", "Collocation", "Phrase" and "Idiom". Equally to

"Semantics", there are input fields for the glosses and translation of the sign example.

### 3.2.2.6. Morphosyntactic information

The morphosyntactic categories are displayed by means of a tree structure (cf. Figure 8). In this way, the syntactic categories can be clearly presented and the structure facilitates the selection by the user. Since signs can often be assigned to more than one word class (e.g. noun and verb), it is necessary that this can also be recorded in the analysis. For this reason, the user has the possibility to select several syntactic categories and transfer them to the general presentation of the sign.

Category	Selection
Noun	<input type="checkbox"/>
Verb	<input type="checkbox"/>
Adjective	<input type="checkbox"/>
Adverb	<input type="checkbox"/>
Pronoun	<input type="checkbox"/>
Preposition	<input type="checkbox"/>
Conjunction	<input type="checkbox"/>
Numeral	<input type="checkbox"/>
Interjection	<input type="checkbox"/>
Particle	<input type="checkbox"/>

Figure 8: Selection options in "Morphosyntax"

### 3.2.2.7. Semantic field

A sign may be assigned one or more "semantic field(s)". At the moment, a user can choose among more than 100 semantic fields (Figure 9 shows a part of these semantic fields). The selection of a semantic field is done by clicking on the box in front of it.

Semantic field	Selection	Semantic field	Selection	Semantic field	Selection
<input type="checkbox"/> Abuse	<input type="checkbox"/>	<input type="checkbox"/> Accommodation	<input type="checkbox"/>	<input type="checkbox"/> Activity/Occupation	<input type="checkbox"/>
<input type="checkbox"/> Advertising	<input type="checkbox"/>	<input type="checkbox"/> Affection/Love	<input type="checkbox"/>	<input type="checkbox"/> Animals/Insects	<input type="checkbox"/>
<input type="checkbox"/> Art	<input type="checkbox"/>	<input type="checkbox"/> Associations/Federations	<input type="checkbox"/>	<input type="checkbox"/> Attributes/Features	<input type="checkbox"/>
<input type="checkbox"/> Bar	<input type="checkbox"/>	<input type="checkbox"/> Behaviour	<input type="checkbox"/>	<input type="checkbox"/> Birds	<input type="checkbox"/>
<input type="checkbox"/> Carnival	<input type="checkbox"/>	<input type="checkbox"/> Cities/Places	<input type="checkbox"/>	<input type="checkbox"/> Clothing/Fashion	<input type="checkbox"/>
<input type="checkbox"/> Computer	<input type="checkbox"/>	<input type="checkbox"/> Construction equipment/utensils/material	<input type="checkbox"/>	<input type="checkbox"/> Cooking	<input type="checkbox"/>
<input type="checkbox"/> Customs/Tradition	<input type="checkbox"/>	<input type="checkbox"/> Day time/Time	<input type="checkbox"/>	<input type="checkbox"/> Dialect	<input type="checkbox"/>

Figure 9: Clipping of selection window "Semantic field"

### 3.2.2.8. Non manual components

In this window, there are input fields for the categories "Facial expression", "Mouthing", "Mouth gesture", "Head", "Shoulder" and "Upper part of the body" (cf. Figure 10).

Figure 10: Selection window "Non manual components"

Values which are in the selection list can be transferred to the input field by clicking on them; new values may simply be entered in the input fields. It is possible to analyse only one (e.g. Facial expression) or more categories (e.g. Facial expression and Mouthing).

### 3.2.2.9. Manual components: Analysis of sign phases

Though the discussion on order/hierarchy of sign parameters is still ongoing (cf. Dotter 2000; 2007), we decided to use a phenomenologically arguable structuring of signs into phases, following the principal observations by Liddell & Johnson (1989). By this solution, the big number of parameters can be distributed into two sets which follow from the observation of the signs rather naturally. This practice is a "phonetic" one which guarantees that the analyses can be read by all linguists independent of their different "phonemic" orientation without obeying any theoretical model. In order to analyse a sign by phases, the user has to open the editing window by clicking on the link "Add phase 1" (cf. Figure 11). Now the user has to decide whether the sign starts with a hold or a movement phase. The list of categories and values is connected to the selection of the phase. In other words, if the phase "Hold" is selected, the selection list of the input field "Category" contains all categories assigned to this phase: Hand shape, Place of articulation, Palm orientation, Direction of the knuckles, Wrist, Point of contact and Type of contact. If now the category "Hand shape" is selected, the selection list of the input field "Value" shows the respective values. In this case, these are the descriptions of more than 150 hand shapes currently available in the database. In order to facilitate the selection of a hand shape, the user may call up the graphic presentations of these hand shapes (by clicking on the green button with the white arrow). Here, the chosen hand shape can be directly selected. In some cases, a "Specification" of values is needed. For example, a place of articulation may be specified by "close to".

Figure 11: Analysis of the individual phases

When the analysis of the category "Hand shape" is finished, the next category (e.g. Place of articulation) may be added by clicking on the link "add new category". The analysis procedure is the same as described for the category "Hand shape". When all desired categories of the phase "Hold" are analysed, the data are transferred to the general presentation of the sign by clicking on "OK". If the hold phase is followed by a movement phase, the user will click on the link "add phase 2" to enter again the analysis area of the individual phases and will choose "Movement" as type of phase. The selections of the categories and the values as well as the transfer of the data to the general presentation of the sign take place in the same way as in the hold phase. The categories and values for the non dominant hand may be entered in the same manner. Figure 12 shows the analysis of the ÖGS-sign "der Abend", as it appears in the general presentation.

Figure 12: General presentation of the ÖGS-sign "der Abend" ["evening"]

In some cases, it may be necessary to add a comment on the analysis of a sign. For this purpose, the field "Comment" may be used.

## 3.2.2.10. Statistics

The last area in the general presentation of a sign contains statistical data, which are automatically generated by the system. They show the assignment of the (analysed) sign to a project, the name of the person who has entered the sign and the name of the person who has last modified the sign (cf. Figure 5). These data cannot be viewed by anonymous users, only by registered ones.

## 3.2.3. Current status

So far, the database has only been used for the description of ÖGS. At the moment, LedaSila contains about 14.000 sign (videos) of Austrian Sign Language. These are regional sign variants as well as 3.400 standardised ÖGS-signs.

The priority objective of ZGH was to enter available sign videos into the database. This explains why only about 400 signs of the complete corpus have been analysed so far. For the near future, it is planned to accelerate the analysis of signs, which will mainly be carried through by the deaf collaborators of ZGH.

## 3.2.4. Invitation to the Sign Language Linguistics Community

We specifically want to invite deaf researchers and, of course, also all other parties interested in sign language to use the free database for documentation and/or analysis of their sign language(s). A first impression of the database can be gained by visiting <http://ledasila.uni-klu.ac.at/>. If you are interested in further information, please contact Klaudia Krammer from ZGH:

[klaudia.krammer@uni-klu.ac.at](mailto:klaudia.krammer@uni-klu.ac.at).

Researchers and other parties who accept this offer and enter sign language data into the database agree to share their data with the Sign Language Linguistics Community. This in turn contributes to improving the international research networking.

## 4. References

- Blees, M. et al. (1996). *SignPhon. A database tool for phonological analysis of sign languages*. Ms. Leiden.
- Brien, D., Brennan, M. (1995). Sign language dictionaries: Issues and developments. In H. Bos & T. Schermer (Eds.), *Sign Language Research 1994. Proceedings of the 4<sup>th</sup> European Congress on Sign Language Research. München, September 1-3, 1994*. Hamburg: Signum, pp. 313--338.
- Dotter, F. (2000). On the adequacy of phonological solutions for sign languages. In A. Bittner, D. Bittner & K.-M. Köpcke (Eds.), *Angemessene Strukturen: Systemorganisation in Phonologie, Morphologie und Syntax*. Hildesheim/Zürich/New York: Olms, pp.47--62.
- Dotter, F. (2007). Methodenprobleme der Gebärdensprachforschung. In *Das Zeichen 21*, pp. 462--479.
- Krammer, K. et al. (2001). The Klagenfurt database for sign language lexicons. In *Sign Language & Linguistics 4*, (1/2), pp. 191--201.
- Liddell, S.K., Johnson, R., E. (1989). American Sign Language: The phonological base. In *Sign Language Studies 64*, pp. 195--277.
- Prillwitz, S. et al. (1989). *HamNoSys, Version, 2.0. Hamburger Notationssystem für Gebärdensprache: Eine Einführung [Hamburg National System of Sign Language: An Introduction]*. Hamburg: Signum.
- Wolf, B. (1992). Wörterbuch und Benutzer - Versuch einer empirischen Untersuchung [Dictionary and user - Attempt at an empirical investigation]. In U. Braue & D. Viehweger (Eds.), *Lexikontheorie und Wörterbuch. Wege der Verbindung von lexikologischer Forschung und lexikographischer Praxis [Lexical theory and dictionary: Ways to connect lexicological research and lexicographic application]*. (Lexicographia 44). Tübingen: Max Niemeyer Verlag, pp. 295--376.

# Digital Deployment of the Signs of Ireland Corpus in Elearning

Lorraine Leeson, Brian Nolan

School of Linguistic, Speech and Communication Sciences, Trinity College Dublin  
Centre for Deaf Studies, 40 Lwr. Drumcondra Road, Drumcondra, Dublin 9, Ireland.

Department of Informatics, Institute of Technology, Blanchardstown, Dublin, Ireland.

E-mail: leesonl@tcd.ie brian.nolan@itb.ie

## Abstract

The Signs of Ireland corpus is part of the School of Linguistic, Speech and Communication Sciences' "Languages of Ireland" project. The first of its kind in Ireland, it comprises 40 male and female signers from across the Republic of Ireland, aged 18-65+, all of whom were educated in a school for the Deaf. The object was to create a snapshot of how ISL is used by 'real' signers across geographic, gendered and generational boundaries, all of which have been indicated as sociolinguistically relevant for ISL (cf. the work of Le Master; also see Leeson and Grehan 2004, Leonard 2005, Leeson et al. 2006). With the aim of maximising the potential of cross-linguistic comparability, we mirrored aspects of data collection on other corpora collected to date. Thus, we include the Volterra et al. picture elicitation task (1984), "The Frog Story", and also asked informants to tell a self-selected story from their own life. To date, all of the self-selected and a quarter of the Frog Story data have been fully annotated using ELAN.

Two institutions (TCD and ITB) have partnered to create a unique elearning environment based on MOODLE as the learning management system, funded under the Irish government's Strategic Innovation Fund, Cycle II. This partnership delivers third level signed language programmes to a student constituency in a way that resolves problems of time, geography and access, maximizing multi-functional uses of the corpus across undergraduate programmes. Students can take courseware synchronously and asynchronously. We have now built a considerable digital asset and plan to re-architect our framework to avail of current best practice in digital repositories and digital learning objects vis-à-vis Irish Sign Language.

This paper outlines the establishment and annotation of the corpus, and the success of the corpus to date in supporting curricula and research. This paper focuses on moving the corpus forward as an asset to develop digital teaching objects, and outlines the challenges inherent in this process, along with our plans and our progress to date in meeting these objectives.

Specific issues include:

- Decisions regarding annotation
- Establishing mark-up standards
- Use of the Signs of Ireland corpus in elearning/ blended learning contexts
- Leveraging a corpus within digital learning objects
- Architecture of a digital repository to support sign language learning
- Tagging of learning objects versus language objects
- Issues of assessment in an elearning context

## 1. Background

This paper outlines the establishment and annotation of the Signs of Ireland corpus, currently the largest digital annotated corpus in Europe insofar as we are aware, and the success of the corpus to date in supporting curricula and research. This paper focuses on moving the corpus forward as an asset to develop digital teaching objects. This paper outlines the challenges inherent in this process, and outlines our plans and our progress to date in meeting these objectives.

### 1.1 A Note on Irish Sign Language

Irish Sign Language is an indigenous language of Ireland. It is used by some 5,000 Irish Deaf people as their preferred language (Matthews 1996) while it is estimated that some 50,000 non-Deaf people also know and use the language to a greater or lesser extent (Leeson 2001). The Signs of Ireland corpus is part of the Languages of Ireland programme at the

School of Linguistic, Speech and Communication Sciences, TCD. It comprises data from

Deaf Irish Sign Language (ISL) users across Ireland in digital form, and has been annotated using ELAN, a software programme developed by the Max Planck Institute, Nijmegen. The corpus is housed at the Centre for Deaf Studies, a constituent member of the School.

While technology has opened the way for the development of digital corpora for signed languages, we need to bear in mind that signed languages are articulated in three dimensional space, using not only the hands and arms, but also the head, shoulders, torso, eyes, eye-brows, nose, mouth and chin to express meaning (e.g. Klima and Bellugi 1979 for American Sign Language (ASL); Kyle and Woll

1985, and Sutton-Spence and Woll 1999 for British Sign Language (BSL); and McDonnell 1996; Leeson 1996, 1997, 2001; O’Baill and Matthews 2000 for Irish Sign Language (ISL)) leads to highly complex, multi-linear, potentially dependent tiers that need to be coded and time-aligned.

As with spoken languages, the influence of gesture on signed languages has begun to be explored (Armstrong, Stokoe and Wilcox 1995, Stokoe 2001; Vermeerbergen and Demey (2007)), while discussion about what is linguistic and what is extra-linguistic in the grammars of various signed languages continues (e.g. Engberg-Pedersen 1993, Liddell 2003, Schembri 2003). While these remain theoretical notions at a certain level, decisions regarding how one views such elements and their role as linguistic or extra-linguistic constituents plays an important role when determining what will be included or excluded in an annotated corpus. Such decisions also determine how items are notated, particularly in the absence of a written form for the language being described.

## 2. ELAN

Originally developed for research on gesture, ELAN has become the standard tool for establishing and maintaining signed language corpora. ELAN (EUDICO Linguistic Annotator) is an annotation tool that allows one to create, edit, visualize and search annotations for video and audio data. ELAN was developed with the aim of providing a sound technological basis for the annotation and exploitation of multi-media recordings.. (Source: *ECHO Project* - <http://www.let.ru.nl/sign-lang/echo/index.html?http&&www.let.ru.nl/sign-lang/echo/data.html>)

## 3. The Corpus

The corpus currently consists of data from 40 signers aged between 18 and 65 from 5 locations across the Republic of Ireland. It includes male and female signers, all of whom had been educated in a school for the Deaf in Dublin (St. Mary’s School for Deaf Girls or St. Joseph’s School for Deaf Boys). None were sign language teachers, as we wished to avoid the collection of data from signers who had a highly conceptualized notion of ‘correct’ or ‘pure’ ISL. All use ISL as their preferred language. While some of the signers are native signers insofar as they come from Deaf families, the majority are not. Several have Deaf siblings. All signers included use ISL as their first or preferred language, and all acquired it before the age of 6 years. The distribution of locations from where data was collected can be seen in Figure 1 below and a breakdown of the gender and age of participants is outlined in Figure 2.

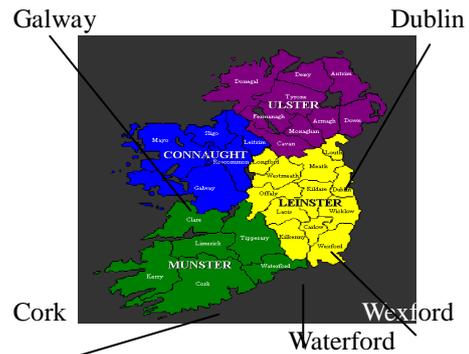


Figure 1: Sites for Corpus Collection (2004)

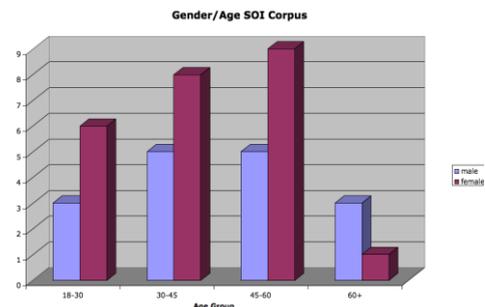


Figure 2: Preliminary gender breakdown within Corpus Collection (2004)

Data was collected by a female Deaf research assistant, Deirdre Byrne-Dunne. This allowed for consistency in terms of data elicitation. It also meant that, due to the demographics of the Irish Deaf Community, Ms. Byrne was a known entity to all of the participants, which is evident in some placed in terms of interaction on-screen between informants and data collector, allowing for some interesting sociolinguistic insights. The fact that Ms. Byrne-Dunne is herself Deaf, and an established member of the Irish Deaf community, meant that the potential for ‘Observer’s Paradox’ (Labov 1969) while not reduced, took on a positive spin: knowing who the interviewer/ recorder of data was, and knowing their status as a community member, lent itself to the informants opening up and using their ‘natural’ signs rather than a variety that they might have assumed a university researcher would ‘expect’ or ‘prefer’.

It also meant that the informants who knew Deirdre, either as a former class-mate or from within the Deaf community, code-switched to use lexical items that would not typically be chosen if the interlocutor was unknown. For example, some ‘school’ signs were used (BROWN). And in other instances, informants, telling stories that they had self-selected, referred to Deirdre during the recounting of their tales.

We have touched on the fact that data collected included self-selected narratives. We also asked participants to tell ‘The Frog’ story, which is a

picture sequence format telling the story of a young boy who, with his dog, searches for his frog, which has escaped from a jar. Informants were also asked to sign the content of the Volterra picture elicitation task, a series of 18 sets of paired pictures showing a series of situations that aim to elicit transitive utterances. Both the ‘frog’ story and the Volterra picture elicitation task have been used widely in signed language specific descriptions and in cross-linguistic comparisons, including ISL (e.g. Leeson 2001 for ISL; Johnston, Vermeerbergen Schembri and Leeson (2007) for Australian Sign Language, Flemish Sign Language and ISL; Volterra et al. 1984 for Italian Sign Language; Coerts 1994 for Sign Language of the Netherlands).

Funding permitting, we would like to expand the data on file to include renditions of Chafe’s Pear Story and Aesop’s fables, dialogues, and interviews with Deaf ISL users regarding how they view ISL in order to record the current status and usage of ISL. We would ideally also like to supplement this with register specific data, such as descriptions of occupational activities to elicit the range of register specific vocabulary available within the community at present. Additional gaps that need to be addressed include dialogues and ethnographic data, the inclusion of child language data and elderly signers. Further, there are a number of locations that we would also like to see represented as they represent particular sociolinguistic situations (e.g. the language situation in Northern Ireland, the Mid-West).

For example, the Mid-West School for the Deaf was established some 20 years ago, with the result that many children from the region were educated locally. This brought an end to the tradition for all Deaf children in Ireland to attend the Catholic Schools for the Deaf in Dublin. This shift in educational provision has also allowed for a ‘regional variant’ to have emerged, brought about by the relative isolation of signers in the Mid-West during their formative schooling years (Conama 2008). To explore this further, we are currently collecting data in the Mid-West region (Limerick, Tipperary, Clare).

#### 4. Annotating the Corpus

One of the myths of annotating data is that the annotators are neutral with respect to the data and that they simply ‘write down what they see’. But it is just that – a myth. As ISL does not have a written form, there is no standard code for recording it. While some established transcription keys exist (HamNoSys, Sign Writing, Stokoe Notation), none of these are compatible with ELAN and none are fully developed with respect to ISL.

Another issue is that these transcription systems are not shared ‘languages’ – that is, in the international sign linguistic communities, these transcription codes are not common place, and to use one in place

of a gloss means limiting the sharing of data to an extremely small group of linguists. However, glossing data with English ‘tags’ is problematic too. Pizzutto and Pietrandrea (2001) point out the dangers inherent in assuming that a gloss can stand in for an original piece of signed language data. They note that “It is often implicitly or explicitly assumed that the use of glosses in research on signed [languages] is more or less comparable to the use of glosses in research on spoken languages ... this assumption does not take into account, in our view, that there is a crucial difference in the way glosses are used in spoken as compared to signed language description. In descriptions of spoken (or also written) languages, glosses typically fulfill an ancillary role and necessarily require an independent written representation of the sound sequence being glossed. In contrast, in description of signed languages, glosses are the primary and only means of representing in writing the sequence of articulatory movements being glossed” (2001: 37). Later, they add that “... glosses impose upon the data a wealth of unwarranted and highly variable lexical and grammatical information (depending upon the spoken/written language used for glossing).” (ibid: 42).

Thus, the glossing of signed data is fraught with potential problems – even when a team is working very consistently and cross-referencing work in a diligent manner, as is the case here. The Signs of Ireland project appears to be unique in that *all* annotated data was verified by a Deaf research assistant who holds a masters degree in applied linguistics. All three annotators held masters degree qualifications in linguistics/ communications as well as Deaf Studies specific qualifications, making them uniquely qualified to work with this data.

While one of the most positive features of ELAN is the fact that the stream of signed language data runs in a time-aligned fashion with the annotations, the problem remains that any search function is restrained by the consistency and accuracy of the annotations that have been inputted and second-checked by the Signs of Ireland team.

For example, several ISL signs may be informally glossed in the same way, but the signs themselves are different, for example, HEARING [1] as used by older signers (“L” handshape at chin) and HEARING [2] (“x” handshape at chin) as used by younger signers. The fact that both of these signs are glossed in the same way demonstrates that any frequency count that would subsequently be carried out using ELAN would not distinguish between the two on the basis of the gloss, HEARING, alone. But the inclusion of both variants glossed in the same way does allow students to search for all possible variants of the signs and find relevant sociolinguistic information as to who typically uses the sign (gender, age, region) and whether it is a borrowed sign or seems idiosyncratic in some way.



HEARING [1]



HEARING [2]

This issue of tagging items according to grammatical function is yet another issue that poses challenges. We have not yet tagged data in this way because we do not yet know enough about the grammatical function of items in ISL to accurately code to that level. Despite this, our annotations do reflect assumptions about the nature and structure of certain items. We also take very seriously the concerns of linguists who have discussed the impact of early codification of signed languages like Flemish Sign Language (VGT) (Van Herreweghe and Vermeerbergen 2004).

Despite the fact that we wanted to avoid making assumptions about word class and morpho-syntax, the act of annotating a text means that certain decisions have to be made about how to treat specific items. For example, it is known that non-manual signals, articulated on the face of the signer, provides information that assists in parsing a message as for example, a question or a statement, or in providing adverbial like information about a verbal predicate (e.g. Leeson 1997; O’Baioill and Matthews 2000 for ISL, Sutton-Spence and Woll 1999, Brennan 1992, Deuchar 1984 for British Sign Language; Liddell 1980 for American Sign Language). When it comes to annotating such features, we had to decide if we would treat non-manual features as dependent tiers, relative to the manual signs that they co-occur with, or as independent tiers containing information that may be supra-segmental in nature. We decided to treat all levels as independent of each other until we

could ascertain a relationship that held consistently across levels.

At the lexical level, there were decisions to be made as to what constitutes a word in ISL. While established lexical items that have citation forms in dictionaries or glossaries of ISL were ‘easy’ to decide on, there was the issue of how to determine if a sign was a ‘word’ or a ‘gesture’ or part of a more complex predicate form, often described as classifier predicates. The fact that some signers used signs related to their gender or age group challenged the annotators – they had to determine whether a sign that was new to them was a gendered variant (Le Master 1990, 1999-2000, Leeson and Grehan 2004), a gendered generational variant (Le Master *ibid*, Leonard 2005), a mis-articulation of an established sign (i.e. a ‘slip of the hand’ (Klima and Bellugi 1979), an idiosyncratic sign, a borrowing from another signed language (e.g. BSL), or a gesture. Our team’s experience and qualifications helped the decision making process here. All decisions were recorded in order to provide a stable reference point for further items that challenged that shared characteristics with items that were discussed previously.

The use of mouth patterns in signed languages provide another challenge for annotators dealing with signed languages. Mouthings and mouth gestures have been recognized as significant in signed languages, and while mouthings are often indicative of the language contact that exists between spoken and signed languages, mouth gestures are not (for example, see Boyes Braem and Sutton-Spence 2001, Sutton-Spence 2007).

Given that the Signs of Ireland corpus will, in the first instance, be used by researchers looking at the morpho-syntax of the language, we opted to not annotate the mouth in a very detailed manner. Instead, we have provided fairly general annotations following from those listed in the ECHO project annotations list.

## 5. Use of the Signs of Ireland corpus in elearning/ blended learning contexts

The Signs of Ireland corpus has been piloted in elearning and blended learning at the Centre for Deaf Studies in the academic years 2006-7 and 2007-8 across a range of courses, but specifically, Irish Sign Language courses, an introductory course focusing on the linguistics and sociolinguistics of Irish Sign Language, and a final year course that focuses on aspects of translation theory and interpreting research. At present the corpus exists on each client-side computer. Students are provided with training in how to use ELAN in order to maximize use of the corpus. The implications of this are that, currently, students must be able to access the corpus in a lab. This presents a challenge for blended

learning delivery where students require internet access to the corpus. This also creates challenges in terms of data protection legislation, distribution, copyright and general access issues which need to be resolved as we move forward. For example, subsets of the data are already used as digital learning objects, but no decision has yet been made regarding optimal management and deployment of the corpus.

Examples of how we have used the corpus include the following:

We have developed assessments to Council of Europe Common European Framework of Reference level B1 (productive/ expressive skill) and B2 (receptive/ comprehension skill) level for ISL. This includes a receptive skills test which includes multiple choice questions linked to data taken from the Signs of Ireland corpus. The corpus data sits amid other test items, which are outlined in Table (1) below:

Test Item	Domain	Duration	Test Format
Multiple Statements	Life Experience	1 1/2 minutes video (10 minutes)	Visual images (10 items)
The Deaf Summer Camp (SOI)	Life Experience Travel Deaf Current Affairs	1 minute video (10 minutes total)	MCQ Paraphrase True/False Qs Pen & paper (10 items)
“My Goals”	Ambitions / Professional Focus	1 minute video (10 minutes total)	MCQ Paraphrase True/False Qs Pen & paper (10 items)

**Table 1: Sample ISL Receptive Test Using Digital Objects**

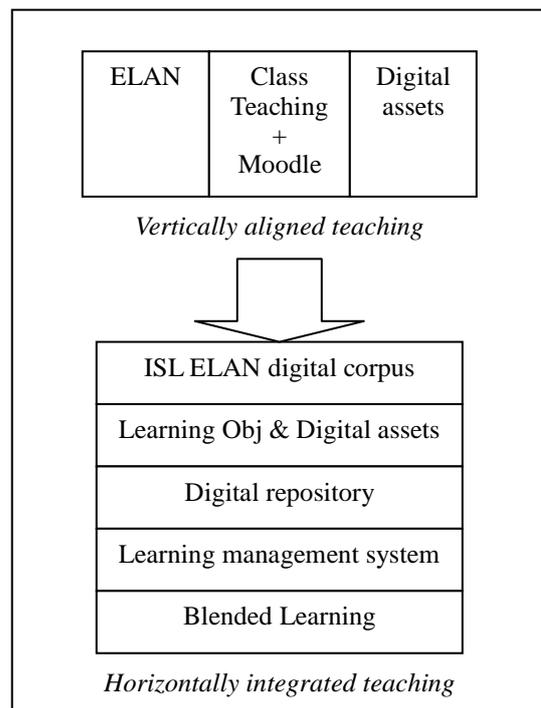
We also use the corpus as part of the continuous assessment of students in our Introduction to the Linguistics and Sociolinguistics of Signed Languages course. For example, students are required to engage with the corpus to identify frequency patterns, distribution of specific grammatical or sociolinguistic features (e.g. lexical variation) and to draw on the corpus in preparing end of year essays.

In the Translation and Interpreting: Philosophy and Practice course, students engage with the corpus to explore issues of collocational norms for ISL, look at the distribution of discourse features and features such as metaphor and idiomatic expression.

## 6. Leveraging a Corpus and Digital Learning Objects

To optimally leverage the Signs of Ireland corpus within a learning environment, we will, in the initial phase of the proposed educational value chain begin by determining what are the actual functional requirements with respect to how the application will be used by both students and academics in the blended learning context.

At the moment we have Moodle populated with a wide variety of modules delivered within the suite of CDS undergraduate programmes. The Signs of Ireland digital corpus is tagged in ELAN. We have traditional classroom and blended delivery of content.



**Figure 3: The integrated model**

The present programme architecture is very vertical in orientation (Figure 3). The challenge is to achieve horizontal integration through the use of information technology, the Internet and a blended learning approach.

## 7. Architecture of a Digital Repository to Support Signed Language Learning

Planning is also required with respect to the overall architecture and framework. We are in the process of determining what profiling and other user related information we require to capture and tag data regarding the user environment and their interaction with the digital classroom and curriculum.

Additionally, we have started the analysis that will indicate what types of learning objects we need for each of the programme modules for each lecture, and

how many and of what type with the intention of making our blended learning Diplomas and Degrees available online from September 2009. We make the initial base assumption that the target client devices are browsers on internet aware laptops and desktops. This assumption can be expected to evolve, over time, into mobile devices such as the Apple iPhone, iPod Touch and similar computing appliances. This will deliver to us a plan for the capture and creation of the respective digital rich media that we intend to deploy within our learning objects.

We are designing and architecting our learning environment to situate the learning objects in a digital repository in such a way to easily facilitate their use in conjunction with a learning management system. The repository will be expected to link the learning objects to the learning management system in a horizontal integrated manner across the appropriate technology hardware and software platforms. We plan to facilitate for searching for learning objects by keyword through standards based tagging. For the associated technology platforms, we are investigating some open source software options, for example, FEDORA [FEDORA-a, FEDORA-b] the Flexible Extensible Digital Object Repository Architecture. We will also investigate the possible use and advantages that an XML ontology may deliver to the project, including the Protégé tools from Stanford University, which are also open source for educational use [Protégé]. Protégé can work with XML, RDF and has some smart visualisation tools built in. We are not certain yet as to the role the ontology might play.

## **8. Tagging of Learning Objects Versus Language Objects**

Even today in the sector, it is an open question as to what are the current best practices in meta-tagging for learning objects. Notwithstanding this, we are of the opinion that the SCORM v2 standard will be applied [SCORM]. We will link the SCORM

standard in a way that is functional and optimal for our project.

As we create our rich media digital assets and leverage the ISL ELAN digital corpus, we are paying particular attention to the tagging of the digital assets to include, for example, some or all of the following, with some private and public user views according to access profiles (Figure 4).

These initial tag labels can be expected to mature and be fine-tuned following the completion of our programme learning outcome and learning object analysis.

## **9. Issues of Assessment in an elearning Context**

We are also working on developing an assessment model, based on best pedagogical practice, that is appropriate to our online blended learning environment. From there, we will then as part of our design phase, determine how to implement this online. We will need to link, in a principled and structured way, the assessments to the learning outcomes of individual modules, for example, An Introduction to the Linguistics and Sociolinguistics of Signed Languages, and to a particular lecture's thematic learning outcomes as appropriate. We also consider the effectiveness of the assessment with students in a blended learning situation.

## **10. Moving Forward**

Our Strategic Innovation Fund (SIF II) Deaf Studies project is scoped for a three-year window commencing in 2008-9. A challenging year one plan has been created that will yield infrastructure changes, achievements and digital assets as well as the approval of a four year degree in Deaf Studies; ISL Teaching, and ISL/English Interpreting.

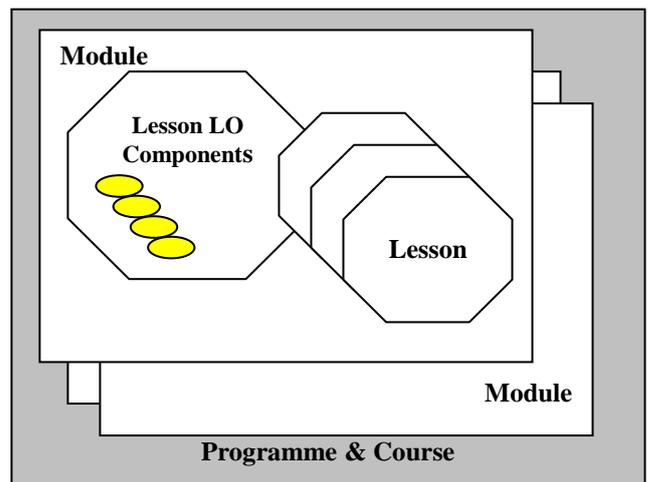
1. Topic
2. Description
3. Sections
4. Media
  - a. Source
  - b. Options for reuse
  - c. Context - 'where used now'
  - d. Proof of availability
  - e. Ownership
    - I. Licensing
    - II. Cost
    - III. Payment Method
  - f. Optimum speed of access and use
  - g. Ability to apply style guide
  - h. Types supported
6. Handle tags: Specific topics covered
7. Context
  - a. Modality for delivery
  - b. Format
10. Conversion speed
11. Assessment of topics
  - a. Assessment of specific areas
  - b. Depth of assessment
  - c. Level of adaptability
  - d. Feedback
16. Author
17. Version number
18. Date Created

**Figure 4: Potential tags of interest**

We are presently engaged in an analysis phase to identify for each of the curriculum modules in year one of the Diploma programmes offered by the Centre, the learning objectives of a particular lecture and its themes on a week-by-week basis. For example, week 1, lecture 1 has learning objectives LO1, LO2 and LO3, etc. Typically, this will broadly equate with a lecture plan that is rolled out over a semester. For example, the module 'An Introduction to the Linguistics and Sociolinguistics of Signed Languages' is delivered over two semesters totaling 24 weeks with 24 2-hour lectures over the academic year. We will need to make explicit the learning objectives of each of these lectures such that each objective may be supported by up to, say, four learning objects initially (Figure 5).

These learning objects are expected to form a composite unit, but will be made up of different

media types. A composite unit, therefore, will be expected to include the lecture notes (.pdf or .ppt), Moodle quizzes and exercises, video data of signing interactions (in Macromedia Breeze, Apple QuickTime and/or other formats), and ELAN digital corpora. To make a composite unit, each learning object needs to be wrapped with proper tagging. This tagging will facilitate searches for these learning objects within a digital repository. We plan that this will be done for all modules across all weeks.



**Figure 5: Learning object components as a unit within a module**

We will identify and implement appropriate assessment models for a blended learning delivery of Sign Language programmes. In addition to an assessment model, we will need to devise a model for determining the overall effectiveness of the programme within the blended learning approach that will take a more holistic and pedagogical perspective to the programme objectives. We intend to deploy this programme nationally across the regions of Ireland following initial Dublin based trials. When this national deployment occurs these effectiveness key performance indicators will assume a greater importance that will enable us to determine the answer to the question: Are we successful with this programme and how can we tell?

Following an initial trial period in the Dublin area and once we have gathered a sufficiency of initial data, we will compare and contrast the assessments with anonymous (but marked for age and social background, gender, hearing status, etc.) and start to compare longitudinal figures with the initial first year outputs for this blended programme.

As this programme is to be modeled for a blended learning environment, we will need to build in a model of student support to include in an appropriate way, online college tutors, peer-learning and mentoring, in order to address any retention issues that may arise and provide the students with the ingredients of their learning success within a

productive and engaging community of practice.

We intend to create a website for this SIF II Deaf Studies Project with links to the learning management system/Moodle, other technology platforms including, for example, Macromedia Breeze, and the rich digital media assets as we determine to be useful in support of the teaching of Irish Sign Language within 3<sup>rd</sup> level education. We will also use this website to disseminate programmatic and research outcomes and other relevant information. We will address the technology related issues pertinent to the design and implementation of the framework for digital learning objects in a repository to facilitate access-retrieval, update, and search. We will determine the tagging standards that will operate across this.

While we will deploy the blended learning approach initially to the Dublin area, we will start planning for the national deployment. We will therefore pilot data in the Centre for Deaf Studies in Dublin from October 2008 as supplementary to traditional modes. We will capture feedback from students and analyse this critically. Following this, we will rollout in selected region/s across the country via local 3<sup>rd</sup> level institutes of higher education in 2009-10. We have agreements with many of these secured at this time.

In terms of the human resources required to build the framework and create the digital assets for the full programme, and the appropriate skill-levels required, we will shortly be seeking to recruit a number of individuals with postgraduate qualifications with a specific research focus. These individuals will be required to determine the appropriate assessment models and how this can be implemented for elearning, backed up by a digital repository of learning objects that leverage the Signs of Ireland digital corpus.

We will also be recruiting a co-coordinating project manager with a relevant post-graduate qualification with people-influencing skills who is bilingual in ISL/English and has good organizational and financial management skills who can leverage key community insights with empathy and diplomacy. We will recruit academic staff for local delivery of ISL in the regions, interpreting lecturer/s and also general Deaf Studies academic/s. We will recruit an elearning/ digital repository/ digital media specialist as well as ISL/English interpreters. We will recruit administrative support to the project.

To contribute to the research of the programme, we intend to recruit at Ph.D level to investigate the following research areas: 1) Assessment models appropriate to ISL in an elearning and blended learning context; 2) Developing and maturing the Signs of Ireland corpus, including meta-tagging and enriching the data; 3) Signed language/spoken language interpreting; 4) Design and build of rich digital media for Irish Sign Language

There are considerations regarding the cultural and work practice implications for academic staff delivering curricula in this manner. There are also corresponding implications for students receiving education in a blended learning approach via elearning technology. What will assume a greater importance immediately for academics and students is the minimum level of computer literacy skills and access to modern computing equipment and a fast broadband network required to engage in this kind of learning environment. We also plan, therefore, to devise a training programme for academic staff to induct them into the new teaching and learning environment and plan for a similar induction for students enrolled on the programme.

## 11. Summary

In this paper we have discussed decisions we have made regarding annotation of the Signs of Ireland corpus. We discussed ongoing work regarding mark-up standards and their application as we move forward. We outlined the range of applications currently made with respect to the Signs of Ireland corpus in elearning/ blended learning contexts. We indicated how we will leverage the corpus within a framework for digital learning objects situated in an architecture with a digital repository to support signed language learning. We outlined issues relating to the tagging of learning objects for deployment in a digital repository versus the tagging in ELAN of language objects for grammatical, morpho-syntactic and sociolinguistic phenomena. We noted that there will be challenges to representing these with a common notation that is digitally accessible. Issues of assessment in an elearning context were also addressed.

## References

- Arms, W. Y. (1997). *An architecture for information in digital libraries*. D-Lib Magazine, February 1997.  
<http://www.dlib.org/dlib/february97/cnri/02arms1.html>. Accessed March 2008.
- [Digital Librarian]:  
<http://www.dlib.org/dlib/november96/ucb/11hastings.html>. Accessed March 2008.
- Armstrong, D.F., W.C. Stokoe and S.E. Wilcox (1995).  
*Gesture and the Nature of Language*, Cambridge University Press, Cambridge.
- Boyes Braem, P. and R. Sutton-Spence (eds.) (2001).  
*The Hands are the Head of the Mouth: The Mouth as Articulator in Sign Languages*. International Studies on Sign Language and Communication of

- the Deaf, 39, Signum, Hamburg.
- Chaudhry, Abdus Sattar and Christopher S.G. Khoo. (2006). *Issues in developing a repository of learning objects for LIS education in Asia*. [http://www.ifla.org/IV/ifla72/papers/107-Chaudhry\\_Khoo-en.pdf](http://www.ifla.org/IV/ifla72/papers/107-Chaudhry_Khoo-en.pdf). Accessed March 2008.
- Coerts, J. (1994). Constituent order in Sign Language of the Netherlands, In M. Brennan and G.H. Turner (eds.) *Word-order Issues in Sign Languages – Working Papers*, The International Sign Linguistics Association (ISLA), Durham, pp 47-71. [FEDORA-a]: <http://www.fedora.info/>. Accessed March 2008. [FEDORA-b]: <http://www.fedora-commons.org/>. Accessed March 2008.
- Emmorey, K. (ed.) (2003). *Perspectives on Classifier Constructions in Sign Languages*, Lawrence Erlbaum and Associates, New Jersey.
- Engberg-Pedersen, E. (1993), *Space in Danish Sign Language*, Hamburg, Signum Verlag.
- Janzen, T. (2005). *Perspective Shift Reflected in the Signer's Use of Space*, Centre for Deaf Studies, University of Dublin, Trinity College, Dublin.
- Hastings, Kirk and R. Tennant. (1996). *How to build a digital librarian*. D-Lib Magazine. November 1996. <http://www.dlib.org/dlib/november96/ucb/11hastings.html>. Accessed March 2008.
- Johnston, T., M. Vermeerbergen, A. Schembri and L. Leeson (2007). 'Real data are messy': Considering cross-linguistic analysis of constituent ordering in Australian Sign Language (Auslan), Vlaamse Gebarentaal (VGT), and Irish Sign Language (ISL). In: Perniss, P., Pfau, R., Steinbach, M. (Eds.) *Visible Variation: Comparative Studies on Sign Language Structure*. Mouton de Gruyter, Berlin. (41 blz.).
- Klima, E. and U. Bellugi (1979). *The Signs of Language*, Harvard University Press, Harvard.
- Kyle, J.G. and B. Woll (1985). *Sign Language: The Study of Deaf People and their Language*, Cambridge University Press, Cambridge.
- Himmelman, N.P. (2006). Language Documentation: What is it and what is it good for?, In Gippert, J., N.P. Himmelman and U. Mosel (eds.), *Essentials of Language Documentation*, Mouton de Gruyter, Berlin and New York, pp 1-30.
- Himmelman, N.P. (2006). The challenges of segmenting spoken language In Gippert, J., N.P. Himmelman and U. Mosel (eds.), *Essentials of Language Documentation*, Mouton de Gruyter, Berlin and New York, pp 253-274.
- Labov, W. (1969). *The Logic of Nonstandard English*, Georgetown University 20<sup>th</sup> Annual Round Table, Monograph Series on Languages and Linguistics, No. 22.
- Leeson, Lorraine, John Saeed, Cormac Leonard, Alison Macduff and Deirdre Byrne-Dunne (2006). *Moving Heads and Moving Hands: Developing a Digital Corpus of Irish Sign Language: The 'Signs of Ireland' Corpus Development Project*. Paper presented at the IT&T conference, Carlow.
- Leeson, L. and C. Grehan (2004). *To the Lexicon and Beyond: The Effect of Gender on Variation in Irish Sign Language*. In M. Van Herreweghe and M. Vermeerbergen (eds.): *To The Lexicon and Beyond: The Sociolinguistics of European Sign Languages*. Gallaudet University Press, pp. 39-73.
- Leeson, L. (1996) *The Marking of Time in Signed Languages with Specific Reference to Irish Sign Language*, Unpublished M. Phil Dissertation, Centre for Language and Communication Studies, University of Dublin, Trinity College.
- Leeson, L. (1997). *The ABC of ISL*. Booklet to accompany the TV series. RTE/Irish Deaf Society, Dublin.
- Leeson, L. (2001). *Aspects of Verb Valency in Irish Sign Language*, Unpublished Doctoral Dissertation, Centre for Language and Communication Studies, TCD, Dublin.
- Leeson, L. and J.I. Saeed (2003). *Exploring the Cognitive Underpinning in the Construal of Passive Events in Irish Sign Language (ISL)*. Paper presented at the 8<sup>th</sup> International Cognitive Linguistics Association Conference, Spain, July 2003.
- Leeson, L. and J.I. Saeed (2004). The Windowing of Attention in Simultaneous Constructions in Irish Sign Language. In T. Cameron, C. Shank and K. Holley (eds.) *Proceedings of the Fifth High Desert Linguistics Conference*, 1-2 November 2002, Albuquerque, University of New Mexico.
- Leeson, L. and J.I. Saeed (2005). *Conceptual Blending and the Windowing of Attention in Simultaneous Constructions in Irish Sign Language (ISL)*, Paper Presented at the 8<sup>th</sup> International Cognitive Linguistics Association Conference, Seoul, Korea, July 2005.
- Leeson, L. and J.I. Saeed (2007). *Conceptual Blending and the Windowing of Attention in Simultaneous Constructions in Irish Sign Language (ISL)*, In Vermeerbergen, M., L.

- Leeson and O. Crasborn (eds), *Simultaneous Constructions in Signed Languages – Form and Function*, John Benjamins, Amsterdam.
- Le Master, B. (1990). *The Maintenance and Loss of Female and Male Signs in the Dublin Deaf Community*, PhD. Dissertation, Los Angeles, University of California.
- Le Master, B. (1999-2000) Reappropriation of Gendered Irish Sign Language in One Family, *Visual Anthropology Review*, 15(2), pp. 1-15.
- Le Master, B. (2002). What Difference Does Difference Make? Negotiating Gender and Generation in Irish Sign Language. In S. Benor, M. Rose, D. Sharma and Q. Shang (eds), *Gendered Practices in Language*, Centre for the Study of Languages and Information Publication, Stanford.
- Leonard, C. (2005). *Signs of Diversity: Use and Recognition of Gendered Signs among Young Irish Deaf People*. Deaf Worlds, Vol. 21 (2), pp. 62-77.
- Liddell, S.K. (2003) *Grammar, Gesture, and Meaning in American Sign Language*, Cambridge University Press, Cambridge.
- Matthews, P.A. (1996). *The Irish Deaf Community-Volume 1*, ITE, Dublin.
- McDonnell, P. (1996). *Verb Categories in Irish Sign Language*, Unpublished Doctoral Dissertation, Centre for Language and Communication Studies, University of Dublin, Trinity College.
- McDonnell, P (ed.) (2004). *Deaf Studies in Ireland: An Introduction*, Doug McLean, Glouc.
- McGreal, Rory (ed.) (2004). *Online education using learning objects*. Open and flexible learning series. London and New York: RoutledgeFalmer.
- [Protégé]: <http://protege.stanford.edu/>. Accessed March 2008.
- Meir, I. (1998). *Thematic Structure and Verb Agreement in Israeli Sign Language*, PhD Dissertation, Hebrew University of Jerusalem.
- Mosel, U. (2006). Fieldwork and Community Language Work. In Gippert, J., N.P. Himmelmann and U. Mosel (eds.), *Essentials of Language Documentation*, Mouton de Gruyter, Berlin and New York, pp 67-86.
- O’Baill, D. and P.A. Matthews (2000). *The Irish Deaf Community, Volume 2 – The Linguistics of ISL*, ITE, Dublin.
- Pizzutto, E. and P. Pietrandrea (forthcoming). The Notation of Signed Texts, *Sign Language and Linguistics*, Vol 4, No. 1/2, pp29-45, 2001.
- Roberts, G., W. A., Martin Feijen, Jen Harvey, Stuart Lee, and Vincent P. Wade. 2005: *Reflective learning, future thinking: digital repositories, e-portfolios, informal learning and ubiquitous computing*. ALT/SURF/ILTA Spring Conference Research Seminar. [http://www.alt.ac.uk/docs/ALT\\_SURF\\_ILTA\\_white\\_paper\\_2005.pdf](http://www.alt.ac.uk/docs/ALT_SURF_ILTA_white_paper_2005.pdf). Accessed March 2008.
- Sallandre, M., *Simultaneity in French Sign Language* Discourse, In Vermeerbergen, M., L. Leeson and O. Crasborn (eds), *Simultaneous Constructions in Signed Languages – Form and Function*, John Benjamins, Amsterdam, (forthcoming).
- Sandler, W. and D. Lillo-Martin, *Sign Language and Linguistic Universals*, Cambridge University Press, Cambridge, 2006.
- Schembri, A., Rethinking ‘Classifiers’ in Signed Languages, In K. Emmorey (ed), *Perspectives on Classifier Constructions in Sign Languages*, Lawrence Erlbaum and Associates, New Jersey, 2003, pp. 3-34.
- [SCORM]: <http://adlcommunity.net/mod/resource/view.php?id=458>. Accessed March 2008.
- Seifart, F. Orthography Development, In Gippert, J., N.P. Himmelmann and U. Mosel (eds.), *Essentials of Language Documentation*, Mouton de Gruyter, Berlin and New York, 2006, pp 275-299.
- Stokoe, W. C., *Language in Hand: Why Sign Came Before Speech*, Gallaudet University Press, Washington DC, 2001.
- Sutton-Spence, R. Mouthings and Simultaneity in British Sign Language, In Vermeerbergen, M., L. Leeson and O. Crasborn (eds), *Simultaneous Constructions in Signed Languages – Form and Function*, John Benjamins, Amsterdam, (forthcoming).
- Sutton-Spence, R. and B. Woll, *The Linguistics of British Sign Language*, Cambridge University Press, Cambridge, 1999.
- Talmy, L., The Windowing of Attention in Language. In M. Shibatani and S.A.Thompson, (eds.): *Grammatical Constructions- Their Form and Meaning*. Oxford: Clarendon Press, 1997, pp235-287.
- Tsunoda, T., *Language Endangerment and Language Revitalisation – An Introduction*, Berlin and New York, Mouton de Gruyter, 2005.
- Wilbur, R., Eyeblinks and ASL phrase structure, *Sign Language Studies* 84: 221-240,1994
- Van Herreweghe and Vermeerbergen 2004, Flemish Sign Language: Some Risks of Codification, In Van Herreweghe, M. and M. Vermeerbergen, *To the Lexicon and Beyond: The Sociolinguistics of European Sign Languages*, Gallaudet University Press, Washington DC, pp. 111-140, 2004.

- Van Valin, R. D. and R.J. La Polla, *Syntax-Structure, Meaning and Function*, Cambridge University Press, Cambridge, 1997.
- Vermeerbergen, M., L. Leeson and O. Crasborn, Simultaneous Structures in Signed Languages: An Introduction, In Vermeerbergen, M., L. Leeson and O. Crasborn (eds), *Simultaneous Constructions in Signed Languages – Form and Function*, John Benjamins, Amsterdam,
- Vermeerbergen, M. and E. Demey (2007). Sign + Gesture = Speech + Gesture? Comparing Aspects of Simultaneity in Flemish Sign Language to Instances of Concurrent Speech and Gesture. In Vermeerbergen, M., L. Leeson and O. Crasborn (eds.), *Simultaneous Constructions in Signed Languages – Form and Function*, John Benjamins, Amsterdam.
- Volterra, V., S. Laudanna, E. Corazza and F. Natale (1984). *Italian Sign Language: The Order of Elements in the Declarative Sentence*. In F. Lonke (ed.) *Recent Research on European Sign Languages*. Svets and Zeitlinger: Lisse, pp.19-48.
- Wilbur, R.B. (1994). Eyeblinks and ASL Phrase Structure, In *Sign Language Studies* 84, pp 221-240.
- Wilcox, S. (2004a). Gesture and Language: Cross-linguistic and Historical Data from Signed Languages, *Gesture* 43, pp 43-73.
- Wilcox, S. (2004b). Cognitive Iconicity: Conceptual Spaces, Meaning and Gesture in Signed Languages, *Cognitive Linguistics* 15-2, pp 119-147.

#### **Acknowledgements:**

We are most grateful for funding that has been made available under the Strategic Innovation Fund, Cycle II programme for this for the Deaf Studies SIF II Project. Funding for the Signs of Ireland corpus came from the Arts and Social Sciences Benefaction Fund, TCD.

## Toward an computer-aided sign segmentation

François Lefebvre-Albaret, Frederick Gianni, Patrice Dalle

IRIT (UPS - CNRS UMR 5505)

Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 9

{lefebvre,gianni,dalle}@irit.fr

### Abstract

The presented article explains an innovating method to process a computer-aided segmentation of sign language sentences. After having tracked the signers hands in a video, the traitment consists in detecting motion attributes such as repetition or symmetries. Those observations are taken into account to process a gesture segmentation. We also discuss about the evaluation of such a segmentation.

### 1. Introduction

Processing French Sign Language (FSL) videos requires a first segmentation step. Nowadays, this tedious task is processed manually and the result is dramatically influenced by the the human operator. We have focused on the segmentation problem to find unified segmentation criteria and to accelerate the segmentation step. The applications of such a research are far beyond the linguistic task of defining where a sign begins or ends, it could also be applied to automatic sign language video processing or used to produce sign language sentences with signing avatars.

### 2. Goal of the paper

We first relate some significant studies concerning sign language video processing and present some commonly used algorithms in sign language video processing. Then, we explain the method we have developed to segment a video into signs. This algorithm is based on dynamic programming and on a one-segment definition of a sign. It processes hand motion, and we will soon include other informations as facial expression, elbow position or hand configuration. After having detailed our evaluation method, we will discuss about the accuracy of our segmentation results, and the way we could improve it.

### 3. Previous studies of sign language video processing

Nowadays, most teams focus on the sign recognition problem. The recognition process sometimes includes a segmentation step (Kim et al., 2001), but the segmentation results are not evaluated. However, those recognition methods are based on several approaches that could also be used for sign segmentation.

The sign recognition methods can be classified into several categories according to the model of sign they refer to. We will distinguish approaches using one-segment or multi-segment sign modelling and other hidden model based algorithms.

In one-segment approach, each gesture is modeled as one single segment. This description refers to Stokoe's sign definition and models a sign as a combination of simultaneous features (hand motion, position, configuration and

orientation) (W. C. Stokoe and Croneberg, 1978). The reasons for such a one-segment model of a sign have been exposed in (Channon, 2002). This one-segment model approach has been used in (Derpanis et al., 2004) in order to characterize isolated gestures. Each gesture is qualified by its motion pattern, its hand configuration, and its location. The purpose of their algorithm is to recognise the primitive combination. There are 14 different movements, 3 body locations and 4 hand shapes. Each primitive can be identified with one or more operators processing the whole video sequence of one elementary gesture. According to the operator results, it is possible to determine which combination of primitives has been used to create the gesture. 148 Movements were correctly classified with a success rate of 86 %.

This method has only been applied to gesture classification and was not employed to process real signs ; but this kind of approach could also be useful in sign language processing.

An other approach would be to consider signs as a succession of segments according to the model proposed by Lidell and Johnson (Liddell and Johnson, 1990). This approach has been successfully employed by (Kim et al., 2001) for sign recognition. The algorithm models a sign as a succession of 5 states (resting state, preparation state, stroke/moving state, ending/repetition state and end state) and uses a Markov Model to segment the signs. After this step, a Hidden Markov Model (HMM) leads to sign recognition. The algorithm is able to recognise signs with an accuracy of 95 %. Unfortunately, there are very few information about the evaluation protocol.

In fact, HMM are commonly used for sign recognition because this model is particularly adapted for temporal signal processing. In such a method, each sign is modelled as a succession of states that are automatically determined during the training phase. Processing signs with HMM does not need any *a priori* explicit sign model, but needs a long training phase to find the optimal states. This method is sometime adapted to process the different sign parameters (hand motion, configuration) separately (Vogler, 2003) or to speed up the recognition phase (Wang et al., 2001)<sup>1</sup>.

<sup>1</sup>These studies make use of a cyberGlove to capture motion

In the field of sign language video processing, several studies also used HMM based methods. Among them (Bauer and Hienz, 2000) uses HMM to process isolated sign recognition. The signer wears coloured gloves, both hands are then easier to track in the video. The recognition rate is 92 % on a 97 sign corpus.

A very interesting study has been realised by (Bowden et al., 2004). As in the previous studies, hands are tracked to find out in each frame of the video sequence their position and their global shape among 5 configurations. The use of a combination of HMM and Principal Component Analysis enable a recognition rate of about 98% over a 43 sign corpus. Those results are very encouraging but the recognition process only works on isolated signs.

We have presented a few studies related to our present research. Some other methods have been achieved to process isolated and continuous sign recognition. Those approaches are listed in (Ong and Ranganath, 2005).

#### 4. Which approach for a sign language segmentation?

Those ways of sign language processing have been successfully used to perform isolated or continuous sign recognition. Even if a segmentation is made at the same time of the sign recognition, the segmentation accuracy is not evaluated.

The HMM based method requires a long training phase to be able to recognize each sign. Such an approach could not be applied for a continuous natural sign language segmentation (we can notice that all the video used previously for the studies only contained highly constrained sentences with a small set of vocabulary). We have noticed that a lot of signs commonly used in FSL are highly iconic. Moreover, standard signs can be transformed according to the context to express spatial relationships (as it is the case for directional verbs). For those two reasons, learning all the possible signs seems unconceivable with such a method.

The solution would be to detect the sign components independently. This is what has been done by (Vogler, 2003). We have chosen to use this approach to build our segmentation method and to base our algorithm on a one-segment definition of a sign.

#### 5. Algorithm presentation

Our segmentation process is composed of four steps :  
 Firstly, the hands and head position are tracked in a video.  
 Secondly , a human operator picks out one frame (that will be called *seed*) of each sign included in the video sequence.  
 Thirdly, according to the *seeds* and the hands trajectories, the algorithm performs the segmentation.  
 Fourthly, the sign language expert can check the segmentation result and make the necessary corrections.  
 Each of those steps depicted in [Figure 1] will be explained in the following sections.

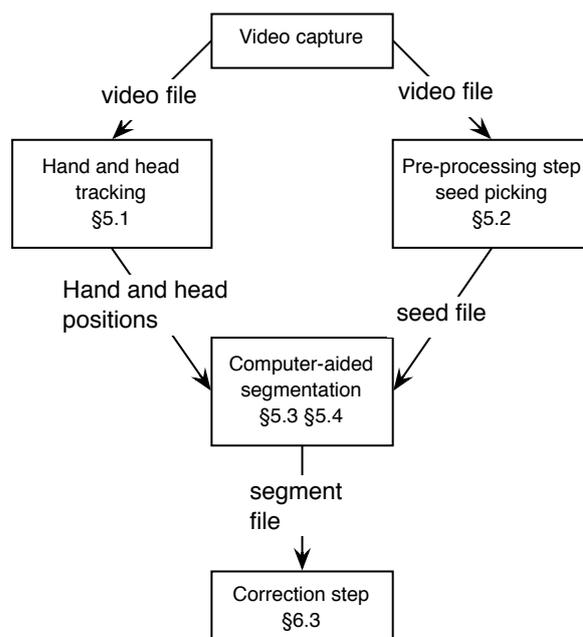


Figure 1: Description a computer-aided segmentation

##### 5.1. Body parts tracking

The first step of the segmentation process consists in tracking the head and the two hands in the video. During a FSL utterance, hand motions can be very fast and brutal direction changes come along. One of the major problem is to design methods that handle those kinds of movements.

In the presented approach, we use the skin color to detect the head and hands, and statistics estimators (via particles filters) for the correspondence. Since particles filter models the uncertainty, it provides a robust framework to track the moving hands of a person telling a story in FSL.

##### 5.1.1. Algorithm description

We used the annealed filtering method presented in (Gall et al., 2006) and applied it in a skin color context, in order to be robust against non-rigid motion and free orientation changes of the observed objects. The observation density is modeled by the skin color distribution of pixels using non-parametric modelling. This model is sometimes referred to as construction of skin probability Map (Brand and Mason, 2000) (Gomez and Morales, 2002).

##### 5.1.2. Results

We have evaluated the tracking method on a FSL video sequence. This sequence is around 3000 frames long (at 25 frames per second) and images have a size of 720 x 576 pixels. The results are presented in the [Figure 2].

##### 5.2. Pre-processing step

By now, a fully automated segmentation, using only motion processing with an unconstrained sign language, gives as a results a 25 % correct segmentation. A short intervention of a human operator in the segmentation process can increase this correct segmentation rate by manually selecting one frame (and only one) of each sign

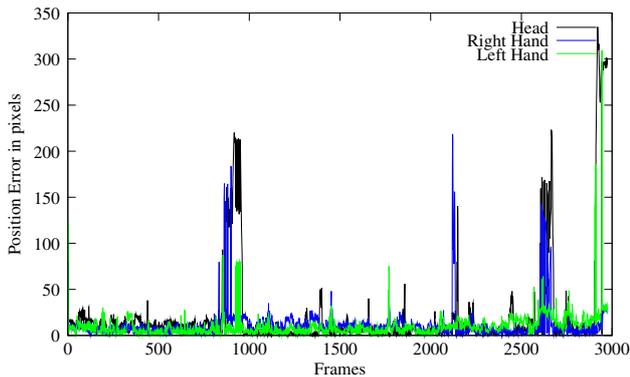


Figure 2: Evaluation the particles filter applied to the FSL video from LsColin.

included in the video sequence. During this step, the video can be displayed with a normal speed or slowed down according to the operator preference. The selected frame can be anywhere in the sign temporal segment.

Each time he recognizes a sign, the operator simply presses a key of his keyboard. The result of this manual pre-processing step is a list of *seed* frames, which is represented as a track at the bottom of the visualization screen [Figure 3]. Naturally, it is possible to make some corrections and to move back if a mistake has been made while pointing the *seeds*.



Figure 3: Tool for the pre-segmentation step

### 5.3. Segment characterisation

Our algorithm is given this list of *seeds* and has to find for each one the beginning and the end of the corresponding sign. The next step consists in processing the hand motion in order to find the temporal segment of each sign that have been marked during the pre-processing step. Some operators characterising a specific feature (motion symmetry, repetition, straight movement) will be applied to each temporal segment. Thanks to those features detections, it will be possible to assign two confidence measures to each temporal segment indicating whether it could be a sign or a transition between two signs. The following part

explains how those operators are applied to calculate those confidence measures.

The algorithm processes the 2D hand motion in the video. It means that depth information will not be taken into account. As a consequence, some different motion patterns will be recognised as the same movement. For instance, it will be the case for a horizontal circle and a back and forth horizontal movement.

In the following explanation, a temporal segment between the frame  $i$  and the frame  $j$  will be noted  $S_{ij}$ . The 2D speed of right and left hand on the frame  $f$  will respectively be written  $\overline{Vr}(f)$  and  $\overline{Vl}(f)$ . The horizontal and vertical components of the right speed will respectively be written  $Vr_h(f)$  and  $Vr_v(f)$ .

Each temporal segment  $S_{ij}$  of less than 50 frames is analysed to find out movement features. We use 9 different kinds of operators divided into four categories:

#### Relational operators :

They detect a specific relationship between the motion of left and right hands during the sign processing:

- Central symmetry :  $\overline{Vr}(f) \approx -\overline{Vl}(f)$ .
- Sagittal symmetry :  
 $Vr_h(f) \approx -Vl_h(f)$  and  $Vr_v(f) \approx Vl_v(f)$
- Translation :  $\overline{Vr}(f) \approx \overline{Vl}(f)$
- Static hand (only the case of a static left hand will be illustrated) : during the temporal segment  $\frac{\|\overline{Vl}(f)\|}{\|\overline{Vr}(f)\|} \ll 1$

#### Structural operators :

- Double repetition (the global movement can be depicted as a juxtaposition of two identical movements)
- Triple repetition

#### Motion operators :

- Straight motion (constant speed)
- Straight motion (accelerated and decelerated)

#### Economy operators :

- A last operator is applied to evaluate the economy of the sign. Assuming the hand to be a punctual mass ( $m = 1$ ). The temporal segment is characterized by the hand speed ( $\overline{V}_b$  and  $\overline{V}_e$ ), and the hand position ( $P_b$  and  $P_e$ ) at the beginning and at the end of the segment and the duration  $d$  of the segment. The minimal energy to realize the transition between the states  $(V_b, P_b)$  and the states  $(V_e, P_e)$  is written  $E_m$ . This energy is calculated thanks to a potential energy difference. The real energy is written  $E_r$ .

$$E_r = \sum_{f=i}^{j-1} (|V_h(f) \cdot (V_h(f) - V_h(f+1))| + |V_v(f) \cdot (V_v(f) - V_v(f+1))|)$$

The economy of a movement is computed by the formula  $E_m/E_r$ . This operator is only used to find out the transition segments. If the result is near 0, it means that the movement is not economical and probably contains a sign.

Not all of the movements have been taken into account. For instance, we do not detect yet curved trajectories or back and forth motion. Those detectors of movement primitives will be added later if they increase the segmentation accuracy.

The results of those  $k$  operators are written  $o_{ij}^k$  ( $o_{ij}^k \in ]0, 1]$ ). An aggregation process, based on weighting rules is applied to assign two confidence measures to each temporal interval. The first one,  $C_{ij}^s$ , indicates whether the segment could be a sign and the second one,  $C_{ij}^t$ , indicates whether it could be a transition. The aggregation process will be improved and studied in detail in the further version of our algorithm.

#### 5.4. Dynamic programming

After having given those two confidence measures to each temporal segment, we want to find the best video segmentation. All sign sentences are modelled by a succession of signs and transitions. The structure of a sign sentence is then represented as the following finite state machine [Figure 4].

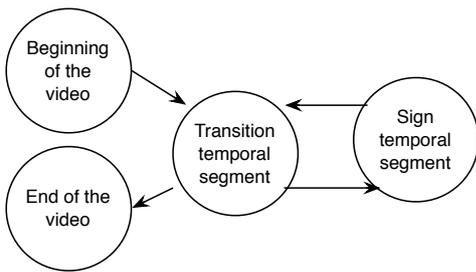


Figure 4: Sentence model

We can draw a parallel with a conventional two states Markov Chain. Each state does not represent a time point but is actually a temporal segment (a succession of frames). We have adapted Viterbi algorithm, which is currently used to solve this kind of problem to take the temporal segment length into account.

The function that should be optimized to find the segmentation is then :

$$\operatorname{argmax} \left( \sum_n \ln(C_{ij}(n)) \cdot l(n) \right)$$

were  $C^{ij}(n)$  and  $l(n)$  represent the confidence measure and the length of the  $n^{\text{th}}$  chosen temporal segment. Other constraints are also added to use the *seeds* :

- All signs must contain exactly one *seed*
- A transition have not to contain any *seed*

The result of the segmentation is a succession of segments, which can be visualized with the annotation software Ancolin (Braffort et al., ). This software allows us to compare the computer-aided segmentation with a manual segmentation [Figure 5].

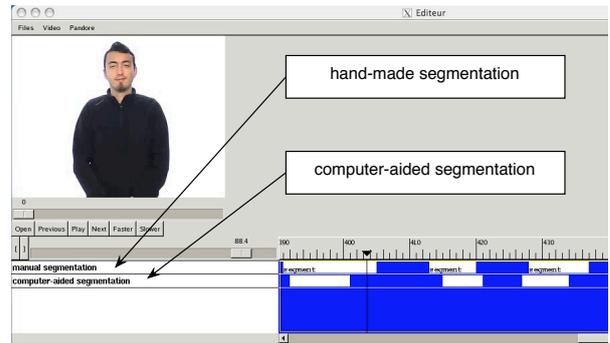


Figure 5: Comparison of manual and computer-aided segmentation

## 6. Result evaluation

### 6.1. Evaluation criteria

How can we evaluate a segmentation? The answer mainly depends of the usage of the obtained temporal segments. Our mid-term goal is to animate a signing avatar. As a consequence, the segmented signs are intended to be used in other sentences. If they do not contain the whole definition of the sign, a truncated sign will always be displayed. At the opposite, if a sign contains a part of the transition from the previous or to the following sign, this parasitic motion or configuration will degrade the quality of the sign language synthesis.

In general, partially segmented signs present one of the following features :

- Some contacts or a speed reversal points of the original sign are not included in the segment.
- The number of repetitions or back and forth cycles is not correct .
- The segment contains a configuration of the preceding or following sign.
- The segment contains a movement of transition from the preceding or to the following sign.

## 6.2. Result analysis

We have used two sequences to test our algorithm. It is very important to notice that those two sequences have been realised by native signers and that the signs are processed at a normal speed. At the contrary of a lot of other studies, the vocabulary used in the video is totally unconstrained.

The first one is a 2 minutes video sequence where the signer explains the 11 September 2001 events from LS Colin corpus (Cuxac, 2002). This sequence is very interesting because a lot of signs are highly iconic. The story is composed of 203 signs. Our success rate is 40% in this video. The second video<sup>2</sup> has been provided by the Websourd society and is a translation of a piece of news in FSL. This video of 30 seconds contains 43 signs. Among them :

- 22 are correctly segmented (54%)
- 16 are partially segmented (39%)
- 3 are not detected at all (7%)

It clearly appears that the segmentation accuracy depends of the type of the processed video. The following hypothesis could account for this accuracy difference between the two computer-aided segmentations.

**Video length :** For small videos, it is possible to watch the video once. The seeds picking is then easier to realise because one can remember the sign included in the video. It was not possible for the 11 September video because the video was too long. As a consequence, a lot of seeds have been misplaced during the pre-processing step.

**Prepared sentences :** The 11 September story was totally spontaneous and had not been prepared. So, there was a lot of hesitation that we have considered as if it were signs. Unfortunately, a lot of them were very short and did not fit our automatic segmentation criteria.

**Iconic signs :** A lot of iconic signs have been used in the 11 September video. For some of them, it was very hard to distinguish elementary sub-signs, which could be automatically segmented. For this reason, we had some granularity problems. Some signs, that were considered as one segment during the hand-made reference segmentation, were assigned two seeds during the pre-processing step (seed-picking) because they seemed to be compound of two elementary signs.

Coming back to our goal of signing avatar animation, the errors made for hesitation and highly context dependant signs are not so penalising because those signs would be hard to reuse to build other sentences. However, such signs have to be taken into account in a computer-aided segmentation tool to be able to process real sign language.

The partially segmented signs of the second video (news translation) have been classified into several categories :

- 1 segment had too little frame. The sign was not recognizable.
- 6 segmented signs were truncated (one contact point, repetition or back and forth motion have been deleted).
- 5 segments contained a configuration of the previous or following sign.
- 4 segments contained a movement of the previous or following sign (or a movement of the transition to those signs).

Those results could be improved by taking other parameters into account in the segmentation process. According to our observations, the hand configuration exploitation would be a good way of improving the segmentation accuracy.

## 6.3. Correction step

Even if the above results are very encouraging, automatic segmentation must be checked and corrected by a human operator. During this step, all the segment must be visualised and corrected if needed.

It is very important to avoid deletion errors, because the presence of ignored signs would oblige the sign language expert to watch all the video to find out those signs.

It would also be interesting to use the confidence measure  $C_{ij}^s$  during the checking phase. By using this value, the sign language expert would be able to focus on the signs, which have the smallest confidence value (and then the highest probability to be bad segmented).

The goal of computer-aided segmentation is to accelerate the manual segmentation phase. It means that the pre-processing step added to the checking step must spend less time than a fully manual segmentation. We have measured those 3 informations for the second video (of 30 second) :

- Seed picking :  $T_s = 3$  minutes
- Checking/correction :  $T_c = 7$  minutes
- Manual segmentation :  $T_m = 13$  minutes

$(T_s + T_c)/T_m = 0.77$ . In our case 23 % time is spared using this semi-automated segmentation process. And we could improve this value by increasing the segmentation accuracy of our algorithm.

## 7. Conclusion

Regarding to the few parameters that have been taken into account to process the segmentation our results are very encouraging and validate our segmentation approach.

However, it is important to improve the segmentation accuracy to make our method be usable in corpora processing. Such an improvement could only be made in taking into account other sign parameters:

<sup>2</sup> This video can be downloaded at the following address <http://websourd.nnx.com/mediav0/IMG/flv/1D001-97.flv>

- A lot of sign ends can be characterized by a configuration change. Using configuration could probably allow us to perform a better segmentation.
- The simultaneous tracking of hand, elbow and shoulder positions could lead to a reconstruction of the whole arm posture as demonstrated in (Lenseigne et al., 2004) and allow us to process 3D motion. The hand motion could be depicted more accurately.
- The analysis of head orientation and facial expression could also decrease the number of wrong segmentation (Parashar, 2003).

## 8. Acknowledgments

This study has been financed by the Websourd society and the Midi-Pyrenees region. It has been done in the frame of the SESCO project (System for Sign Language Pedagogy and Communication with Avatars) of the group PRESTO (Tolosan Sign Reserch Group). Thanks to J. Dalle who has been our LSF expert and performed all the manual segmentation that were refered to in this article.

## 9. References

- Britta Bauer and Hermann Hienz. 2000. Relevant features for video-based continuous sign language recognition. In *FG '00: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, page 440, Washington, DC, USA. IEEE Computer Society.
- Richard Bowden, David Windridge, Timor Kadir, Andrew Zisserman, and Michael Brady. 2004. A linguistic feature vector for the visual interpretation of sign language. In Tomás Pajdla and Jiri Matas, editors, *ECCV (1)*, volume 3021 of *Lecture Notes in Computer Science*, pages 390–401. Springer.
- A. Braffort, A. Choisier, C. Collet, P. Dalle, F. Gianni, B. Lenseigne, and J. Segouat. Toward an annotation software for video of Sign Language, including image processing tools & signing space modelling .
- Jason Brand and John S. Mason. 2000. A comparative assessment of three approaches to pixel-level human skin-detection. In *15th International Conference on Pattern Recognition*, volume 1, pages 1056–1059.
- R.E. Channon. 2002. *Signs are single segments: phonological representations and temporal sequencing in ASL and other sign languages*. Ph.D. thesis, University of Maryland, USA.
- Cuxac. 2002. "langage et cognition" , rapport de fin de recherche. <http://www.irit.fr/LS-COLIN>.
- Konstantinos G. Derpanis, Richard P. Wildes, and John K. Tsotsos. 2004. Hand gesture recognition within a linguistics-based framework. In *Computer Vision - ECCV 2004 8th European Conference on Computer Vision*. Springer.
- J. Deutscher, A. Blake, and Reid I. 2000. Articulated body motion capture by annealed particle filtering. *Computer Vision and Pattern Recognition*.
- J. Gall, J. Potthoff, C. Schnoerr, B. Rosenhahn, and H.P. Seidel. 2006. Interacting and annealing particle filters: Mathematics and a recipe for applications. Technical Report MPI-I-2006-4-009, Max-Planck Institute for Computer Science.
- G. Gomez and E. Morales. 2002. Automatic feature reconstruction and a simple rule induction algorithm for skin detection. In *ICML Workshop On Machine Learning in Computer Vision*, pages 31–38.
- J.-B. Kim, K.-H. Park, W.-C. Bang, J.-S. Kim, and Z. Bien. 2001. Continuous korean sign language recognition using automata based gesture segmentation and hidden markov model. In *ICCAS2001*, Jeju, Korea. Jeju National University.
- Boris Lenseigne, Frédéric Gianni, and Patrice Dalle. 2004. Estimation mono-vue de la posture du bras en utilisant un modèle biomécanique, méthode et évaluation. In *14ème Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle - RFIA 2004, Toulouse, 28/01/04-30/01/04*, volume 2, pages 957–966. AFRIF-AFIA, janvier.
- S. K. Liddell and R. E. Johnson. 1990. American sign language: the phonological base. *Sign Language Studies*, 64.
- Sylvie C.W. Ong and Surendra Ranganath. 2005. Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6):873–891.
- Ayush S. Parashar. 2003. *Representation and interpretation of manual and non-manual information for automated American Sign Language recognition*. Ph.D. thesis, University of South Florida, USA.
- C. P. Vogler. 2003. *American sign language recognition: reducing the complexity of the task with phoneme-based modeling and parallel hidden markov models*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA, USA. Supervisor-Dimitris N. Metaxas.
- D. Casterline W. C. Stokoe and C. Croneberg. 1978. *A dictionary of American Sign Language on Linguistic principles*. Linstok Press.
- C. Wang, W. Gao, and Z. Xuan. 2001. A real-time large vocabulary continuous recognition system for chinese sign language. In *PCM '01: Proceedings of the Second IEEE Pacific Rim Conference on Multimedia*, pages 150–157, London, UK. Springer-Verlag.

## Development of Sign Language Acquisition Corpora

Diane Lillo-Martin<sup>\*</sup>, Deborah Chen Pichler<sup>#</sup>

<sup>\*</sup>University of Connecticut and Haskins Laboratories; <sup>#</sup>Gallaudet University  
Storrs, CT 06269-1145; Washington, DC 20002

E-mail: lillo.martin@uconn.edu, deborah.chen.pichler@gallaudet.edu

### Abstract

Longitudinal, spontaneous production data have long been a cornerstone of language acquisition studies, but building corpora of sign language acquisition data poses considerable challenges. Our experience began with the development of a sign language acquisition corpus more than 15 years ago and has recently included a small-scale experiment in corpus sharing between our two research groups. Our combined database includes regular samples of deaf and hearing children between the ages of 1;06 to 3;06 years acquiring ASL as their native language. The process through which we generate and share transcripts has undergone dramatic changes, always with the triple goal of creating transcripts with sufficient information for the reader to locate regions of interest, while keeping the video fully accessible and minimizing the time required to generate transcripts. In this paper we summarize the various incarnations of our transcription system, from simple Word documents with minimal integration of video, to a combination of FileMaker Pro software integrated with Autolog, to a fully integrated transcript+video package in ELAN. Along the way, we discuss the potential of ELAN to surmount several obstacles that have traditionally stood in the way of large-scale corpus sharing in the sign language acquisition community.

### 1. Longitudinal Spontaneous Production Corpora in Language Acquisition Research

Longitudinal, spontaneous production data have long been a cornerstone of acquisition studies, offering a wealth of information on the processes by which children develop language. Research based on longitudinal spontaneous production data has already led to significant discoveries about the acquisition of a number of languages.

Spontaneous production data provide several advantages to the researcher. (a) A particular child participant is observed in a natural environment, interacting with people she is very familiar with. (b) The child's development over a period of time is carefully documented. (c) Researchers working within a wide variety of theoretical frameworks are able to use such data to address a great range of theoretical issues. For example: (d) The researcher can use the data to address research hypotheses concerning acquisition sequence (the ordering of constructions hypothesized to have particular pre-requisites) and hypotheses concerning simultaneous acquisition (constructions related by an underlying common principle). (e) The researcher can investigate hypotheses about non-target structures used by the child. (f) The input provided to the child can also be sampled and studied, when the child is recorded interacting with a parent. Useful resources on the use of spontaneous production data are found in Stromswold (1996) and Snyder (2007), among others.

The development and distribution of multiple corpora of child language data through the CHILDES project (MacWhinney 2000) has resulted in thousands of publications (see <http://childes.psy.cmu.edu/bibs/>). Corpus sharing through CHILDES allows multiple

researchers to independently examine the same data, making it possible to test each other's analyses for reliability, or propose alternative approaches to interpreting the data. Such activity greatly increases the scientific rigor of the research community.

For the sign language acquisition community, corpus sharing on the scale of CHILDES is still far in the future, due to a number of challenges that we discuss below. However, experiments in smaller-scale corpus sharing can allow us to begin addressing these challenges now. In this paper, we will discuss our experiences in small-scale sharing of sign language longitudinal corpora between our respective research groups at the University of Connecticut and Gallaudet University. We will point out some of the difficulties we have encountered over the course of our collaboration, and the modifications we have adopted in response to them. Although many challenges remain, we are encouraged by the success of our experiment so far and by the enormous potential benefits of corpus sharing for the field of sign language acquisition.

### 2. Challenges of Creating and Sharing Sign Language Acquisition Corpora

Research on sign language acquisition has expanded significantly over the last thirty years, propelled in large part by a growing number of video corpora of signing children. Corpora have now been developed for an ever-increasing number of natural sign languages, creating the potential for fruitful cross-subject and cross-linguistic comparison. Yet much of what we know about sign language development remains limited to a small number of reports on a very small number of children. For example, our collective understanding of early word order acquisition in American Sign Language

(ASL) was for decades determined by a single study based on longitudinal data from three deaf children. Furthermore, details on how data are coded and analyzed are often unavailable, making it impossible for other researchers to test the reliability of analyses presented in the literature. In short, the sign language acquisition community has yet to enjoy the important benefits of corpus sharing that databases such as CHILDES have brought to spoken language acquisition researchers.

There are a number of reasons why sharing of sign acquisition corpora has been slow to catch on. One is that the filming of signing children and their families raises extra concerns about confidentiality. The faces of children and their families must be clearly visible for linguistic analysis to be possible (i.e. their identities can not be concealed by masking or distorting their faces). This increases the already high probability that subjects will be recognized by members of the research community, which draws heavily from comparatively small Deaf communities. Recently, we have noted that video is quickly replacing audio and written records as the standard for acquisition studies, for spoken as well as signed languages, and with this trend has come a general increased tolerance for the inevitable accompanying loss of anonymity. Still, acquisition researchers using video corpora have the responsibility of ensuring as high a degree of confidentiality for their subjects as possible, and this poses a challenge for which we cannot offer any solution at this time.

Instead, we will focus our discussion on two other major obstacles to sharing sign corpora. The first is an absence of standards for annotation or transcription of signed data. Although CHILDES supports the Berkeley Transcription System (Slobin et al. 2001) as a new standard for sign transcription, this system has not been universally adopted, and researchers continue to employ a wide variety of largely idiosyncratic notational conventions. This has made cross-corpus comparisons difficult, if not impossible. Second, our field has until recently lacked a standardized system for efficiently linking annotation or transcription files to large amounts of video data. In the next sections of this paper, we summarize the various ways in which we have addressed these two concerns for our sign acquisition corpus over the past decade.

### 3. Our Corpus

Although we refer to our corpus as a single entity in this paper for ease of exposition, it is actually composed of two distinct sets of naturalistic, longitudinal corpora, one focused on Deaf signers and the other on hearing, bimodal bilingual (coda) signers. The former was collected as part of the *Cross-Linguistic Early Syntax Study* (CLESS) at the University of Connecticut, Department of Linguistics. This project, funded by the U.S. National Institutes of Health (NIH), and the National Science Foundation (NSF), supported data collection of early child English, Spanish, Japanese, and ASL (Lillo-Martin & Snyder

2002). Over the years, data from the acquisition of Russian, Brazilian Portuguese, and Brazilian Sign Language (LSB) have also been included in the project. For this paper, we will focus on the ASL corpus, which includes data from Deaf children (ages 1;6-3;6) acquiring ASL from Deaf, signing parents, and Deaf children of hearing parents (ages 5;9-10;0) whose exposure to ASL began only after the age of five years. Data from hearing, bimodal bilingual children (ages 1;6-4;0) acquiring both ASL from Deaf, signing parents and spoken English are currently being collected as part of a separate project, *Effects of bilingualism on word order and information packaging in ASL*, at Gallaudet University, Department of Linguistics.

As is clear from the names of our projects, our initial focus has been on early syntactic development, beginning at or before the point when children first combine words into 2-word phrases. A great deal of syntactic development occurs within two years from this point, so we extend data collection until the children are about 3-and-a-half to 4-years old. The children are/were filmed regularly (in most cases, on a weekly basis) for about 30-60 minutes at a time.

Information about the age range and amount of data collected for each child in our combined corpora is shown in Table 1. (NB: as coda data collection is still in progress, the information for those children is projected.)

	Child	Age Range	#Sessions	#Hours (approx)
<b>ASL</b>				
<b>D/D</b>	Abby	1;05 – 3;04	79	75
	Jill	1;07 – 3;07	77	79
	Ned	1;05 – 4;02	44	40
	Sal	1;07 – 2;10	18	16
<b>D/H</b>	Cal	6;10 – 10;01	115	50
	Mei	6;07 – 10;0	111	50
<b>H/D</b>	Ben	1;04 – [4;04]	[100]	[80]
	Tom	1;04 – [4;04]	[100]	[80]
	Pete	1;07 – [4;07]	[100]	[80]

Table 1: Data collection – ASL participants.

### 4. Early Transcription System

The very first incarnation of our sign transcripts took the form of Word documents, with a format patterned loosely after the CHAT format used in the CHILDES database.

Each child or adult utterance appeared on its own line, accompanied by information about context and phonological form. Time code was noted every ten lines or so to help users of the transcripts locate regions of interest. However, as may researchers found, entering time code was tedious and actually did very little to facilitate data mining, since video data was stored on analog VHS tapes that had to be manually rewound and fast forwarded to find specific utterances. Furthermore, the video integration of this early system was rather unwieldy, so we quickly sought a way to allow easier and more rapid access to video.

## 5. FMP+Autolog System

The next incarnation of our transcription system featured File Maker Pro (FMP) software integrated with Autolog, a program that allowed us to control a VCR via the computer and link each utterance to its corresponding SMPTE time code on the VHS tape. We designed different interfaces (coding screens) so that different transcribers could focus on the children's sign utterances, adult utterances, non-manual markers (for both child and adult utterances), and non-linguistic context (action and comments). A screen shot of one coding screen, with spaces for all of this information, showing one sample child utterance, is given in Figure 1.

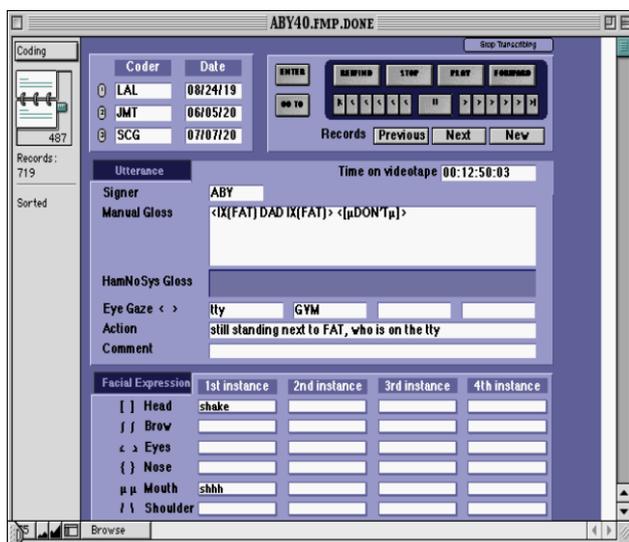


Figure 1: FMP Screen Shot

Drop-down menus and semi-automated time code grab made the job of transcribing easier and greatly decreased the time required to generate transcripts. In addition, FMP included useful features for searching and sorting data,

and could print out sections of transcripts for quick reference. Most importantly, this system dramatically increased the speed and ease with which we could locate video for specific utterances of interest, leading to more accurate data analysis.

Unfortunately, this system required access to Autolog and specially-modified VCRs, and was not widely adopted by linguistics researchers. Furthermore, the VHS tapes on which data was stored and viewed deteriorated quickly from heavy use, making it necessary to dub copies. If subsequent copies started a few seconds later than the original, all our time code stamps on the corresponding transcript would then be off, a small annoyance that eventually caused considerable inconvenience for analysis.

## 6. ELAN System

More recently, we have converted our transcription system to ELAN (<http://www.lat-mpi.eu/tools/elan/>), which enables our transcripts to be time-locked to corresponding digitized video data. This makes the relation between the transcription and the video image much tighter than in either of the previous systems and eliminates the problematic dependence on tape media.

We continue to use traditional upper-case English glosses for transcribing ASL signs, a convenient system, but one with well-known limitations. Nevertheless, we chose this system for its relative readability and ease of use. We keep a running list of glossing conventions, developed and modified through transcriber discussion, that ensures relatively consistent use of glosses across transcribers. However, we recognize the limits of this system, and are willing to accept them only because it is now so easy to consult the video for any given utterance in any transcript in ELAN.

Although ELAN also offers drop-down menus ("controlled vocabularies") and other time-saving features for transcription such as on-the-fly segmentation, generating transcripts is still a time-consuming endeavor, and we have elected to focus our attention on manual activity only, leaving our nonmanual tiers blank for the time being. Again, thanks to the tight integration of transcript and video, researchers will still have access to nonmanual information by watching the video. A screen shot from a sample ELAN transcript is provided in Figure 2.

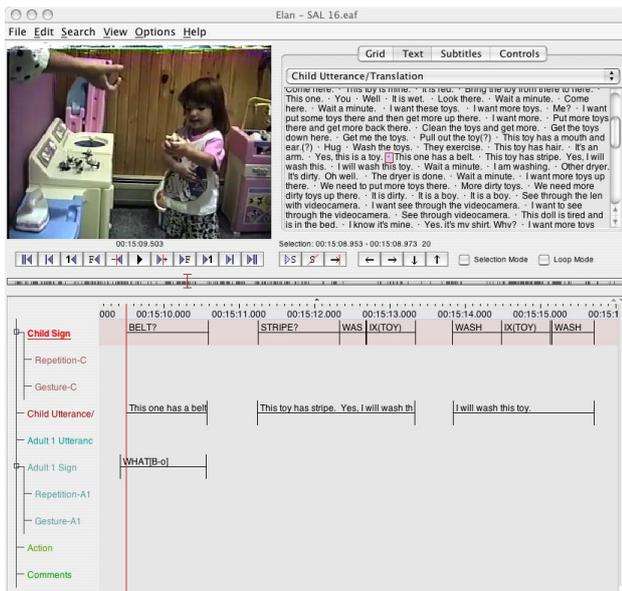


Figure 2: ELAN Screen Shot

With ELAN, corpus sharing among sign language researchers is finally becoming an attainable reality. Because the video is integrated so completely with the transcript, idiosyncratic notation conventions no longer pose as great an obstacle as they once did; researchers using transcripts generated by another research team can readily see which signs are represented by which glosses, then modify them with convenient Find and Replace features. ELAN offers a host of powerful features to facilitate annotation and analysis, yet is free to all researchers. In addition, it is available in both Mac and PC versions and is compatible with a variety of commonly used video file formats. These attractive features have established ELAN as the new standard for sign annotation. The existence of a widely-used standard has already opened the door for corpus sharing of adult sign data (eg. the European Cultural Heritage Online/ECHO site for Case 4: Sign Languages at <http://www.let.ru.nl/sign-lang/echo/>), paving the way for similar sharing of child sign data in the near future.

## 7. Continuing Challenges

Our experience of generating and sharing sign acquisition transcripts between our two research groups has been very promising so far, but of course, certain challenges remain. Some of these challenges are common to anyone in the business of generating corpora. Transcription of naturalistic video remains a long and tedious job, despite the welcome improvements that ELAN has brought us. We also struggle with decisions of how much of the video data to transcribe. We aim to generate transcripts that are as neutral as possible with respect to analysis, but invariably, the way we choose to gloss a sign or assign utterance breaks will reflect the analysis of the transcriber.

Others challenges are perhaps unique to naturalistic first language acquisition work. For instance, we film children

as they play, which means that they are constantly in motion. To increase our chances of keeping the children on camera, we avoid tight shots. The result is that the children's hands and faces look very small on the ELAN video, even when viewed in detached mode. To maximize video resolution in ELAN, we generally rely on H.264 compression rather than the .mpg format that is the default standard for ELAN. Whereas .mpg versions of our video files would be relatively small and portable, our H.264 video files are over 1GB each, posing difficulties for storage and transfer of data. We are optimistic, however, that solutions to this and other challenges lie in technological advances that we have yet to exploit. For example, we are currently in the process of establishing a server to house our corpus data at Gallaudet, which would allow us to overcome the challenge of sharing large video files. For us, the benefits of corpus sharing between our two research groups have clearly outweighed the challenges, and we will continue to seek ways to streamline and refine the process.

## 8. Conclusion

In this short paper, we have traced the evolution of our sign acquisition corpora from a transcript-centered format with little video integration to a fully integrated transcript and video format. Our goal has always been to efficiently create transcripts that are rich enough for the reader to locate regions of interest by scanning text. At the same time, we require ready access to video, to mitigate the limitations of English-based glosses and guard against analyses based on the transcripts alone. ELAN has facilitated enormous progress towards these goals and made it possible for our two research groups to share our sign acquisition corpora with great success. In the long term, we are hopeful that experiments in small-scale corpus sharing such as ours will one day lead to sharing on a much broader scale, of the type currently available for the spoken language acquisition community through online databases such as CHILDES. As mentioned at the start of this paper, many of the seminal studies on early sign language development are based on tiny sample sizes of two or three children. Given the enormous resources of time and money required to collect and code longitudinal acquisition data, shared databases are absolutely crucial to achieving larger sample sizes, which will permit replication and expansion of basic studies as well as increased possibilities for statistical analyses.

## 9. Acknowledgements

Collection of monolingual ASL data at the University of Connecticut was supported by grants to Diane Lillo-Martin from the National Institutes of Health, National Institute on Deafness and Other Communication Disorders, Grant #DC-00183; the National Science Foundation, Division of Behavioral and Cognitive Sciences, Grant #BCS-0078788; and the University of Connecticut Research Foundation. The collection of bimodal bilingual data by Deborah Chen Pichler and her research assistants is supported by a priority grant from the Gallaudet Research Institute. Both

groups wish to thank the signing children and their families who have graciously allowed us to film them over the last fifteen years, as well as the numerous research assistants who have been crucially involved in collecting and particularly in transcribing the data.

## 10. References

- Lillo-Martin, D., and Snyder, W. (2002) Cross-linguistic study of early syntax.  
<http://web.uconn.edu/acquisition/papers.html>
- MacWhinney, B. (2000). The CHILDES project: Tools for analyzing talk. Third Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- Stromswold, K. (1996). Analyzing children's spontaneous speech. In D. McDaniel, C. McKee, & H. Smith Cairns (Eds.), *Methods for Assessing Children's Syntax*. Cambridge, MA: MIT Press, pp. 23-53.
- Slobin, D. I, Hoiting, N., Anthony, M., Biederman, Y., Kuntze, M., Lindert, R., Pyers, J., Thumann, H., & Weinberg, A. (2001). Sign language transcription at the level of meaning components: The Berkeley Transcription System (BTS). *Sign Language & Linguistics*, 4(1-2), pp. 63-104.
- Snyder, W. (2007). *Child Language: The Parametric Approach*. Oxford University Press.

## Use of sign language materials in teaching

**Johanna Mesch, Lars Wallin**

Department of linguistics

Stockholm University

Sweden

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

### Abstract

In Sweden, we have started to use a digital version of the Swedish Sign Language corpus for teaching purposes. Some of the material is now used with students in two separate courses: Swedish Sign Language for beginners, and Swedish Sign Language Linguistics (for deaf and hearing signers). In this workshop we will present some teaching methods and technical problems. Selected examples are shown to demonstrate how students use the sign language corpus through the dictionary database, the corpus database and a learning platform for studying and analyzing sign language texts, like e.g. the small corpus in Bergman and Mesch (2004) and also some old and new recordings. Students have the opportunity to practice sentences, analyze the entries and annotate the texts or their own recordings. Bergman's earlier transcription system for Swedish Sign Language (Bergman 1982) has been updated continuously, and partly adapted for possible use as a standard annotation system. Problems with storing and using sign language material are also discussed.

Bergman, Brita. & Mesch, Johanna. 2004. *ECHO data set for Swedish Sign Language (SSL)*. Department of Linguistics, University of Stockholm.

Bergman, Brita. 1982. Teckenspråkstranskription. Forskning om teckenspråk X. Stockholms universitet, Institutionen för lingvistik)

### 1. Introduction

In this paper we describe how we have developed courses in text analysis and taught students to use ELAN<sup>1</sup> as a tool for reading, analysing, discussing and annotating. The teaching method was tested 2006, but here we discuss our teaching experiments during spring 2008. Since 1996 we have produced Swedish Sign Language dictionaries in digital version (see 'teckenlexika' on the web [www.ling.su.se/tsp]). We have compiled a large amount of materials in diverse places, such as in universities, national deaf association and its local clubs. There are TV-programmes and web information available in Swedish Sign Language, and some annotated materials in sign language researchers' collections. Today we are in the initial phase of building the Swedish Sign Language corpus in a new way due to a good combination of corpus sign linguistics and technology. A corpus with annotations and films is required not only for research but also for teaching and studying.

### 2. Course in text analysis

In Sweden, higher education is divided into three cycles: the undergraduate or first cycle (3 years, Bachelor's), the Master's level, or second cycle (2 years), and the doctoral (PhD), or third cycle (4 years). At Stockholm University we have two different types of courses in Sign Language: Swedish Sign Language for beginners and Swedish Sign

Language Linguistics (for deaf and hearing signers). The purpose of the course for beginners in Swedish Sign Language differs from the one for students of sign linguistics.

In this paper we will present some examples from the courses in sign linguistics that are called Sign Language I, 30 HECs (higher-education credits) and Sign Language II, 30 HECs. In both Sign language I and Sign Language II there is a course module in text analysis, 7,5 HECs. The course Sign Language I has the following four course modules:

#### Sign Language I

- Introduction to sign language and the sign language community, 7.5
- Form and meaning of the sign, 7.5
- Grammar, 7.5
- Text analysis, 7.5

The course module in text analysis includes:

The course module aims to give basic knowledge of conversation structure and narrative texts. In addition, the course module provides skills in using annotation tool for documentation and analysis of sign language materials.

The course module description defines the expected learning outcomes: what a student is be expected to know, understand, master or perform after having successfully completed the module.

After completing the course module Text analysis, 7,5 HECs on level Sign Language I the student has shown that

---

<sup>1</sup> ELAN (EUDICO Linguistic Annotator), today called Multimedia Annotator, see <http://www.lat-mpi.eu/tools/elan/>

s/he can:

- use a tool for transcription/annotation of sign language texts
- annotate sign language texts
- analyse main aspects of narrative and conversational structure in sign language texts

#### Sign Language II

- Sign structure, 7.5
- Grammar, 7.5
- Text analysis, 7.5
- Own work and linguistic production, 7.5

The course module Text analysis, 7.5 HECs in Sign Language II includes analysing sign language texts from linguistic and sociolinguistic perspectives.

Here it is important to observe the different levels of expected learning outcomes. The course module in text analysis in Sign Language I is directed at knowledge and skills in transcription of the manual and nonmanual forms according to annotation conventions. The course module in text analysis in Sign Language II entails deeper transcription with annotations of the morpho-phonological, syntactical and textual entries according to conventions for corpus work.

### 3. Method and materials

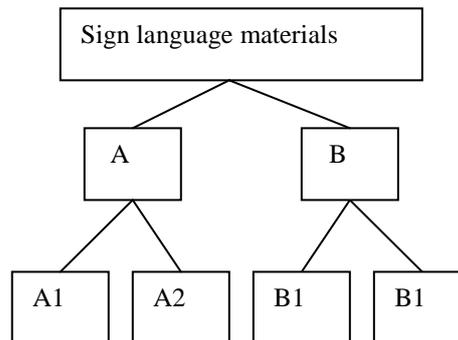
The annotation programme ELAN is gaining ground as a tool in sign language research, particularly because of its ability to facilitate collaboration between research groups who work separate from each other. By using ELAN it is possible for researchers to exchange data and to work together with similar methodology (Crasborn et al, 2007). In addition to research and corpus work, we have tested the use of ELAN and a small corpus in teaching sign linguistics.

Students were expected to learn how to use ELAN for searching and analyzing and annotating entries. A concise manual for ELAN was written in Swedish. Students received sign language materials in the format of mpeg and mov files with annotations in eaf files on DVD or HD in the class room or through Mondo<sup>2</sup> on the web. Then students could use the materials in their computers or in the class room.

The students used the sentences in the Swedish Sign Language dictionary on the web and the small corpus in Bergman and Mesch (2004) that is a starting point for studying and analysing sign language texts. Also some materials from the old and new recordings were available for annotation of texts, for example conversation, lecture, and stories of old deaf people, tactile sign language and

<sup>2</sup> Mondo is a learning- and collaboration system which is available to instructors, students and researchers at Stockholm University. The system is reached by using a web browser and is meant to help collaboration between users.

poetry. Some of the materials are not annotated, as shown in figure 1.



A – annotated material with ELAN

B – material without annotation (some annotated in another way)

A1 – search for the entries

A2 – analyze the entries on phonological, syntactical and textual level

B1 – annotate the material

B2 – analyze and discuss the results and the annotation problems

Figure 1: Structure in work with sign language texts

The students were told which parts of the sign language materials they had to search for information and to practice annotations. There were both annotated and non-annotated conversations to work with, and students had to annotate one part. We selected four different levels for working with ELAN: conversation analysis, lexicon/phonology, syntax, and text.

### 4. Some teaching and technical problems

During the teaching-learning process we have observed some problems in teaching and technical difficulties in text analysis. The annotation work has raised some questions, e.g. which tiers to use for annotation. We have been careful not to let students annotate too many tiers, because annotation work is time consuming and, consequently, would easily frustrate beginners. The most important tier of all is naturally the gloss. It is not difficult to annotate some lexical signs as BRA ‘good’ and FINNS-INTE ‘there is none’. It is more difficult to gloss some signs that have no easy translation in Swedish. It has been discussed if the Swedish Sign Language Dictionary should have a gloss-ID for annotating in order to make it easier for students to master signs and annotate gloss.

Some simple tiers are eyebrow, eye gaze and eye aperture. With the beginners we have skipped mouth movements because of their unclear definitions and the time-consuming annotation work. After annotating a while we discussed which signs are hard to gloss, as homonyms as well as compound signs. For the eyebrows, we have tried to describe how to separate form from

function in only annotating 'raised', i.e. not eyebrow as a marker for question.

At the Sign Language II level we also have conversation analysis, where students search for topics as well as topic change, turn taking, overlap and feedback. For annotating the entries, we have compared three different annotation conventions (Bergman, 1982; Johnston & Beuzeville, 2007; Nonhebel et al., 2004). We have studied how similar or different the annotation conventions are, and how Bergman's conventions fit in ELAN. Furthermore, we have discussed how much our conventions differ from other conventions, and if our future corpus should be translated in English and/or Swedish. We have found that Bergman's conventions differ from Auslan conventions where there is no division for homonyms, like VARFÖR 'why' and GÖRA 'make' (Swedish Sign Language examples). We have preferred to have two different glosses though the signs have same form.

In phonology we have discussed how exact the annotations of the glosses are, for example INDEX, no matter which hand shape (index finger, flat hand, thumb finger) they have, and whether there is a long or short, or no movement to breast. We have pointed out that it is the function, and not the form that changes annotation.

There are some variants, for example the sign SPRINGA 'run' and how to annotate these. The students have not agreed with the Auslan conventions with GLOSS-1, GLOSS-2 and GLOSS-3. They have preferred other examples, such as SPRINGA-knutenhand 'RUN-clenched hand', SPRINGA-krokfingerhand 'RUN-hook finger hand' and SPRINGA-dubbelkrokhand 'RUN-double hook hand'. Should we have such a SPRINGA-group for common gloss or specific gloss for each other sign in purpose to searching gloss or word in ELAN or Swedish Sign Language Dictionaries?

We have tested a combination of the conventions from Bergman, Auslan and ECHO. Purposefully we have looked for a user-friendly version for teaching, searching information and working with a large corpus. We have found the Auslan conventions interesting concerning dialogue materials. Head movement is one of the things that students have shown varying learning results in annotating. Usually the head is directed to the camera in a monologue text, but directed towards the conversation partner who sits on the right side in a dialogue text, though the signer's body is nearly directed to a camera. One student annotated that the signer in the movie has his/her head directed to the left, but another student annotated forward. Which of these directions is neutral? This is one of the many discussion topics. Another example of the topics discussed is how to annotate eye gaze on a signer, for example 'p' (person), and how often. If all eye gaze is annotated there will be too much information, and what is happening may not be understood. It is preferable not to annotate 'p' if it adds no

meaning. If the signer's head is directed to the addressee, not to the camera, the students have no longer annotated it.

When annotating eye gaze, we find the Auslan conventions work well, though not exactly describing the direction the eye gaze is directed to. It depends on if the signer is directed to the addressee, and how the addressee's mind perceives it, i.e. not as we look at it in a movie and see the signer directing her/his eye gaze, regardless of where the addressee is sitting. Still, it is important to annotate the direction of the eye gaze far left or right, up or down. We have discussed these directions with the students, and pointed out that they must observe the addressee's use of INDEX, head and eye gaze, and see if the addressee is telling things from a here-now-perspective. In addition, we have talked about possible surrogates (Liddell 2003).

One of the functions in ELAN is the possibility of expanding a transcription file with an individual student's home work (figure 2), as well as extra tiers for semantics and perspective changes of narrator and actors. A movie tells more than the annotations and compared to the transcribed sentences in earlier research reports.

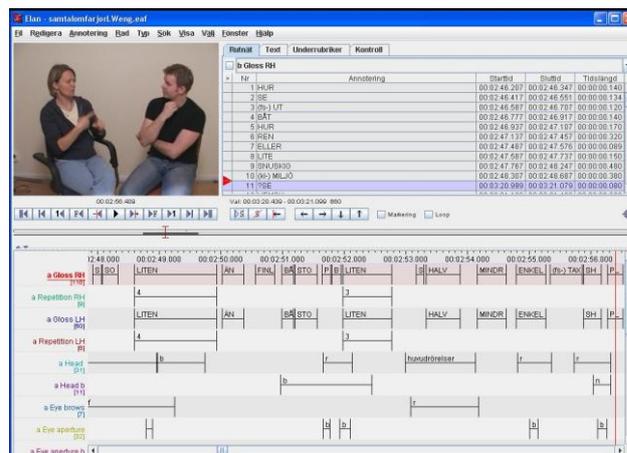


Figure 2: Student's work in ELAN

Another important function in ELAN is the possibility for a student to send an eaf file to a teacher as a completed assignment through Mondo (figure 3). The instructor is able to see the result of the students' homework individually. If there are many students in a course, it becomes hard to compare the students' eafs. It is preferable to have all tiers of many students' homework together in one and the same eaf file for comparing results. This is really a challenge we want to overcome in the future. So far we have decided to search for one word, for example 'captain' that we know is referred to several times in the text. The result, after searching for this word with all students' eafs, it shows how the students have



## Lexique LSF

Cédric Moreau(\*), Bruno Mascret(\*\*)

INS HEA \*

Institut National Supérieur de Formation et de Recherche pour  
l'Éducation 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

LIRIS (UMR 5205 CNRS) \*\*

Laboratoire d'InfoRmatique en Image et Systèmes d'information  
Université Claude Bernard Bâtiment Nautibus (710)  
43, Boulevard du 11 Novembre 1918  
69622 VILLEURBANNE CEDEX France

cedric.moreau@inshea.fr, bruno.mascret@liris.cnrs.fr

### Abstract

The French Sign Language (LSF) was banned in 1880 from all teaching institutions. From then on, it continued expanding in an uncoordinated way throughout special schools. In 1991, a new French law allowed deaf people to choose a bilingual education (French and sign language), and since February 2005 each school is required to integrate every devoted child who wishes it, no matter his handicap. All public websites must also become accessible.

With this new context, the LSF grows using regional differences, and users invent new signs to translate new concepts. However, the sign language cannot count on traditional media to spread out new expressions or words, since it is not spoken nor written. Therefore the sign vocabulary differs depending on geographical and social situations, furthermore if the concept is specific and elaborate. The website LexiqueLSF wishes to propose users a contributing and efficient tool, allowing a large diffusion of new signs and concepts.

A short analysis of the existing supports will lead us to present the main issues and to describe precisely the technical and linguistic solutions we chose, as well as some of the problems we met.

Likewise, all the elements composing the website should be considered as a concept in order to imagine complete accessibility to deaf people, and not only to blind people. We do not wish to make a simple dictionary. Our aim is to allow exchanges between users, to encourage them to invent and spread neologisms, and to make sure that the represented concepts are clear and understandable.

Publishing a new notion requires to create a number of descriptors (in French and in sign language, illustrations, examples... ) and to relate this notion to others already existing (opposite or similar concepts...). Each new sign proposed will be completely described, therefore it can easily be appropriated. Thanks to this organization, the same concept can be shown in different ways depending on the role it must play in context (classification, illustration, rendering, etc.).

One of the most important design patterns is the possibility of dynamically changing the classification system. Users will be able to choose various descriptors to build a classification view.

A reliable, but not compulsory, validation system will guarantee only serious suggestions. Three steps are needed: grammatical validation, sign validation (both require experts) and community validation.

Our production is thus very different from already existing paper or digital dictionaries, containing only everyday life vocabulary and almost no definitions, nor use examples. The best ones sort words according to the space location and configuration of the sign, but do not recognise morphological variations. Let us also observe that these dictionaries are not "bilingual" since they are accessible only to French speakers.

There are two discursive enunciation strategies according to (Cuxac, 2000). Signer may choose to show without saying, or to say in showing. In the future LexiqueLSF will try to manage both of this kind of signs: standards signs from dictionary and structures having a great iconic representation from morphemic elements.

### 1. Context presentation

The French Sign Language (LSF) was banned in 1880 from all teaching institutions. From then on, it continued expanding in an uncoordinated way throughout special schools. In 1991, a new French law allowed deaf people to choose a bilingual education (French and sign language), and since February 2005 each school is required to integrate every

devoted child who wishes it, no matter his handicap. All public websites must also become accessible.

With this new context, the LSF grows using regional differences, and users invent new signs to translate new concepts. However, the sign language cannot count on traditional media to spread out new expressions or words, since it is not spoken nor written. Therefore the sign vocabulary differs depending on geographical and social situations, further-

more if the concept is specific and elaborate.

## 2. Poster's aim

The website LexiqueLSF wishes to propose users a contributing and efficient tool, allowing a large diffusion of new signs and concepts.

A short analysis of the existing supports will lead us to present the main issues and to describe precisely the technical and linguistic solutions we chose, as well as some of the problems we met. This website must absolutely have a relevant and sharp classifying system, must be accessible to everyone, and offer new entries to satisfy all users.

The poster will explain the most interesting technical aspects of our work and main models, and will introduce our approach of the sign language representation, including browsing and enunciation strategies.

We will conclude in giving some information about the future of LexiqueLSF.

## 3. Technical aspects

We are presenting three main aspects: the content management system, the classification system and the validating system.

### 3.1. Giving sense through associations: the content management system

We needed a method providing the possibility to build a real accessible and dynamic website in French sign language. Our solution is a different use of the same content, depending on the role it takes in context.

To achieve this goal we have to manage the associations between the content elements, possible contexts (classifiers) and descriptors for those contexts (Bénel, 2003). For example, the content "Mathematics" can play several roles: classification (it includes all other mathematical terms), concept (the concept of "mathematics" itself), illustration ("mathematics" can be used as an illustration of the "science" concept), etc.

Other important roles are antagonistic concepts, similar concepts, similar signs, descriptions, examples of use, etc.

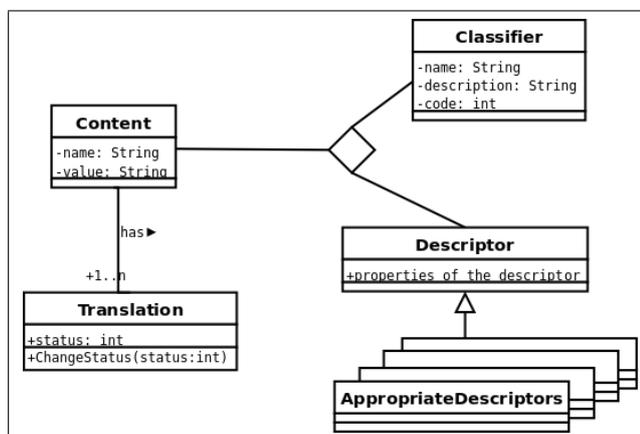


Figure 1: UML view of the data structure.

### 3.2. Offering several possibilities of classification

LexiqueLSF offers several possibilities of classification depending on the user's preferences, his goals or his intentions.

The main purpose is to avoid a unique classification system which cannot suit all users (for example, French lexicographic entries will be difficult for signers) nor purposes (Bénel, 2003; Bush, 1945).

That's why classification is represented as a role for content with custom descriptors. Implementing a new classification system is very easy: you create a new classifier and add (or reuse) descriptors. Then you can link content to this new role.

You may choose to use existing classifiers as well.

The last step is to build a new view for the classification system.

Unfortunately users cannot do this themselves. It would be interesting to add this feature in future.

We implemented only one system, for demonstration purpose, based on the Dewey's classification for library. This system is quite good for the beginning but has a lot of inconvenients and cannot remain the only way to access concepts, especially for signers.

Each content playing a Dewey's classifiers role has three descriptors: his own Dewey code, the Dewey code of his parent and his name as classifier (frequently his own name as content).

We are now implementing other classification systems based on sign configuration.

### 3.3. Validation and authoring tools

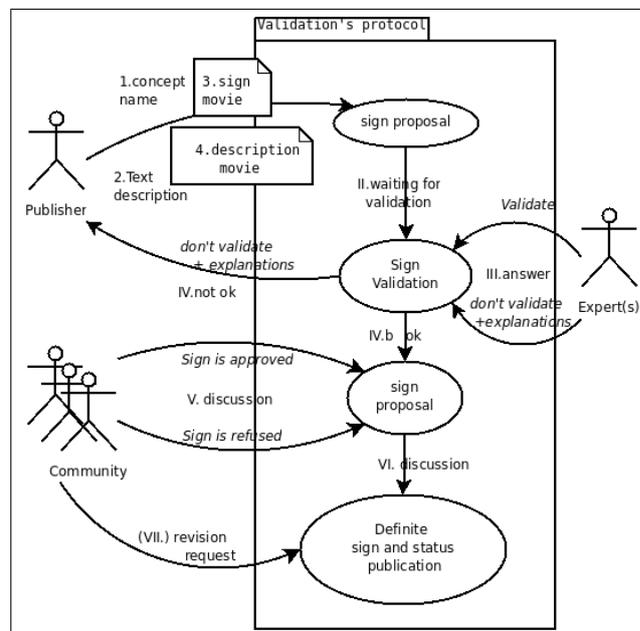


Figure 2: UML Case of utilisation: Validation system

The validation system will guarantee only serious suggestions. Three steps are needed: grammatical validation, sign validation (both require experts) and community validation.

This system improves wikipedia's one.

User's role is described in a descriptor for each classifier. Thus a user may be publisher for mathematics, expert for music and simple user for theology. User's privileges are obtained from the existing experts/publishers. It is explained in figure 2.

#### 4. Future prospects

To look for the existence or/and the definition of a sign (in FSL), most solutions require to know French. No bilingual French/FSL media exist. The lexicon is always defined starting from a French word to a FSL sign, using as much as possible a conceptual link between them. This artificially freezes FSL lexicon (Cuxac, 2004).

Best attempts in resolving this problem include other starting points : configuration (Moreau et al., 2007), spacial position and moves (Stokoe et al., 2000); orientation (Battison, 1974). However facial expression (Baker and Padden, 1978) has never been used as a research criterion. Moreover, none of these approaches consider the morphological variations of FSL signs.

According to (Cuxac, 2004), most standard FSL signs (i.e. those existing in dictionaries) include at least one internal morphemic component. He upholds the theory that the parametric elements composing a sign, these elements having no morphemic value, have a "phonetic" function and are necessary to realize the sign in good shape.

In this paper, we argue that parameters used in FSL signs

R. Battison. 1974. Phonological deletion in american sign language. *Sign Language Studies*, 5:1–9.

A. Bénel. 2003. *Chapters 5 and 7 in Consultation assistée par ordinateur de la documentation en sciences humaines : considérations épistémologiques, solutions opératoires et applications l'archéologie*. Ph.d. thesis, Institut National des Sciences Appliquées, LIRIS, Lyon, France, <http://www.revue-texto.net/Inedits/Benel/Benel.html>.

V. Bush. 1945. As we may think. *The Atlantic monthly*, 176:101–108.

C. Cuxac. 2000. In *La Langue des Signes Française : les voies de l'iconicité*, PARIS. OPHRY.

C. Cuxac. 2004. Phonétique de la Lsf: Une formalisation problématique. In Anne-Marie Bertonneau & Georgette Dal, editor, *La linguistique de la LSF: recherches actuelles*, pages 93–113. Série Silexicales 4.

C. Moreau, C. Tranchant, and al. 2007. Modèle de recherche de signes standard en langue des signes française. TOULOUSE, June. Traitement Automatique des Langues des Signes TALS.

W.C. Stokoe, D.C. Casterline, and C.G. Croneberg. 2000. In *A dictionary of american sign language*, WASHINGTON D.C. Gallaudet College Press.

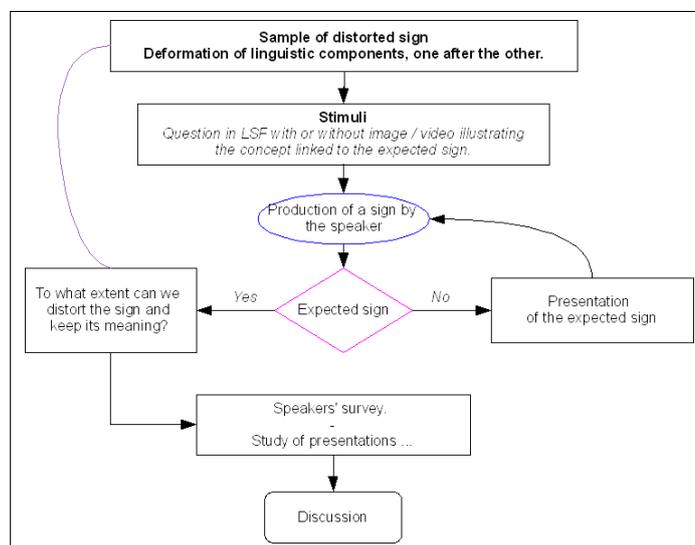


Figure 3: Questionnaire and interview protocol

are classified hierarchically. In order to test our theory, we propose a questionnaire and an interview of deaf signing speakers, as shown in figure 3.

#### 5. References

C. Baker and C. Padden. 1978. Focusing on the non manual components of asl. In P. Siple, editor, *Understanding Language Through Sign Language Research*, pages 27–57, New York. Academic Press.

## Construction of Japanese Sign Language Dialogue Corpus: KOSIGN

**Yuji Nagashima, Mina Terauchi, Kaoru Nakazono**

Kogakuin University, Polytechnic University, NTT Network Innovation Laboratories,

2665-1 Nakano, Hachioji, Tokyo 192-0015, Japan

Hashimoto-dai 4-1-1, Sagami-hara, Kanagawa, 229-1196, Japan

3-9-11 Midoricho, Musashino-shi, Tokyo 180-0012, Japan

E-mail: nagasima@cc.kogakuin.ac.jp, terauchi@uitech.ac.jp, nkzn@mac.com

**Abstract** This report presents a method of building corpuses of dialogue in Japanese Sign Language (JSL) and the results of the analysis in co- occurrences of manual and non-manual signals using the corpus. We have built the sign dialogue corpus by video recording the dialogues between native JSL signers. The purpose of building corpus is deriving electronic dictionaries such as morphological dictionary, different meaning word dictionary, allomorph dictionary and example dictionary. Example sentences are recorded for every word (key sign) those were recorded in the sign language word data base KOSIGN Ver.2. We were able to confirm a correlation of manual and non-manual signals or a characteristic appearance of sign language dialogue. As a result of the analysis, the pointing occurred to the end of sentence at high frequency. It suggested that pointing be one of the ends of sentence, and clarified the role as the conjunctive pronoun. The co-occurrence relation between the manual and non-manual signals acquired confirmed an important role to make the meaning of the expression sign language limited was achieved. Moreover, "Roll shift" and "Sandwich construction" that was the linguistic feature of sign language were confirmed, too.

### 1. Introduction

The engineering field has been applied to sign language by means of the development of a translation system between Japanese and sign language, construction of an electronic dictionary system, and study of a potential sign language recognition system. Moreover, linguistic analysis has been performed based on phonology, morphology, syntax, and structural analysis. In comparison with American Sign Language (ASL), Japanese Sign Language (JSL) has been studied much less and there is little data on the structural analysis of conversation and communication; further study is required.

Unlike spoken languages, which are expressed in one-dimensional fashion, sign languages are visual, expressed in a three-dimensional space. Sign language consists of manual signals and non-manual signals. Manual signals consist of hand shapes, palm directions, exhibited positions, and overall movements; sign language is mainly constructed by showing these motions simultaneously or continuously. Non-manual signals consist of facial expressions, mouth formations, nods, and line-of-sight motions, all of which are considered to work syntactically and semantically. Thus, various body motions are compounded to express information.

It is uncertain that which information is meaningful to which vocabulary word, and which information is a grammatical marker; in other words, the linguistic structure of sign language is unknown. Since a sign language notation system has not yet been systematically constructed, linguistic studies have made little progress. However, the construction of a sign language notation system has been tried by the SIGINDEX Workgroup. In order to explicate the linguistic structure, semantic structure, and cognitive mechanism of sign language

using linguistic analysis, accumulation of linguistic data called the corpus is essential. By the end of last year, the authors had photographed JSL dialogue samples in the participation of native signers and children of deaf adults (CODA) based on the sign language word database (KOSIGN Ver. 2), which contains 1,096 signs. One of our purposes was to digitalize dictionaries, such as the morpheme dictionary, homograph dictionary, synonym dictionary, and citation dictionary. Moreover, the authors have been analyzing inflectional endings and assimilation of words, grammatical functions of non-manual motions, and cognitive mechanism in a dialogue. Collocation in dialogue corpora was analyzed, and its characteristics were reported.

### 2. Collection method for dialogue corpora

Here, the outline of a dialogue corpus collection method is described. The dialogists participated in making dialogue corpora were the following two females: (1) Signer A, a Child of Deaf Adult (CODA) whose first language was sign language, and (2) Signer B, a congenitally hearing-impaired child whose parents were normal listeners. A dialogue sample was created against each vocabulary word (hereafter referred to as a key vocabulary word) in KOSIGN Ver. 2, according to the following procedure:

- (1) Before recording, signers A and B discussed the dialogue's scenario while looking at an image of the key vocabulary word.
- (2) The scenario was created, in which signer A asked a question (hereafter referred to as C1) and signer B gave an answer including the key vocabulary word (hereafter referred to as C2).
- (3) Videotaping was initiated.
- (4) Signer A asked C1 and signer B answered C2.

(5) In order to create a natural dialogue, signer A immediately gave an answer (hereafter referred to as C3) to C2, without any scenarios after the dialogue between C1 and C2.

(6) Signer B gave an answer (hereafter referred to as C4) to C3 without any scenarios.

Thus, a dialogue sample, with two correspondences regarding a key sign, was created. While the conversation was active, the dialogue continued. With this method, a natural dialogue could be created in the second correspondence, that is C3 and C4. When the sign language system adopted in KOSIGN Ver. 2 key differed from that used daily by signer A or B, that of signer A or B was preferentially adopted. Vocabulary words with low frequency of use in KOSIGN Ver.2, such as proper nouns, were deleted as candidates for key vocabulary words.

In the recording signers A and B did not directly dialogue. Two cameras recorded two signers, respectively, and each signer dialogued while looking at the projected front image of the other signer. Actually, signer A looked at signer B's life-size image, which was projected by a projector, and signer B looked at signer A's image, which was projected on a prompter. The size and position of each projected image were adjusted so that the signer's line-of-sight position was almost the same as that of a normal dialogue.

A camera (BVP-550, SONY Corporation, Japan) was used for recording at a shutter speed of 1/125 s. The front image of each signer was synchronously recorded using a recorder (DVW-500, SONY Corporation, Japan). During recording, in order to reduce the Japanese influence as much as possible, the prepared sample manuscript was not exhibited. The dialogue was initiated with expression of a KOSIGN sign, which would become a key sign, in sign language by one signer. Approximately 3,800 samples have so far been recorded, which include 944 KOSIGN signs.

### 3. Analysis method for dialogues

With reference to the dialogues created by the participants, and roughly extracted sign information, manual signals and non-manual signals were temporally extracted, together with the two signers. This extraction was performed using a dialogue analysis support system called Movie Analysis Tool (MAT). The characteristics of MAT are that multiple animated images can be controlled, each image can be tagged, and statistical information for each tag can be obtained in real time. Since tags of multiple images can be displayed on the same window, the relationship between the manual or non-manual signals of both signers can be grasped.

As part of the analysis, the meaning (reason) of a manual or non-manual signal during a dialogue was checked and noted as subsidiary information. The signal, expressed by a signer against a matter, could become to be important in interpretation of sign language expression. So far dialogue samples for 125 key signs have been

completely analyzed. Figure 1 shows an analysis image of MAT.

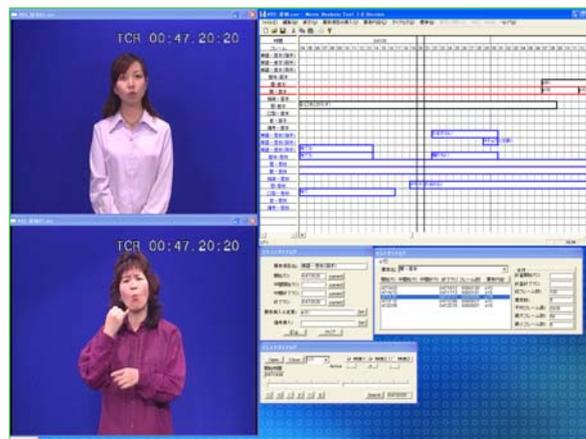


Figure 1: Analysis image of MAT

## 4. Results

Here, the characteristics are described of dialogue samples which have so far been completely analyzed are described. The classification and the numbers of words and samples, which have been recorded, are shown in Table 1.

Number of key signs	Number of samples	Number of signs	Number of sign types
125	563	4260	964

Table 1: Classification and numbers of signs and examples

The average number of vocabulary words used in one sample by each signer is shown in Table 2.

	Signer A	Signer B
Average number of signs used in one sample	14.5	19.5

Table 2: Each signer's average number of vocabulary words used in one sample

As shown in Table 2, the average number of vocabulary words used for one sample of signer B was larger than that for signer A. This was because the both sentences of signer A (C1 and C3) were questions, and a question usually consists of a short sentence. In the dialogue, signer B gave an answer (C2) against a question of signer A (C1), and signer A gave a question (C3) against C2. The frequency of hand usage (dominant hand, non-dominant hand, and both hands) used in one sample by each signer was investigated. The result is shown in Table 3.

	Dominant hand	Non-dominant hand	Both hands
Signer A	8.8	0.7	5
Signer B	10.4	0.9	8.3

Table3: Each signer's frequency of hands

When the dominant hand pointed at “that man,” third person finger-pointing (pT3), while the non-dominant hand simultaneously pointed at the man (oTOKO), this case was not classified into a group of both hands, and a vocabulary word was assigned to each hand.

sIGNDEX: pT3[oTOKO]

Most of the non-dominant hands were remainders of previous sign language. Non-dominant hand expressed alone was finger-pointing (pT) such as I, first person finger-pointing (pT1), or you, second person finger-pointing (pT2). Regarding this result, the signers were interviewed. They answered that they did not intentionally use their non-dominant hands. They unconsciously used their non-dominant hands under the influence of the final hand and finger positions of the previous sign language.

## 5. Collocation in sign language

Collocation is generated temporally or simultaneously in the same sentence. In linguistics, temporal collocation is called a syntagmatic relation and simultaneous collocation is called a paradigmatic relation. Here, the former is called temporal collocation and the latter is called simultaneous collocation following the literature. In this study, several collocations were observed.

### 5.1 Temporal collocation in pT

In JSL, pT is often observed. The frequency of pT in dialogue samples analyzed in this study is shown in table 4.

Sign	Frequency	Ratio (%)
pT2 (you)	465	12.51
pT1 (I)	345	9.28
pT3	201	5.41

Table 4: Frequency of pT

In JSL, it is said that a signer often finger-points to him/herself or a third person at the beginning of a sentence, and finger-points again at the end of a sentence. It is uncertain whether the second pT is always performed at the end of a sentence. However, since pT is often performed together with non-manual signals, such as nodding and winking, pT is considered to be a marker at the end of a sentence. Then, we noticed a repeated expression of pT. Among 563 samples against 125 key vocabulary words, pT was repeated in 129 samples (approximately 23%). The distribution of repeated pT is shown in Table 5.

	pT1	pT2	pT3
Signer A	17	50	15
Signer B	30	6	11
Total	47	56	25
Distribution	36.4%	43.4%	20.2%

Table 5: Distribution of repeated pT

In 96 samples of JSL as described in sIGNDEX V. 2, pT

was repeated in 20 samples (approximately 20.8%)[7]. Among the 20 samples, pT1 was observed in 17 samples, and pT2, pT3, or specific third person finger-pointing (pTR) were observed in 1 sample each. Since this analysis was not performed for dialogue samples, signers mainly finger-pointed themselves (pT1).

On the other hand, since dialogue samples were analyzed in the current study, the distribution of pT2 was the highest. However, the distribution significantly differed according to the signer; the distribution of pT2 was highest in signer A but the distribution of pT1 was the highest in signer B. The reason was considered to be that signer A finger-pointed signer B, who was a second person, in order to give a question (C1) and to express the completion of C1 to signer B. On the other hand, since signer B gave an answer (C2) regarding herself, pT1 was frequently performed. Further study is required to investigate the function of the first pT in temporal collocation.

### 5.2 Simultaneous collocation in pT

In spoken language, the meaning of a sentence can be changed by adding prosodic factors to the sentence. When intonation at the end of a sentence is changed, the grammatical meaning of the sentence may change. However, intonational changes cannot be expressed by written language.

In sign language, multiple words can be simultaneously expressed using both hands. Moreover, various meanings can be also expressed by the directions of body, face, and line-of-sight. Therefore, complicated content can be simultaneously expressed by adding non-manual signals to manual signals.

When both signals are simultaneously expressed, functions similar to prosodic factors in spoken language become effective. Moreover, by changes in speed, magnitude, appearance, iteration, and shape, many adverbial functions become effective. Thus, collocation in sign language is more complicated than that in spoken language.

We noticed the end of a sentence, and tried to determine the factors shown at the end of a sentence. In 563 samples analyzed in this study, pT was confirmed to exist at the end of a sentence in 197 samples (35%). The details are shown in Table 6.

	Question sentence	Answer sentence	Total
Signer A	82	28	110
Signer B	4	83	83
Total	86	111	193

Table 6: Frequency of pT at the end of a sentence

Signer A's pT were mostly observed in question sentences to signer B, and pT3 and pTR was occasionally performed. On the other hand, signer B's pT were mostly observed in answer sentences, in which the signer defined herself (pT1), and pT3 or pTR was performed only in one sample. Since the frequency of pT shown at the end of a sentence

was high, pT was suggested to be one of sentence end factors. Moreover, since dialogue data were used in this study, it was clarified that signers often used pT to confirm each other.

## 6. Synchronization phenomenon and collocation between signers

In order to synchronize the same manual signals, signers often confirm the contents of the conversation with each other. In sign language, they often exchange an agreement response {onaji (oNAJI)}. When confirming a size, signers often express the size mutually and simultaneously. For example, signer A confirms the size by a manual signal, which is shown by signer B, who also sometimes corrects it.

Many phenomena were observed in which signers' non-manual signals were synchronized. Among the non-manual signals, hDN was the most frequently observed signal.

Many phenomena were observed in which signers' non-manual signals were synchronized. Among the non-manual signals, hDN was the most frequently observed signal.

In order to express {katadukeru (KATADUKERU)} in sign language, hDN is indispensable. Speaker A understood speaker B's intention from the flow of dialogue and predicted that {katadukeru (KATADUKERU)} would come next. Therefore, their hDN rhythms were perfectly synchronized. In the dialogue corpora analyzed in this study, synchronization of non-manual signals, such as eyebrow motions, was also often observed in addition to hDN. Synchronized non-manual signals include emotional displays resulting from the other speaker's intention. However, collocation between manual and non-manual signals, which is indispensable in sign language expression, is considered to be strongly related to the synchronization phenomenon. Whether this synchronization phenomenon due to collocation between manual and non-manual signals occurs only in a dialogue between native signers is uncertain. Therefore, further investigation is required.

## 7. Conclusion

This study analyzed dialogue images of the sample corpora. Consequently, collocation and synchronization phenomenon were observed in the dialogues. By using a dialogue analysis support tool, correlation between manual and non-manual signals, and the relationship between two signers, could be analyzed, and characteristic phenomena in sign language could be investigated.

Collocation between manual and non-manual signals, which is obtained by native signers, was confirmed to play an important role in defining the meaning of sign language expression. Such information is indispensable in order for normal listeners to learn sign language. Therefore, further hearing investigation is

required. Moreover, collocation in sign language was suggested to influence the synchronization phenomenon between signers.

In the future, the grammatical functions and cognitive mechanism will be analyzed in addition to the further analyses of sample corpora.

Kazuyuki KANDA, et al. (2001). Notation System and Statistical Analysis of NMS in JSL, *Proceedings of HCI*, Vol.1, pp.25--27.

D. Mutoh and Y. Nagashima, (2001). Conversation analysis support system for sign video images, *Transactions of Human Interface Society*, Vol.3, No.3, pp. 153-162(in Japanese).

Prillwitz, S., Leven, R., Zienert, H., Hamke, T., Henning, J.(1989). HamNoSys Version 2.0;Hamburg Notation System for Sign Languages; *An Introductory Guide, volume 5 of International Studies on Sign Language and Communication of the Deaf*, Signum Press.

Kanda, K., Ichikawa, A., Nagashima, Y., Kato, Y., Terauchi, M.(2000). Signdex V.1 & V.2 The Computerized Dictionary of Japanese Sign Language; *Proceedings of TISLR(7th International Conference on Theoretical Issues in Sign Language Research)*, p.195.

# Documenting an Endangered Sign Language: Constructing a Corpus of Langue des Signes Malienne (CLaSiMa)

Victoria Nyst

Leiden University, Centre for Linguistics

Van Wijkplaats 4

2311 BX Leiden

The Netherlands

E-mail: v.a.s.nyst@let.leidenuni.nl

## Abstract

Langue des Signes Malienne (LaSiMa) is the native language of non-educated adult, mostly male Deaf signers in Bamako, Mali. It is currently endangered as the Deaf community in Bamako shifts to American Sign Language. The *Projet LaSiMa* aims at documenting and describing this language. One of the aims is to construct a corpus of LaSiMa discourse. This paper describes the methodology used so far to collect data. Methods developed for collecting data on Western sign languages appear to require adjustments at various points. Thus, filming signers in an unknown setting appears to hamper spontaneous language production with some signers. Also, some of the materials used in other sign language research appear to be too culture-specific. Criteria used to select signers need adjustment as well. The concept of *native signer* is less straightforward in the case of LaSiMa, due to its endangered status, the multilingual setting among hearing people in Bamako, and the influence of home signers. Methods for data annotation appear to need adjustment to the LaSiMa context as well.

## 1. Introduction

This paper discusses the construction of a discourse corpus of Langue des Signes Malienne (LaSiMa). In §2, I present this language, including its spontaneous evolution outside of the context of Deaf education, the influence of culture on gender differences, and lastly its endangered status. In §3, the *Projet LaSiMa* is presented, which aims at the documentation and description of LaSiMa. One of the aims of the project is the construction of the Corpus of Langue des Signes Malienne, CLaSiMa. In §4, the aims and methodology of the corpus project are presented. Most of the sign language corpora having been or being constructed concern (often Western) sign languages that have evolved in the context of Deaf education. The particular setting of LaSiMa creates circumstances that do not allow the direct transfer of methodologies developed for earlier sign language corpora projects. In §5, I discuss the particular features of the social context of LaSiMa that required the adjustment of methodologies used in corpus projects of other sign languages. These adjustments pertain to the gendered use of LaSiMa, cultural specificity of stimuli, lack of literate Deaf LaSiMa signers and the endangered status of the LaSiMa.

## 2. Langue des Signes Malienne

Langue des Signes Malienne (LaSiMa) is used by the adult Deaf community in Bamako, Mali. The exact number of users is unknown. However, given the UNICEF (1985) estimate of 0,5% for the incidence of moderate-severe hearing loss in developing countries, about 5000 Deaf people may be found to live in Bamako. Research is needed to establish to

what extent LaSiMa is used outside of Bamako.

Unlike most sign languages studied so far, LaSiMa has arisen outside of a school context, as a result of regular interaction between Deaf people in the streets of Bamako. Particularly the Malian *grins*; meeting places where men gather in the afternoon to chat and drink tea seem to be the cradle of this language. These tea meetings usually take place outside, in front of the building where one or more Deaf people work. Women rarely participate on a regular basis in these tea meetings and tend to spend most of the day in the more private sphere of a family compound. As such, men seem to have had a predominant role in the development of LaSiMa (cf. Pinsonneault, 1999).

As late as 1994, the first school for the Deaf was established in Bamako. This school initially used Langue de Signes Française, but switched to ASL at the end of the nineties. Soon after that, the second school for the deaf was established by a Canadian linguist, Dominique Pinsonneault, who also published a vocabulary of about 570 LaSiMa signs (Pinsonneault, 1999). The second school used LaSiMa as the medium of instruction. The difference in the language of instruction was deemed as impeding cooperation between the schools and in 2001 the second school replaced LaSiMa by a variety of ASL that is also used in other Francophone countries in Africa (Tamomo, 1994).

The decision to use ASL negatively impacts the usage of LaSiMa. World-wide an estimated 95% of deaf children are born to hearing, non-signing parents so that Deaf schools are a crucial factor in the transmission of a sign language. Although LaSiMa

has arisen outside of an educational context, present and future deaf children in Bamako grow up using ASL. Within the Deaf community, LaSiMa has a low status and is considered the language of the uneducated and illiterate. Several Deaf adults complain about being put under pressure to shift to ASL; they are urged to drop their LaSiMa style name signs (often depicting personal characteristics) and to take on new initialized, ASL-style name signs. LaSiMa signers are being mocked because of their signing. As a result, adult Deaf, native LaSiMa signers too are eager to learn and use ASL and most of them use a mix of LaSiMa and ASL at present. Thus, the Deaf community in Bamako is shifting to a variety of ASL at a considerable speed. Having virtually no child users and a lower status compared to ASL, LaSiMa should be considered a seriously endangered language (Cf. Wurm, 1998).

In view of its “natural” emergence outside of Deaf education, its endangered position, together with the lack of knowledge on African sign languages in general, LaSiMa needs to be documented and described before this is no longer possible. To this end, the *Projet LaSiMa* has been initiated.

### 3. *Projet LaSiMa*

The *Projet LaSiMa* aims at the documentation and description of LaSiMa as used in Bamako. In addition the project hopes that the research activities will lead to an increased interest for LaSiMa within the Malian Deaf community as well as in the wider hearing Malian society.

The main results we hope to produce are

- a corpus of LaSiMa discourse, using the ELAN software developed at the Max Planck Institute for Psycholinguistics, Nijmegen
- a lexical database of LaSiMa, using the LEXUS software, also developed at the Max Planck Institute for Psycholinguistics, Nijmegen
- articles on selected features of LaSiMa

### 4. The Corpus of LaSiMa

Thus, the aim is to collect diverse types of texts, i.e. spontaneous monologues and conversations. These texts should include different registers, such as informal and formal discourse, instructive discourse, joking stories, etcetera. Also, the topics of the discourse should be diverse to some extent including personal narratives, views on deafness and the sign language situation in Bamako, professional skills, etcetera. In addition to spontaneous discourse, data elicited with the aid of stimuli material should be part of the corpus, to facilitate cross-signer and cross-linguistic comparisons.

Not only the discourse types are to display a balanced variety, the signers to be filmed should show a representative variety with respect to age, gender, education, hearing status, age of onset of language acquisition and degree of multilingualism.

The recordings should be of good quality, both technically and with respect to its data. The data will be stored in several archives, among which at least one that is accessible through internet. We hope the *CLaSiMa* will be useful for sign linguists wanting to study LaSiMa, for other academics such as ethnologists and historians wanting to study the topics discussed in the discourse of the corpus, for professionals wanting to develop materials in LaSiMa and last but not least for the Deaf community who may want to use the corpus for several purposes.

#### 4.1 Data collection

The current corpus project is building on earlier work done in the construction of signed language corpora. This work mostly entails national sign languages, such as Sign Language of the Netherlands (Crasborn & Zwitserlood, 2007), Australian Sign Language (Johnston & Schembri, 2005) and Irish Sign Language (Leeson & Saeed, 2007). One of the advantages of sharing the approaches used in other sign language corpora is that it is likely to facilitate the comparative analysis of the data across corpora. The Dutch and the Australian corpora both have used a standard format to ensure a representative balance in discourse type and signers. Thus, a preset number of signers in a preset number of locations were asked to participate in a data collection session with a more or less fixed program of linguistic tasks. In the case of the Dutch corpus, pairs of signers were invited to a filming location where they discuss statements and responds to tasks involving stimuli that have been used in research on other sign languages as well.

As the construction of a corpus of an endangered sign language of a Deaf community with no formal education in a non-Western culture brings about specific circumstances, that may significantly differ from circumstances in the corpus work done so far on national sign languages, I will discuss in some detail the approach taken so far in the building of the *LaSiMa* corpus.

Data collection was initiated during my visit to Bamako from November 2007 till January 2008. Prerequisite for the data collection was the formation of a team of Deaf signers of *LaSiMa* who would lead the interviews. Identifying talented co-workers was facilitated by the linguistic research done on *LaSiMa* before by Dominique Pinsonneault. As a result of her

research, there were a handful of Deaf signers with some experience in participating in linguistic research. However, most of them were now predominantly using ASL or a mix of LaSiMa and ASL. Only one relatively younger signer (in his thirties), Siaka Keita, was found to be predominantly using LaSiMa, having made the decision not to give in to the pressure put on him to stop using LaSiMa and instead switch to the use of ASL. A second co-worker was found in the person of Moustapha Magassouba, who became deaf around the age of seven. He developed a home sign language until the age of fifteen, when he joined the Deaf school where LaSiMa was used. He is now bilingual in ASL and LaSiMa. Despite my efforts, I have not been able to find a female signer with a good command of LaSiMa and the time/possibility to join the project. Being Dutch and hearing myself, the first step towards any research activity was Siaka and Moustapha teaching me to communicate at a basic level in LaSiMa. During these LaSiMa classes, we managed to establish common ground with respect to the aims and methods of the project.

Initially, an approach was designed for the data collection of the LaSiMa corpus that was more or less similar to the approaches taken in the corpus projects for the sign languages of the Netherlands and Australia. That is, we formed a small studio outside of the building where I was staying using mats to provide an even background and to prevent people passing by from entering in the recordings. Our idea was that Moustapha would be filming, while Siaka would be interviewing two signers. Due to the advanced shift to ASL in the Bamako Deaf community, one of our main concerns was to find signers who would still be highly proficient in LaSiMa, be they monolingual or bilingual in ASL in addition to LaSiMa. As Siaka is renowned for his LaSiMa skills and his monolingualism in this language, we assumed that his role as an interviewer would motivate signers to use LaSiMa rather than ASL.

The standard format I had designed for the first try-out sessions included retelling the Canary row cartoon fragments used in earlier sign language and gesture research and free conversation on several topics including the onset of deafness, the attitude of the family and marriage. Interestingly, Siaka and Moustapha thought that the signers we wanted to film would not be very interested in talking about the sign language situation in Bamako.

However, this approach was challenged in several ways. As in earlier research in West Africa with signers without formal education (Nyst, 2007), monolingual signers appeared to be uncomfortable

retelling the canary row cartoon fragments shown to them. The “studio” set up of the recording sessions appeared to hamper monolingual signers in their spontaneous language production to such an extent that it was deemed favorable to displace the recording to the place where LaSiMa is naturally used: the *grins*, or tea groups of Deaf adults (see §2). We decided that interviews should be held in or close to the *grins*, in my absence. To this end, Moustapha and Siaka were trained in filming techniques. The interviews however were mostly held by Moustapha, due to the limited availability of Siaka. These interviews concentrate mostly on the topics mentioned above, i.e. the personal experiences of Deafness. In my absence, and in a familiar setting, there was no problem in capturing spontaneous sign production.

Obviously, this approach has advantages as well as disadvantages. The circumstances in terms of light and camera location can be manipulated less easily in the *grins*. Also, the fact that filming was a newly acquired skill for the filmers has influenced the recordings. This all led to lesser quality in the recorded images than originally aimed for. These disadvantages were counterbalanced by the advantages mentioned of more natural data. Moreover, the central role of Moustapha and Siaka is crucial for embedding the project in the community.

Issues requiring a modification of the initial approach for data collection included the age and gender balance, the “nativeness” of signers as well as the cultural appropriateness of the material and the tasks, and the spontaneity of the signers.

#### 4.2 Variation in signers and in language use

Sign language researchers working on national sign languages of countries with Deaf education are faced with the struggle to distinguish whether the sign language data they collected is on the right side of the continuum between the “native” end and the speech-supporting end. Although LaSiMa has emerged outside the context of Deaf education, a continuum of LaSiMa varieties ranging from little to more influence of Bambara, the dominant spoken language in Bamako is found for this language as well. Interestingly, young ASL signers in Bamako assert that they value LaSiMa, because it allows them to communicate with their family. As a consequence, these signers do not consider LaSiMa an endangered language. However, the signing referred to by them is likely to differ greatly in structure, prosody and lexicon size from the LaSiMa variety used among monolingual signers of LaSiMa. In the Deaf community in Bamako, the degree of hearing impairment is not a strict criterion for

membership. There does not seem to be a taboo for Deaf people to speak when they can. Deaf (and hearing) people considered good signers may actually sign and speak at the same time in their communication with other Deaf signers.

In the case of LaSiMa, and probably any sign language endangered by another one, an extra continuum of variation is added; the continuum of varieties more or less influenced by the dominant sign language. In the recordings made so far, almost all signers use ASL signs in their LaSiMa.

One may argue that there is yet a third continuum to be discerned. The majority of deaf Malian children, especially in rural areas, are likely to be isolated from regular interaction with a group of Deaf people. A number of these children –from within and outside of Bamako- will enter the Deaf community at an advanced age, after the critical period for language acquisition. Though prelingually deaf in many cases, these late learners of LaSiMa are likely to enlarge the scope of linguistic variation in the language as well.

Typically, sign language corpora focus on the language production of native signers, whereby the ultimate native signer is defined as being prelingually deaf and born to Deaf parents. In Bamako, the percentage of deaf children born to Deaf parents is likely to be relatively low for several reasons. Marriages between two Deaf persons are a new phenomenon in Bamako and there seem to be no family in Bamako well-known for multigenerational deafness. In short, it is hard to define the profile of the ideal signer who is representative for the “native” variety of LaSiMa. Variation according to the continua outlined above is a characteristic feature of LaSiMa as it is used today. Therefore, I aim to reflect this variation in the documentation project as well.

It was easy to find signers displaying the diverse linguistic backgrounds described above. It was much harder to find signers displaying diversity in age and gender. As in most cases of language endangerment, LaSiMa is mostly used by older people, in this case of 35 years or older. Thus, the rarity of LaSiMa discourse of younger people in the corpus is representative of the user community. More striking is the low number of women we have been able to record so far. Factors contributing to this low number may be the gender of the Deaf interviewers/filmers, as well as a dislike in some women of their image being taken. However, the most important factor seems to be the predominance of male signers in the LaSiMa user community. As described in §2, LaSiMa has evolved in public spaces where deaf people have a chance of meeting and gathering on a regular basis irrespective of the family,

neighborhood, social or ethnic group they belong to. However, many women spend most of their social life inside family compounds, where the chance of meeting another deaf person is small. In view of this pattern, it is likely that the sign language production of elder Deaf women in Bamako will be quite different from the LaSiMa production of elder Deaf men. On the one hand, one would expect that the female signing will tend towards the home sign end, showing structural difference with the signing common among male signers, as well as a smaller size of the lexicon. On the other hand, the female signing is likely to be less influenced by ASL, being more conservative in that respect.

Recordings during the remainder of the project will focus 1) on filming female signers of LaSiMa and 2) on finding ways to record LaSiMa responses to stimuli used in other sign language research to allow cross-linguistic analysis.

### 4.3 Annotation

So far, twenty hours of discourse have been filmed. The filmed discourse is to be translated in French using ELAN software. A selection of the discourse in the corpus will be glossed. The material will be deposited in a digital archive, where it will be accessible through internet for the academic as well as the Malian Deaf community.

The annotation process has not been started yet. Issues complicating the transcription are the rarity of Deaf LaSiMa signers with a good command of a written language. Educated Deaf signers typically use ASL. Hearing signers with a good command of LaSiMa are rare. The lack of trained hearing interpreters makes the option of providing voice-over translations impossible. The approach that will be taken now is to work with a team of a Deaf signer with a good command of LaSiMa and a second hearing or Deaf signer with a moderate command of LaSiMa, but good writing and typing skills.

## 5. Acknowledgements

I thank the numerous LaSiMa signers who have contributed to the corpus with their signing. I also want to express my gratitude to the Hans Rausing Endangered Languages Project for providing the funds for the *Projet LaSiMa*.

## 6. References

- Johnston, Trevor & Adam Schembri (2005) ‘The use of ELAN annotation software in the Auslan Archive/Corpus Project’ Paper presented at ‘The Ethnographic Eresearch Annotation Conference’ University of Melbourne, February 15-16, 2005.
- Leeson, Lorraine & John I. Saeed ‘Signs of Ireland:

- Developing a corpus of Irish Sign Language'  
Paper presented at the workshop 'The emergence  
of corpus sign language linguistics', BAAL 2007,  
Edinburgh, September 6, 2007.
- Nyst, Victoria (2007) *A descriptive analysis of  
Adamorobe Sign Language (Ghana)* Utrecht:  
LOT.
- Onno Crasborn & Inge Zwitterlood (2007) 'The  
Corpus NGT project: challenges in publishing  
sign language data'. Paper presented at the  
workshop 'The emergence of corpus sign  
language linguistics', BAAL 2007, Edinburgh,  
September 6, 2007.
- Pinsonneault, Dominique (1999). *Lexique des signes  
utilisés par les sourds du Mali*. Bamako: Editions  
Duniya.
- Tamomo, Serge. 1994. *Le Langage des Signes du  
Sourds Africain Francophone*. Cotonou: PEFISS.
- UNICEF (1985) *UNICEF report on prevention of  
Deafness: hearing aids*. London: UNICEF.
- Wurm, Stephen (1998) 'Methods of Language  
Maintenance and Revival, with Selected Cases of  
Language Endangerment in the World' In: Kazuto  
Matsumura (ed.) *Studies in Endangered  
Languages*. Tokyo: Hituzi Syobo, pp. 191--211.

## The representation Issue and its Multifaceted Aspects in Constructing Sign Language Corpora: Questions, Answers, Further Problems

Elena Antinoro Pizzuto<sup>1</sup>, Isabella Chiari<sup>2</sup>, Paolo Rossini<sup>1,3</sup>

<sup>1</sup>Istituto di Scienze e Tecnologie della Cognizione, Consiglio Nazionale delle Ricerche, Via Nomentana, 56 – 00161, Roma

<sup>2</sup>Dipartimento di Studi Filologici, Linguistici e Letterari, Università di Roma “La Sapienza”, P.le A. Moro, 5 – 00185 Roma

<sup>3</sup>Istituto Statale Sordi di Roma, Via Nomentana, 56 – 00161, Roma

E-mail: elena.pizzuto@istc.cnr.it, [isabella.chiari@uniroma1.it](mailto:isabella.chiari@uniroma1.it), [paolo.rossini@istc.cnr.it](mailto:paolo.rossini@istc.cnr.it)

### Abstract

This paper aims to address and clarify one issue we believe is crucial in constructing Sign Languages (SL) corpora: identifying appropriate tools for representing in written form SL productions of any sort, i.e. lexical items, utterances, discourse at large. Towards this end, building on research done within our group on multimedia corpora of both SL and spoken or verbal languages (vl), we first outline some of the major requirements and guidelines followed in current work with vl corpora (e.g. regarding transcription, representation [mark-up], coding [or annotation] Chiari, 2007; Edwards & Lampert, 1993; Leech & al, 1995; Ochs, 1979; Powers, 2005, among others). We highlight that a basic requirement of vl corpora is an easily readable transcription that, aside from specialist linguistic annotations, allows anyone who knows the object language to reconstruct its forms, and its form-meaning correspondences. Second, we point out how this basic requirement is not met in most current work on SL, where the ‘transcription’ of SL productions consists *primarily* of word-labels taken from vl, inappropriately called ‘glosses’. As argued by different authors (e.g. Cuxac, 2000; Pizzuto & al, 2006; Leeson & al, 2006), the use of such word-labels as a *primary representation tool* grossly misrepresents SL, even when supported by specialist linguistic annotations. Drawing on a relevant work on SL lexicon and discourse (e.g. Cuxac, 2000; Brennan, 2001; Cuxac & Sallandre, 2007; Russo, 2004; Pizzuto & al, 2008), we illustrate how the ‘transcriptions’ most widely used for SL are especially inadequate for representing complex sign units that are very frequent in SL discourse, and exhibit highly iconic, multidimensional/multilinear features that have no parallel in vl. Third, we discuss findings from ongoing research on Italian Sign Language (LIS) in which experienced deaf signers explore the use of SignWriting (SW: Sutton, 1999) as a tool for both composing texts conceived in written form -- thereby creating a corpus of written LIS -- and for transcribing corpora of face-to-face LIS discourse (Di Renzo & al, 2006; Di Renzo, in press; Lamano & al, in press). The results show that deaf signers can easily represent the form-meaning patterns of their language with an accuracy never experienced with other representation or annotation systems. We illustrate examples of SW-encoded vs. ‘gloss’-based transcripts which suggest that SW can be a valuable tool for addressing the representation issue in constructing SL corpora. However, the present computerized form of SW poses problems that constrain its use. We conclude specifying some of the problems that need to be faced on the route towards identifying more appropriate written representations of SL.

### 1. Premises

The observations reported in this paper are based on one assumption we wish to make explicit. We assume that, in order to qualify as a ‘corpus’, any collection of linguistic and communicative productions must include not only the ‘raw data’ themselves (as recorded, stored and more generally accessible via and on, for example, audiovisual tools such as digital video, videotapes, CD, DVD, multimedia tools of various sort) but also, as a necessary requirement, an easily readable transcription that aside from specialist linguistic annotation, and in the absence of the raw data, allows anyone who knows the object language to reconstruct its forms, and its form-meaning correspondences.

We also believe that such a requirement remains (and most likely will remain) a substantial tool for the linguistic analysis and description of any language, in spite of substantial advancements in computer-assisted treatments of language data. In the last decade or so, several researchers have pointed out how the vast majority of Signed Languages (SL) corpora currently available do not meet the requirement specified above.

The ‘transcriptions’ of SL data (even within multimedia tools) consist primarily of word-labels taken from verbal languages (vl), inappropriately defined ‘glosses’. It has been argued from different perspectives that the use of such word-labels as a *primary representation tool* grossly misrepresents SL, and renders extremely difficult to capture and analyze distinctive properties of SL lexicon and grammar, most notably with respect to complex, highly iconic structures, and multidimensional / multilinear features that have no parallel in vl (see among others Cuxac, 2000; Brennan, 2001; Leeson & al, 2006; Russo, 2004; 2005; Di Renzo & al, 2006; Pizzuto, Rossini & Russo, 2006; Vermeerbergen, 2006; Pizzuto & al, 2008, Slobin, 2008 – and references cited therein).

In this paper, we first provide an overview of general transcription requirements in linguistic corpora, and of available tools created for the most for spoken language (multimedia) corpora. We then focus on issues concerning the selection of relevant data, and of linguistic models in the construction of SL corpora, highlighting the representation problems posed by features that are unique of SL. We illustrate how we are currently addressing these problems in ongoing work using SignWriting (SW: Sutton, 1999) for representing

Italian Sign Language (LIS) texts. We conclude specifying technical and methodological problems that need to be faced on the route towards identifying more appropriate written representations of SL.

## 2. Transcription Requirements in Linguistic Corpora

SL corpora share with vl corpora the need of different layers of representation of linguistic and communicative aspects of discourse. The design and construction of language corpora make the issue of representation of linguistic data a crucial element both from a theoretical and a practical point of view, determining the need of explicit linguistic and logical models of representation and of formal standards employed in linguistic annotation. Some of the problems involved in transcription, mark-up and annotation of vl are relatively comparable with similar issues in SL corpora planning, even though the levels of representation of signed discourse are still open to debate and far from evident and shared in the scientific community.

In spoken and multimedia corpora the issue of transcription has been faced from a theoretical point of view (e.g. Ochs, 1979; Edwards & Lampert, 1993; Powers, 2005), from a descriptive point of view (e.g. Edwards & Lampert, 1993; Chafe, 1995; Cook, 1995; Leech, Myers & Thomas, 1995; Derville, 1997; Lapadat, 2000; Pallaud, 2003), and from a psycholinguistic perspective (e.g. Lindsay & O'Connell, 1995; Pallaud, 2003; Chiari, 2007), noting how it is inherently infused with linguistic theory and interpretation.

Recently the debate has been focused on different aspects of annotation of multimodal corpora, involving not only speech as the main linguistic signal, but also communicative information conveyed by posture, gesture, visual elements of context, and their interaction with verbal communication. *Linguistic annotation* is defined most generally as “any descriptive or analytic notations applied to raw language data [...]. In the simplest and commonest case, ‘linguistic annotation’ is an orthographic transcription of speech, time-aligned to an audio or video recording.” (Bird & Liberman, 2001). As Leech (1993: 275) states, corpus annotation is a procedure “for adding interpretative information to a text corpus”, centering on the *interpretative nature* of the process as defined by specific choices made by the annotator. The term annotation thus is generally used to cover both transcription practices and activities of addition of information of other nature (mainly metalinguistic glosses) such as part of speech, rhetorical, semantic description. The nature and typology of information deriving from transcription and description is nevertheless radically different even when coded within a similar format.

Transcription always involves a set of representational choices (Ochs, 1979), even when the aim is the reproduction of the spoken words, and not the more complex metalinguistic task of linguistic annotation. Transcribers' errors are common and, to a certain extent, unavoidable, following regular patterns of substitution, deletion, insertion and inversion

typically semantically-driven (Chiari, 2007). The mere act of converting spoken language into written language often involves practices of *naturalization*, such as conventions used to make speech conform to written standards. The process of approximating to speech thus involves the production of a text, the transcript, which becomes less readable in a conventional way: “the more a text reflects the oralness of speech, the less transparent it becomes for readers unaccustomed to encountering oral features in written discourse” (Bucholtz, 2000: 1461).

Annotation tools are thus the direct product of a specific linguistic model that declares the typology of layers required and the possible interconnection of the layers that the system can manage. Under this point of view many papers have focused on the issue of requirements for multimodal annotating tools, mostly devoted to vl (e.g. among others Ide & Brew, 2000; Bird & Liberman, 2001; Dipper, Götze & Stede, 2004; Garg et al., 2004). Among the properties of annotation that need to be fulfilled are reusability, flexibility and extensibility. Moreover the need of levels of annotation that cover new aspects and meet different purposes generally poses the question of developing specific tools that live and die within one project, or to use common frameworks that share an exchange format and API (Application Program Interface) (Ide & Brew, 2000).

## 3. Available tools and requirements for multimodal corpora

A number of different tools have been specifically created to perform the task of annotating multimodal audiovisual corpora. Among these: ATLAS<sup>1</sup>, MediaTagger (Senghas, 2001), TASX<sup>2</sup>, Anvil<sup>3</sup>, SyncWRITER (Hanke, 2001), NITE<sup>4</sup>, MMAX<sup>5</sup> - EXMARaLDA - EXtensible MARKup Language for Discourse Annotation (Schmidt, 2004), ELAN<sup>6</sup> and ANNEX, the web-based ELAN upgrade. Among tools specifically designed for SL corpora (but adequate for multimodal VL corpora too) are SignStream (Neidle, Sclaroff & Athitsos, 2001) and partly ELAN and ANNEX.

The suggestion of keeping transcription and mark-up separated is a capital issue both for verbal and SL corpora, determining the need of what has been called *stand-off annotation*. Most tools nowadays share the preference for the XML format, still controversy is open on what information to encode in this framework and how to encode it for general purposes of scientifically accurate corpus-based research and for possible commercial uses. Annotation schemes should be implemented using XML coding, or at least should envisage the possibility of exporting annotation in XML format as suggested in current guidelines for linguistic corpora following the TEI, Text Encoding

<sup>1</sup> ATLAS: <http://jAtlas.sourceforge.net/>

<sup>2</sup> TASX: <http://tasxforce.lili.uni-bielefeld.de/>

<sup>3</sup> Anvil: <http://www.dfki.uni-sb.de/kipp/anvil/>

<sup>4</sup> NITE: <http://nite.nis.sdu.dk>

<sup>5</sup> MMAX: <http://www.eml-research.de/english/research/nlp/download/mmax.php>

<sup>6</sup> ELAN: <http://www.lat-mpi.eu/tools/elan/>

Initiative (Barnard & Ide, 1997), Eagles and CES, Corpus Encoding Standard (Ide & Brew, 2000). Stand-off annotation is a need for it provides separate storage for data (audio, video signal) and description (transcription, annotation at different levels).

Common to most of these tools is an architecture that involves multiple layers of annotation, seen as tracks, that are filled with time-anchored elements. Layers are generally separate elements of annotation that represent different aspects of the communicative acts in a linear way, as for Anvil (Figure 1):

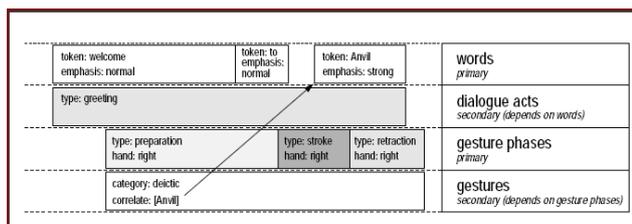


Figure 1: Anvil tracks (Kipp, 2001)

Anvil, a Java-based XML package for audiovisual annotation, enables multiple tracks that can be dependent on each other, and also links to mark co-reference. The direction in tool development runs towards the use of an object oriented system (like Java) and XML mark-up. A similar solution is that of EXMARaLDA and TASX, with the “single timeline, multiple tiers model”, where elements are connected directly to a basic transcription tier, which is connected to audio by absolute time-values (Figure 2).

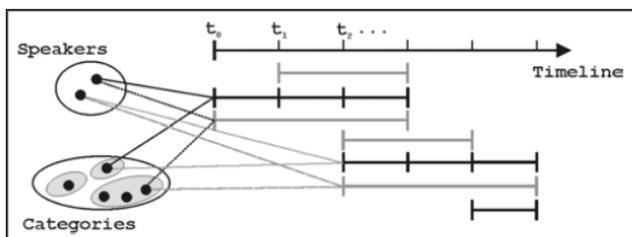


Figure 2: EXMARaLDA timeline model from (Schmidt, 2004: 2)

TASXS adds to the single timeline a multiple tiers model, with the capability of including hierarchical annotation structures, enabling the possibility of linking any annotation element belonging to one layer to the timeline or elements from other layers. Annotation should support labelling of time-aligned annotation layers and it should provide a supporting annotation scheme that takes into account the spatial and tri-dimensional space of the signed interaction. The annotation should be directly related to the video signal files.

While in spoken corpora specific instructions and conventions and annotation layers involve the representation of non-verbal data, such as contextual information, paralinguistic features, pauses, overlaps and other vocal phenomena, in SL corpora is presumable that processing, representation and annotation of these elements are involved in an often radically different manner, and interact in the discourse process in new ways that need to be taken into account

specifically.

From the reading of the annotation only it should be possible to predict the exact communicative gesture that was performed without looking at the video, taking into account the inevitable variability of the linguistic act itself. This task cannot be fulfilled by a representation system that contains only glosses. Some tools make a distinction between a basic, canonical or primary layer of annotation, which mainly consists of a word level transcription, and secondary or dependent levels of annotation.

The annotation format should be as neutral as possible from a theoretical point of view, simple, and it should be based on choices generally shared by the scientific community (Barnard & Ide, 1997). This means that transcription of linguistic data should be more descriptive than interpretative when dealing with basic or primary levels of annotation (that corresponding to spoken language transcription for verbal corpora). Müller & Strube (2003: 2) in the development of the MMAX tool for multilevel annotation argue, for example, that the annotation of what they call *base* (transcription of words) for spoken corpora “can be performed on a mere formal (i.e. surface-based) level, we believe these elements to be sufficiently objective to serve as the structure for what we call annotation *base data*”. Even if the notion of base data is not an undisputable and definite concept, it is clear that a formal description of linguistic data is an inevitable task to be performed on any kind of corpus data, while description of further levels of interpretation, from semantic to morpho-syntactic levels, is only a further step in linguistic data description.

In vl corpora tools, the base transcription (the verbal-tier) can be also used to represent a sort of temporal point of reference for all other entities at other annotation layers, like in the EXMARaLDA tool, if time-alignment is not directly linked to source video, as happens in ELAN. Using the v-tier as temporal reference would not be the best solution for SL corpora, since the lack of superficial linearity of transcription for the formal properties of signs can cause problems in relating other tiers to the base, which is by definition multidimensional. To overcome this problem an appropriate transcription system for SL should include the possibility of linking other annotation layers to single portions of the complex gestural sign (e.g. the formal elements representing eye direction or body positioning used to express reference should be individually linked to the co-reference annotation layer, even if they are part of a larger compositional and simultaneous whole sign). A general requisite is the possibility of using a common timeline independent from transcription and annotation layers, and directly connected with primary linguistic data source, namely digital video recordings.

The question to be posed would then be: what can constitute base data for SL corpora? Is there an annotation scheme that can be used? Are there appropriate tools to perform this annotation?

A separate question involves the different users on an annotation scheme, as claimed by Dipper et al. (2004), among which the annotator himself with different skills and training needs, and the corpus

explorer who needs readable, clearly understandable transcripts and annotations. Under this perspective while for vl corpora we have a number of usual standards for transcript presentation, SL corpora need to develop representation systems which are specific for the formal properties of the language, and can be easily recognized as such by signers, given the absence of a tradition in a shared written SL.

#### 4. Base data and linguistic models for SL corpora: some key issues

The question of what constitute base data for SL corpora cannot be appropriately addressed without taking in due account broader, strictly intertwined issues concerning the theoretical models adopted, and how they influence the data (and inevitably also the base representations) we select. Within the limits of this paper we can only touch these questions. With respect to theoretical models, we would like to recall here two major theoretical perspectives in past and current work on SL. One can be defined ‘assimilationist’, and exhibits a strong tendency to focus on SL data and features that demonstrate how, beyond ‘surface differences’, SL are deeply similar to vl. The other is a ‘non-assimilationist’ view highlighting several structural properties that sharply differentiate SL from vl. This latter view has been articulated most explicitly in extensive research conducted on French Sign Language (LSF) discourse and grammar, but also in work independently developed, in similar and compatible directions, on LIS and other SL (e.g. American, British, Irish, Flemish SL -- see for overviews, among others, Cuxac, 2000; Brennan, 2001; Cuxac & Sallandre, 2007; Pizzuto, 2007; Russo, 2004; Slobin, 2008; Vermeerbergen, 2006, Vermeerbergen, Leeson & Crasborn, 2007).

These two different perspectives provide equally different descriptions of two major kinds of units that can be identified in SL discourse: ‘standard’ signs, more or less easily translatable with spoken language ‘words’ (the so-called ‘frozen lexicon’), and complex, highly iconic constructions, consisting of manual and nonmanual elements arranged in a multidimensional and multilinear fashion that appears to be unique of SL (the so called ‘productive’ lexicon and morphology, e.g. Brennan, 2001). Following Cuxac (2000), we describe these complex sign units as Highly Iconic Structures (HIS).

A crucial point to be noted is that, in the theoretical framework proposed by Cuxac, the key feature for distinguishing HIS from standard signs (and the different, metalinguistic communicative intentions they express) is a SL-specific use of eye-gaze, which in fact renders eye-gaze a constituent parameter of sign units, and more generally of signed discourse, at several different structural levels (Cuxac & Antinoro Pizzuto, 2007). Standard signs are preceded or accompanied by eye gaze directed towards the interlocutor, whereas HIS are marked by gaze patterns directed towards the hands (in the production of two major subtypes of HIS characterized as Transfer of Form [TF] and of Situation [TS]), or via a gaze which mirrors the gaze of the referent(s) represented, in producing a third major type of HIS characterized as

Transfer of Person (TP).

Focusing on different formal and functional aspects of sign production, what we define here HIS have been characterized in most SL literature with different terms. For example, most researchers have focused on the manual components of TF and TS structures, describing them for the most as “classifiers” or, more recently, “property markers” (see Emmorey, 2003; Slobin, 2008, among others). TP structures are most often characterized as ‘role taking / shifting’, or ‘impersonation’ devices (but see also Dudis, 2004; Liddell, 2003; Slobin, 2008, for different proposals).

Figures 3a-3c below illustrates, with examples taken from LIS texts, the differences between the standard signs for ‘dog’ (3a) and ‘child’ (3b), and a TP construction (3c) used in the frame of a story to represent a ‘dog’ and ‘cat’ referents and the actions they performed. The structure in (3c) actually corresponds to what is characterized as a double TP, in which several different manual and nonmanual constituents are simultaneously arranged in time and space to represent two distinct referents and their actions, encoding the meaning “the child holds the dog in his arm, while the dog licks him on his cheek”.



3a: ‘dog’

3b: ‘child’

3c: TP

Research on LSF but also LIS and, more recently, crosslinguistic work on LSF, LIS and ASL, has documented that HIS are very frequent in SL discourse, with important variation related to discourse genre (e.g. Cuxac, 2000; Sallandre, 2003; 2007; Russo, 2004; Pizzuto, 2007). For example, Sallandre’s detailed analyses of a LSF corpus of narrative and prescriptive (cooking recipes) texts produced by 19 signers show that HIS constitute on average as much as 70% of the referential expressions identified in narrative texts, and 30% of those found in prescriptive texts. Similar patterns have been reported (using a different terminology), in Russo’s (2004) analyses of LIS poetic and non-poetic texts. A recent crosslinguistic study of narrative ASL, LSF and LIS texts shows that across these three SL HIS are by far the primary means for carrying out anaphoric reference in SL discourse: they constitute between 80% and 95% of the referential expressions used for this purpose (Pizzuto & al, 2008).

In summary, it is clear in our view that these data, framed within the non-assimilationist theoretical approach sketchily recalled above, challenge any description of SL that tends to underestimate (and underdescribe) the extent to which SL differ from vl. It is equally clear, however, that appropriate investigations of the structural features that distinguish SL from vl, and the creation of adequate corpora that can also be used for probing alternative theoretical models, require representation and annotation tools that are still to be developed. Towards this end, we believe

that it is crucial to develop a more profound reflection on *both* ‘annotation’ and ‘data editing’ problems, and how these are faced in vl as compared to SL research (see for example Blanche- Benveniste, 2007), *and* the tools that are used as primary means for creating the ‘basic transcripts’ of SL corpora. In our view, any basic transcript must satisfy at the very least the requirement of allowing the transcriber, and all researchers who use the transcript, to reconstruct the forms of the language (and not just the interpretation and/or analyses performed). It is useful to consider in some detail how different representational means succeed or do not succeed in meeting this requirement, and how they influence the analyses that can be done.

### 5. Experimenting with SignWriting: results achieved and perspectives

As reported elsewhere (Pizzuto & al, 2006; Di Renzo & al, 2006), since 2005 we have begun experimenting the use of SignWriting (SW: Sutton, 1999), in our research on LIS, as a tool for both composing LIS texts conceived directly in written form, and transcribing videorecorded corpora of narrative texts originally produced in a signed, face-to-face form. This research activity has been developed with the direct involvement of experienced Italian deaf signers who have produced all written texts, and transcriptions of LIS (signed) texts, realized thus far. Compared with any other written annotation tool previously experimented for LIS, SW has proven to be a tool much more effective, and easy to use, for representing the form-meaning patterns of the language (Di Renzo, in press; Di Renzo & al, 2006; Lamano, Luciola & Gianfreda, in press).

Drawing on research conducted by Di Renzo (2006; in press), transcripts (1) and (2) below, adapted from handwritten and computerized transcriptions made by Di Renzo, illustrate some of the insights that can be gained comparing the information encoded in two different transcriptions of the same, 9” fragment of LIS text. Transcript (1), was made with the graphic symbols of SW (hereafter ‘glyphs’, after Di Renzo et al, 2006); transcript (2) is realized with spoken-word labels (i.e. what are inappropriately defined ‘glosses’)

N	TIME	SIGN	TP	N	TIME	SIGN	TP
21	0.02.11		1	27	0.02.17		1 1
22	0.02.11		1	28	0.02.17		1 1
23	0.02.12		1 1	29	0.02.18		1 1
24	0.02.13		1	30	0.02.19		1 1
25	0.02.14		1 1	31	0.02.20		1
26	0.02.16		1				

Transcript (1)

N	TIME	GLOSS	SIGN	TP
21	0:02:11	GO-TO-BED (center to left)	1	1
22	0:02:12	(Conf B) SURFACE + (Conf 2) FEET slip inside	1	1
23	0:02:13	DOG	1	1
24	0:02:14	IMPERS (PAWS) REACH-FORWARD (centre to front)	1	1
25	0:02:15	IMPERS (PAWS under chin) FALL-ASLEEP	1	1
26	0:02:16	(Conf C G) JAR	1	1
27	0:02:17	FROG (NMC. sense-opportunity)	1	1
28	0:02:18	IMPERS (LEGS 3 3) CLIMB (alternated movem)	1	1
29	0:02:19	IMPERS (LEGS 3 3) JUMP-FORWARD (centre to right)	1	1
30	0:02:20	ESCAPE-AWAY (Conf L)	1	1

Transcript (2)

The text represented is taken from a classical ‘Frog story’ narrative produced by a 15 yrs old deaf girl in a study on the development of narrative skills in 15 deaf signing children and adolescents. The fragment can be translated in English as follows: “(the child) goes to bed, slips under the blanket. The dog searches for a place where to lie and falls asleep with his head on his paws. In the jar, the frog senses the opportunity, climbs up the jar, jumps out, escapes away”.

One of the aims of Di Renzo’s study was to ascertain the incidence of HIS compared to standard signs, and to assess at the same time the extent to which different transcription methodologies influenced the analysis and coding of signed texts. Di Renzo’s transcripts include numbers for each sign unit progressively identified in the text (e.g. N 21-31 / 21-30 in the leftmost columns of transcripts [1] and [2], respectively, where the number sequences reflect the temporal sequencing of the signs in the original signed text), annotation of the point in time marking the beginning of each unit (in minutes and seconds, column 2 in the transcripts), coding for HIS. For the present purposes, we focus only on the coding for TP constructions (marked by a ‘1’ under the TP column,

while no marks in the same column indicate that the sign produced was a standard one). The '1' under the 'SIGN' column are to be ignored: they were used for calculations performed on excel sheets linked to the transcriptions. All coding for HIS vs. standard signs was performed on the basis of the transcripts.

Both transcripts must be read from top to bottom. For anyone unfamiliar with the SW glyphs (and the way they have been adapted for realizing transcripts of LIS), the graphemic units in transcript (1) would require more explanations than it is possible within the limits of this paper. It will be hopefully sufficient to note the following. Each graphemic unit within a box aligned with the unit number (N) and time code identifies a major sign unit parsed by the transcriber. Within each such unit, the glyphs on the top (i.e. circle-like shapes, and the symbols within them) stand for head postures and facial expressions and their constituent elements (e.g. significant eye-gaze pattern, head movements). Especially important are the glyphs for eye-gaze patterns used in this transcript, partially modified with respect to the set provided within the SW system. Di Renzo uses a 'ii' graphic symbol for representing eye gaze directed to the interlocutor (e.g. see sign N. 22, 24, 27). The glyphs below the 'circle-shapes' represent hand configurations, location, movements and orientation patterns proper of the manual constituents of the signs identified.

In transcript (2) the signs are represented via 'glosses' in CAPS which render their basic meaning (e.g. N. 21: GO-TO-BED). Specific annotations before or after the 'gloss', in CAPS or in smalls letters, are annotations on structural/functional features the transcriber felt useful to mark (e.g. the annotation for sign N. 22 indicates that the sign unit produced was not a standard sign but a complex unit produced with a 'B' configuration representing a SURFACE and a '2' or 'V' configuration representing somebody's FEET, conveying the meaning 'slip inside').

An appropriate contrastive analysis of these two transcripts would require far more space than it is available here. We point out only some of the major differences. A first, striking difference is the following. Transcript (1) actually allows any LIS signer familiar with the SW system to reconstruct ("read aloud", if you wish) easily and accurately the signs, in the absence of the original videorecorded data. We found this could be done not only by the transcriber of the text (and even after one year since he made the transcript), but also by other signers in our group who had never seen the original data. The same was not true for transcript (2): no one (not even the transcriber himself!) was able to produce an accurate reconstruction of the forms of the signs 'represent' via 'glosses' and 'ad hoc' annotation. In our view, and in agreement with Di Renzo & al (2006), Di Renzo (in press); Lamano & al (in press), this shows an unquestionable advantage of SW-based representations that deserves to be highly valued and which, as far as know, has no parallel in any other form of written representation that has been proposed for SL. The possibility of reconstructing the original forms of the signs, hence also the relevant sign-meaning correspondences, is obviously crucial for any metalinguistic reflection on the text represented which, in turn, is a prerequisite for segmenting, analyzing and

coding it.

A second, very important difference concerns the prominence and high visibility/readability of nonmanual elements, in transcript (1), and more generally of the multidimensional and multilinear features of the signs. In the SW-encoded representation, all sign units include glyphs not just for the hand-articulators, but also for head/face/gaze/mouth patterns (indeed these proved to be essential for both reconstructing the signs and analyzing/coding them). The multilinear arrangements of the different manual and nonmanual articulators, and the distinct functions they play is immediately apparent, for example, in sign N. 23: the marked gaze (represented by upwards arrows) and facial expression easily identify the non manual component of the TP, and how this co-occurred with a complex manual construction where the two hands, and their simultaneously arrangement in space, concur to encode the meaning of a human referent 'slipping in under a flat surface'. In contrast, none of this information is easily readable or reconstructable from the representation of the same sign unit as annotated in transcript (2) under N. 22: the forms of the elements implicated simply are not there, but must be imagined/inferred (of course not necessarily correctly) from linearly arranged words which, in turn, provide a mix of translation, interpretation and annotation. This rendition of the sign unit also ignores nonmanual elements. Yet, it is interesting to note that in both transcripts (1) and (2) the sign construction under discussion is coded as a TP. The relevant difference is that transcript (1) provides us information on the form of the TP structure, expressed by a specific gaze, and represented via the SW glyphs, while transcript (2) does not. We have thus just to trust the transcriber on the accuracy of his coding. Similar remarks could be made for all the sign units represented in the two transcripts.

There are also relevant differences in the ways the text is parsed in the two transcripts, and these affect not simply the total number of meaningful units identified (10 in transcript [1] vs. 11 in transcript [2]), but especially, and more interestingly, what for lack of better terms we can define more qualitative aspects of the segmentation process. For example, unit 21 in transcript (2) is represented as one unit (following the usual convention that hyphenated glosses correspond to a single sign requiring more than one word to be translated with spoken language words). On this ground, one could make the generalization that the sign transcribed is a single lexical unit encoding a meaning comparable to the Italian (or English) complex verbal locution 'go to bed' (made of a verb, a preposition and a noun). However, transcript (1) provides a different representation, distinguishing two units (N. 21 and 22) for, respectively, 'go' and 'bed'. The different segmentation certainly does not tell us whether or not, in LIS, the meaning expressed is categorizable as a 'verbal locution' (as in Italian or in English), or as a simple verb + simple noun combination which has not the properties of a verbal locution. This remains to be decided on the basis of further analyses. Transcript (1), however, is more accurate because it allows us to reconstruct the forms that were produced, indicating that two (not one) lexical elements were implicated,

whereas transcript (2) provides a misleading parsing, which, upon reflection, the transcriber felt was influenced by the use of word-labels.

Similar segmentation differences emerge comparing units 27-30 of transcript (1) with units 27-29 of transcript (2). Space constraints do not allow us to describe these differences as it would be necessary, but it is important to note them because they highlight the crucial relevance that the representation system adopted has in describing and analyzing the elements that constitute a corpus.

Di Renzo (2006; in press) has provided clear evidence that, in the corpus of 15 narrative texts he analyzed, the use of two different transcription systems led to different results. These concerned both the global number of sign units identified in the corpus, and their structural properties. For example, there were significant differences between the number of TP identified via SW-based transcripts (markedly higher), and that of TP identified via 'gloss'-based transcripts (markedly lower). In our view, these data and results cannot be ignored if we wish to construct appropriate SL corpora, and adequate tools for analyzing and describing such corpora.

For example, suppose that we could implement as needed the computerized version of SW (e.g. to allow an easy transposition of handwritten SL texts in a computerized format), and that we could incorporate SW in multimedia tools currently available for handling multimodal language data, creating appropriate retrieval and searching devices whereby we could search the texts transcribed in a similar fashion as we normally do with spoken language multimedia data, i.e. by written representations of the forms of the language (linked as needed, depending upon the specific project, to both the original 'raw data', and the related annotation and coding that may have been introduced). The appropriate implementation we have in mind would of course have to be flexible enough to permit easy retrieval of *both* the constituent elements of sign units (an utility already available for lexical databases -- see the 'SignPuddle' tool created by Sutton and collaborators: <http://www.signbank.org/signpuddle/>), and the 'global gestalt' of simple signs (i.e. standard signs) and complex constructions such as HIS (e.g. N. 23 in transcript [1]).

If we could do this, we could obtain basic information on type-token frequencies (an information that is still so difficult if not impossible to obtain with the tools currently available), and begin to search for a wealth of morphological, lexical, morphosyntactic elements and structural regularities that are undoubtedly present in SL discourse, as it is evident from signers' actual processing and segmentation of signed discourse, and the metalinguistic analyses they are able to provide. Yet these regularities are extremely difficult to detect and describe, in a reliable manner, with the tools currently at our disposal. Most importantly, we could do our searches bypassing the strong limitations that any annotation / interpretation / coding *not accompanied by a representation of the forms of the language* inevitably imposes on any analysis.

## 6. Problems to be faced

At present, to the extent that we have been able to ascertain, the computerized implementation of SW still poses problems that constrain its use, especially for composing and analyzing *texts*. It is not by chance that our example of SW-encoded transcript was given above in its original, handwritten form. Transposing this text into a computerized format would have taken us more time than we could afford within the time constraints we had. More importantly, the tools available at present appear to be designed primarily for storing and retrieving primarily individual lexical items, rather than texts.

This bias towards creating lexical (rather than text) corpora, a bias that has a long history in SL research, may be particularly undesirable, especially if the individual lexical items that come to constitute 'databanks' are identified 'out of context', instead of being drawn from actual usage, i.e. from actual SL discourse. The danger is that of involuntary introducing significant distortions in the corpus, and in the sign representation / annotation process. In research on LIS, we have found that if the signs to be included in a databank are identified 'out of context', there is a strong tendency to: a) exclude a wealth of complex sign constructions that are very frequent in discourse (most notably HIS), and thus provide description limited to so-called 'standard' signs; b) disregard relevant nonmanual components which appear to be crucial for an appropriate description of the signs, as it is evident when the same signs (and even 'standard' signs) are drawn from discourse (Di Renzo & al, 2006; Di Renzo, in press -- see also Leeson & al, 2006). These differences linked to the 'sources' from which signs are drawn, along with more general considerations linked to the 'face-to-face' status of SL (see Pizzuto & al, 2006; Di Renzo & al, 2006; Vermeerbergen, 2006), support Russo's (2005) views on the need of constructing usage-based corpora of SL, i.e. of relying much more systematically than it is often done on actual SL discourse as a *primary* source for corpus construction.

With respect to storing and retrieving SW-encoded signs, we wish to note the following. We have not conducted a sufficient amount of research to evaluate whether the difficulties we have encountered are linked to relatively 'trivial' technical problems, or to more serious 'design-features' problems implicated in encoding and handling the complex graphic symbols of SW.

We would like to conclude posing a question that is admittedly naïve: is it possible that, in order to face (if not solve) the problems we all encounter and recognize in developing appropriate written representations for SL, we could all benefit from: a) knowledge stemming from the history and evolution of non-alphabetical writing systems (e.g. written Chinese), and b) the large repository of experience and 'know-how' implicit in the tools that have been developed for adapting non-alphabetic and/or non-roman alphabet writing systems (e.g. including Hangul) to a computerized format?

## 7. References

Barnard, D. T. & Ide, N. (1997). The Text Encoding

- Initiative: Flexible and Extensible Document Encoding. *Journal of the American Society of Information Science*, 48(7), 622-628.
- Bird, S. & Liberman, M. (2001). A formal framework for linguistic annotation. *Speech Communication*, 33, 23-60.
- Blanche Benveniste, C. (2007). Linguistic units in signed and verbal languages. In Pizzuto, E., Pietrandrea, P. & Simone, R. (eds.), *Verbal and Signed Languages - Comparing structures, constructs and methodologies*. Berlin / New York, Mouton De Gruyter, pp. 81-106.
- Brennan, M. (2001). Encoding and capturing productive morphology. *Sign Language and Linguistics*, 4: 1/2, 47-62.
- Bucholtz, M. (2000). The politics of transcription. *Journal of Pragmatics*, 32, 1439-1465.
- Chafe, W. (1995). Adequacy, user-friendliness, and practicality in transcribing. In G. Leech, G. Myers & J. Thomas (eds.), *Spoken English on computer: Transcription, mark-up, and application*. Harlow, England: Longman.
- Chiari, I. (2007). Transcribing speech: errors in corpora and experimental settings. *Proceedings of the 2007 International Corpus Linguistics Conference*, Birmingham  
<http://ucrel.lancs.ac.uk/publications/CL2007>
- Cook, G. (1995). Theoretical issues: transcribing the untranscribable. In G. Leech, G. Myers & J. Thomas (Eds.), *Spoken English on Computer: Transcription, Markup and Applications*. Harlow: Longman.
- Cuxac, C. 2000. *La Langue des Signes Française (LSF). Les voies de l'iconicité*. Faits de Langues, n. 15-16. Paris: Ophrys.
- Cuxac, C. & Sallandre, M.-A. 2007. Iconicity and arbitrariness in French Sign language: Highly iconic structures, degenerated iconicity and diagrammatic iconicity. In Pizzuto, E., Pietrandrea, P. & Simone, R. (eds.), *Verbal and Signed Languages - Comparing structures, constructs and methodologies*. Berlin / New York, Mouton De Gruyter, pp. 14-33.
- Cuxac, C. & Antinoro Pizzuto, E. (2007), The role of eye gaze direction in the construction of reference and the semanticization of the space of enunciation in signed languages. Paper presented at the 10<sup>th</sup> International Cognitive Linguistics Conference (ICLC), Krakow, Poland, July 15-20, 2007.
- Derville, B. (1997). Expression orale: du bon usage des transcriptions. *Le Français dans le Monde*, 287 (Feb-Mar), 29-32.
- Dipper, S., Götze, M. & Stede, M. (2004). Simple Annotation Tools for Complex Annotation Tasks: an Evaluation. *Proceedings of the LREC Workshop on XML-based Richly Annotated Corpora*, Lisbon, 54-62
- Di Renzo, A. (2006). *Le produzioni narrative in LIS di bambini e ragazzi sordi*. Tesi di laurea, Università di Roma "La Sapienza", Facoltà di Psicologia 2 (undergraduate thesis).
- Di Renzo, A. (in press). Esperienze e riflessioni su metodi di trascrizione della LIS. *Atti del 3° Convegno Nazionale sulla Lingua dei Segni*, Verona, March 2007.
- Di Renzo, A., Lamano, L., Lucioli, T., Pennacchi, B., Ponzio, L., (2006), Italian Sign Language: Can we write it and transcribe it with Sign Writing ? In C. Vettori (ed.), *Proceedings of the Second Workshop on the Representation and Processing of Sign Languages, LREC-2006*, Pisa: ILC-CNR, pp. 11-16.
- Dudis, P. (2004). Body partitioning and real-space blends. *Cognitive Linguistics*, 15: 223-238.
- Emmorey, K. (2003). *Perspectives on Classifiers Constructions in Sign Languages*. Mahwah, NJ: Erlbaum.
- Edwards, J. A. & Lampert, M. D. (1993). *Talking data : transcription and coding in discourse research*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Garg, S., Martinovski, B., Stephan, S. R. J., Tetreault, J. & Traum, D. R. (2004). Evaluation of Transcription and Annotation Tools for a Multi-modal, Multi-Party Dialogue Corpus. *Proceedings of Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, 1707-1710.
- Ide, N. & Brew, C. (2000). Requirements, Tools, and Architectures for Annotated Corpora. *Proceedings of Data Architectures and Software Support for Large Corpora*, Paris: European Language Resources Association, 1-5.
- Hanke, T. (2001). Sign language transcription with syncWRITER. *Sign Language & Linguistics*, 4, 275-283.
- Kipp, M. (2001). Anvil - A Generic Annotation Tool for Multimodal Dialogue. *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, 1367-1370.
- Lamano, L., Lucioli, T. , Gianfreda (in press) I segni sulla carta : analisi e riflessioni sui primi testi di LIS scritta. *Atti del 3° Convegno Nazionale sulla Lingua dei Segni - Verona*, March 2007.
- Lapadat, J. C. (2000). Problematizing transcription: purpose, paradigm and quality. *International Journal of Social Research Methodology*, 3, 203-219.
- Leech, G. (1993). Corpus Annotation Schemes. *Lit Linguist Computing*, 8(4), 275-281.
- Leech, G. N., Myers, G. & Thomas, J. (1995). *Spoken English on computer : transcription, mark-up, and application*. Harlow, Essex, England ; New York: Longman.
- Leeson, L., Saeed, J., Byrne-Dunne, D., Macduff, A. & Leonard, C. (2006). Moving heads and moving hands: Developing a digital corpus of Irish Sign Language. The 'Signs of Ireland' Corpus Development Project. *Information Technology and Telecommunication Conference 2006*, Carlow, Ireland, 25-26 October 2006.
- Liddel, S.K. *Grammar, Gesture and Meaning in American Sign Language*. Cambridge: Cambridge University Press.
- Lindsay, J. & O'Connell, D. (1995). How do transcribers deal with audio recordings of spoken discourse? *Journal of Psycholinguistic Research*, 24(2), 101-115.
- Muller, C. & Strobe, M. (2003). Multi-level annotation in MMAX. *SigDial,2003, Proceedings of the 4th SigDial Workshop on Discourse and Dialogue*, Sapporo, Japan.
- Ochs, E. (1979). Transcription as theory. In E. Ochs & B. Schieffelin (Eds.), *Developmental pragmatics*. New York: Academic Press.
- Neidle, C., Sclaroff, S. & Athitsos, V. (2001).

- SignStream: A tool for linguistic and computer vision research on visual-gestural language data. *Behavior Research Methods, Instruments, & Computers*, 33, 311-320.
- Pallaud, B. (2003). Splendeurs et misères de la transcription. *Cliniques méditerranéennes*, 68(2), 59-73.
- Pizzuto, E. (2007). Deixis, anaphora and person reference in signed languages. In E. Pizzuto, P. Pietrandrea & R. Simone (eds.), *Verbal and Signed Languages - Comparing structures, constructs and methodologies*. Berlin / New York: Mouton De Gruyter, 275-308
- Pizzuto, E., Rossini, P., & Russo, T. (2006). Representing signed languages in written form: questions that need to be posed. In C. Vettori (ed.). *Proceedings of the Second Workshop on the Representation and Processing of Sign Languages, LREC-2006*, Pisa: ILC-CNR, 1-6.
- Pizzuto, E., Rossini, P., Sallandre, M-A, Wilkinson, E. (2008). Deixis, anaphora and Highly Iconic Structures: Cross-linguistic evidence on American (ASL), French (LSF) and Italian (LIS) Signed Languages. In R. M. de Quadros (ed.), *Sign Languages: Spinning and Unravelling the Past, Present and Future - TISLR 9*, <http://www.editora-arara-azul>
- Powers, W. R. (2005). *Transcription techniques for the spoken word*. Lanham, MD: AltaMira Press.
- Russo, T. (2004). *La mappa poggiata sull'isola – Iconicità e metafora nelle lingue dei segni e nelle lingue vocali*. Rende: Università della Calabria, Centro Universitario e Librario.
- Russo, T. (2005). Un lessico di frequenza della LIS. In De Mauro, T. & Chiari, I. (eds.), *Parole e Numeri – Analisi quantitative dei fatti della lingua*. Roma: Aracne.
- Sallandre, M-A. (2003). *Les unités du discours en Langue des Signes Française (LSF) – Tentative de de categorization dans le cadre d'une grammaire de l'iconicité*. Thèse de Doctorat en Sciences du Langage, Paris, Université Paris 8.
- Sallandre, M-A. (2007). Simultaneity in French Sign Language Discourse, In Vermeerbergen, M., Leeson, L. & Crasborn, O. (eds.), *Simultaneity in Signed Languages. Form and function*. Amsterdam: John Benjamins. 103-125.
- Schmidt, T. (2001). The transcription system EXMARaLDA: An application of the annotation graph formalism as the Basis of a Database of Multilingual Spoken Discourse. *Proceedings of the IRCS Workshop On Linguistic Databases, 11-13 December 2001*, 219-227.
- Senghas, A. (2001). Spatial and temporal coding of Nicaraguan Sign Language in MediaTagger: Documenting three dimensions with a two-dimensional tool. *Sign Language & Linguistics*, 4, 229-240.
- Slobin, D.I. (2008). Breaking the Molds: Signed Languages and the Nature of Human Language, *Sign Language Studies*, 8:2, 114-130.
- Sutton, V. (1999). *Lessons in SignWriting. Textbook & workbook*. La Jolla, CA: Deaf Action Committee for Sign Writing (2<sup>nd</sup> edition, 1<sup>st</sup> edition 1995).
- Vermeerbergen, M. (2006). Past and current trends in sign language research. *Language and Communication*, 26, 168-192.
- Vermeerbergen, M., Leeson, L. & Crasborn, O. (eds.), (2007). *Simultaneity in Signed Languages. Form and Function*. Amsterdam: John Benjamins.

## 8. Acknowledgements.

We gratefully acknowledge partial financial support from the Istituto Statale Sordi di Roma (ISSR) and the Association “Progetti Felicità” (Joint ISTC-CNR/ISSR project “Writing LIS & SignWriting”), and from CNR and CNRS (French- Italian Project “Language, its formal properties and cognition: what can be learned from signed languages” - ISTC-CNR, Rome & UMR 7023 CNRS, Paris, 2006-2007).

# DGS Corpus Project – Development of a Corpus Based Electronic Dictionary German Sign Language / German

**Siegmund Prillwitz, Thomas Hanke, Susanne König,  
Reiner Konrad, Gabriele Langer, Arvid Schwarz**

Institute of German Sign Language and Communication of the Deaf (IDGS), University of Hamburg  
Binderstr. 34, 20146 Hamburg, Germany

{Siegmund.Prillwitz,Thomas.Hanke,Susanne.Koenig,Reiner.Konrad,  
Gabriele.Langer,Arvid.Schwarz}@sign-lang.uni-hamburg.de

## Abstract

This paper introduces a 15-year project, which aims to combine the collection of a large corpus with the development and production of a comprehensive, corpus-based electronic dictionary of German Sign Language (DGS). The scope, aims and the methodological approach of this large-scale project, accepted for funding by the Hamburg Academy of Sciences, are discussed.

## 1. Introduction

In modern lexicography there is an increasing awareness that dictionaries should be corpus-based. In sign language lexicography only few corpora exist and none of the larger corpora have been thoroughly transcribed and analysed for dictionary production. This gap is going to be filled by the project presented here. Different sources are used first to collect language data and later to analyse and validate the data on an empirical basis. The primary source is a sign language corpus to be built during the project. A voting system as well as a focus group complement the corpus data. In addition, previously published sign collections are taken into account.

The corpus is completely annotated and lemmatised. Further analysis and detailed annotations of selected lemmas are made particularly with regard to the compilation of the dictionary and the dictionary grammar. If necessary, others of the sources mentioned above support the ongoing analysis.

The primary products are a large reference corpus, partly published with English translation, and a comprehensive, corpus-based electronic dictionary of German Sign Language (DGS) – German, including a dictionary grammar. The dictionary will be preceded by a preliminary collection of basic vocabulary of DGS published after five years.

In the following, we outline properties of the corpus and the elicitation settings as well as the different annotation processes. The various sources of information we use, and their respective functions are introduced and distinguished. Features of the analysis up to the composition of the dictionary entries are presented. Finally, the products resulting from the project are briefly described.

## 2. Corpus design

A corpus of approximately 350–400 hours from 250–300 informants will be collected. We anticipate a number of

approx. 2.25 m. tokens. This is, in size and scope, comparable to large spoken language corpora.

The corpus should reflect a representative and well-balanced part of everyday communication of competent deaf signers. Informants from all over Germany of different age, sex and social status are included. They are selected on grounds of their language competence, their membership in the deaf community, and their regional rootedness. For logistic reasons the elicitations are carried out in ten cities, all of which are areas of relatively high deaf population density and are easy to reach from surrounding rural areas. A mobile studio is set up and is moved from place to place during the first three years of the project.



Figure 1: Preliminary set of locations

The design allows the use of the corpus for various tasks, such as the validation of the basic vocabulary, thorough research on DGS grammar based on the transcriptions and the identification of different meanings and collocations of a sign by appropriate contexts. Furthermore, the design anticipates a comparative sociolinguistic study compara-

ble in kind and quality to Lucas et al. (2001) and Schembri/Johnston (2004).

### 3. Elicitation

The elicitations follow best current practice, considering not only academic but also social, political, ethical and legal aspects. To describe filmed material on a metadata level, we make use of the IMDI standard (IMDI 2003, Crasborn/Hanke 2003)

Pairs of two informants are appointed for each elicitation session. The elicitations are carried out using a peer-to-peer procedure where two informants change roles according to situations. The interviews are conducted by a deaf contact person from the respective region to secure an elicitation of regional sign variants with as little influence from the interviewer as possible.

The elicitation consists of

- a standardized interview covering language and social data (approx. 20 min./informant)
- the filming of spontaneous conversations on a given topic (selected from a list of topics, 60-90 min. for each pair of informants)
- different tasks with selected stimuli (approx. 120 min. for each pair).

The major aim of the spontaneous conversation is to elicit as much basic vocabulary as possible, i.e. signs, which are used in every-day situations on a regular basis. Therefore, a list of topics is supplied, covering about 20 subject areas. From this list, the informants are asked to select 2-3 topics.

The different tasks with selected stimuli aim at capturing special phonological, morphological, syntactic and lexical phenomena. This explains why the stimuli tasks are given more time compared to the spontaneous conversations. We use stimuli from already carried out or planned elicitations and, if necessary, adapt them for DGS in order to enable parallel cross-linguistic analysis.

The informants are seated more or less facing each other, following a preliminary survey of informants' preferences. Each informant is filmed by a high-definition camera, two additional DV-cameras allow for a bird's eye view for an in-depth analysis of the use of space and a third camera films the whole scene to get a complete picture of the informants' interaction with each other.

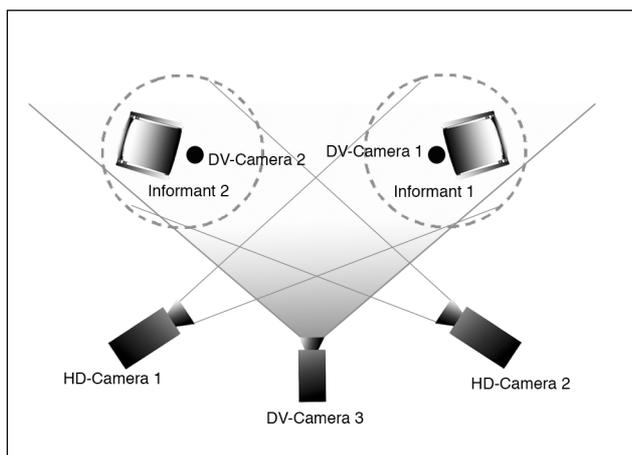


Figure 2: Elicitation setting

Elicitation sessions are expected to last approximately four hours. We anticipate about 80 min utilisable video material per informant.

The elicitation process is recorded, the obtained data is digitised on-site and specific (esp. linguistic) features are noted. The films are successively made available for the basic transcription.

### 4. Annotation

The corpus data will undergo different annotation and transcription processes aimed at identifying signs and documenting their properties. After a translation of the conversations into written German, a basic transcription serves to segment utterances and to identify lexical items and thus to provide a first access to the data. Second, approximately 50% of the transcriptions will be transcribed again in more detail.

#### 4.1 Translation and Segmentation

As a very first step the elicited DGS material is translated into German by trained interpreters and synchronised with the DGS texts. Thus, the content of the interviews is captured and becomes searchable via written German. The alignment also provides a first rough segmentation of the signed text into meaningful sections. Furthermore passages which are special with respect to technical, linguistic or content matters are marked and documented in a report by the interpreters. These remarks, among other criteria, are taken into account when assigning priorities for further analysis to certain sections. They are also taken into consideration when selecting parts of the corpus for the detailed transcription and for publication.

#### 4.2 Basic transcription

The basic transcription serves to segment utterances and to identify lexical items. It provides a first and easy access to the individual signs. The previous segmentation resulting from the translation is reviewed and refined from a sign language perspective. The transcription is carried out according to guidelines and criteria described in a manual. Tokens are assigned to types which are collected in a lexical database that is part of the iLex transcription environment (cf. Hanke/Storz, this volume). Each type is labelled by a unique gloss and its form is described by means of a HamNoSys notation. During the transcription, interesting or special passages are marked for further analysis.

At this stage of transcription, productive sign tokens as well as special signs (such as index, manual alphabet, etc.) are assigned to larger groups and coded only in a very broad way. For productive signs, such groups are determined by handshape/hand configuration and visualisation techniques (cf. Langer 2005).

##### 4.2.1 Lemma selection and management of detailed transcription

After the basic transcription, the data are accessible from a type-token perspective. This allows lemmas to be selected for the dictionary and stretches of text as well as single tokens to be assigned to further detailed transcription.

One major criterion for the inclusion of a sign into the dictionary is the frequency of its occurrence and its distribution among the pool of informants. Other criteria considered include votes from the focus group or the deaf community (see below: public voting). Taking into account the structural characteristics of signed languages and the current state of sign language lexicography, 6000 entries look like a realistic number for a comprehensive DGS sign language dictionary (cf. König, Konrad, Langer, in preparation).

For each selected sign, a linguist will review all occurrences and decide on whether all of them or just selected tokens have to be transcribed in more detail. For this decision the number of tokens, the preliminary assessment of the sign kind (i.e. invariant or highly modifiable) and the probable differentiation of its meanings are taken into consideration. Other reasons influencing the selection for detailed transcription are the suitability of utterances as dictionary examples, and instances of grammatical phenomena, which can serve as a basis for the compilation of the dictionary grammar. A detailed transcription allows for a closer look and well-founded analysis of these grammatical phenomena.

### 4.3 Detailed transcription

In the second transcription stage, selected occurrences of sign lemmas are transcribed again in more detail and with their surrounding context. We expect that approximately 50% of the basic transcriptions will be refined in more detail. The annotation and transcription of the corpus will be closely intertwined with the requirements of the lexical analysis needed for dictionary production. Not only a given occurrence, but also the surrounding utterance of a token is transcribed in order to be able to pin down the contextual meaning of the sign, collocations and other relevant context information. The following aspects are going to be coded:

- mouthing or mouth gesture,
- (lexical) facial expression,
- notation and classification of the form of the token,
- contextual meaning of the token,
- syntactic category,
- aspects of spatial use: establishing of spatial scenes, positioning of objects at special places in signing space and relating back to established places,
- productive elements for the visualisation of objects, processes, etc.

During the detailed transcription, utterances are segmented into smaller units (phrases, sign strings). Thus, sentence analysis regarding functions of single signs (e.g. illustrative function) is made possible. Interesting passages with regard to content or language are, again,

noted in a report. Short passages that can be well understood without further context are marked as potential example sentences or references for the dictionary.

Types are also further differentiated, their tokens are described more closely and classified, e.g. into

- phonological variants,
- forms, which are the result of grammatical processes, such as spatial orientation or positioning, marking of aspect, plural marking and others,
- forms, which result from sign modification, e.g. as a result of metaphoric use or of re-iconisation.

Productive signs are an important component of signed texts and have to be taken into consideration during the process of detailed transcription. Alongside the established lexicon, these have to be coded and annotated appropriately.

### 4.4 Transcription team

By opening the field of professional transcription work to students at the IDGS on such a large scale, we enter new ground. Both, basic and detailed transcriptions are mostly carried out by students. The students are trained and constantly supervised by experienced deaf transcribers. This procedure contains several advantages. In such a process, the transcriptions are continuously checked and verified by a team of experienced native signers, i.e. transcriptions are looked at by at least two persons working independently. This ensures a high quality of transcription without doubling the costs. For the students, we provide a first access to corpus linguistics and a first-hand insight into practical sign language lexicography, which is resumed by a long-term perspective via a PhD position. In addition, the combination of experienced deaf transcribers and students allows us to transcribe larger amounts of data within the timeframe of the project than by staff transcribers only.

iLex models the whole the transcription process, and especially supports the consistency of token-type matching. Transcribers are enabled to communicate with each other within the transcription environment by means of video chat and web 2.0 technologies, contributing to a steady flow of information within a rather large group of transcribers.

## 5. Data Sources

As outlined at the beginning of this paper, we rely on different sources to gather, verify and analyse language data, all of which complement each other to different degrees. Additionally, language knowledge and intuition of the deaf team members come in at all points. In the following, the four sources and their main functions are briefly introduced.

### 5.1 Published sign collections

For DGS there are many published collections of signs, differing in size, standard and degree of documentation (e.g. Metzger et al. 2000, 2003; 777 *Gebärden* 2002;

*Grundgebärden 1 & 2* 1999, 2000). Nevertheless, these are a valuable source for basic DGS vocabulary as it can be assumed that many of the published one-to-one translations of German basic vocabulary consist of agreed-upon signs, which are actually in use in the language community. The function of such sign collections in the project is to provide a pool of signs, which are evaluated by linguists and native signers to compile a preliminary basic vocabulary for DGS on this base.

## 5.2 Corpus of natural language

The corpus, which is elicited during the first phase of the project, serves as the major source to draw information from and is the very heart of the project. Lexemes can be identified and annotated with regard to the dictionary and the empirically based dictionary grammar. Every corpus, however, is limited. Not every topic can be covered in the elicitations, especially spontaneous conversation is unpredictable. Lexical gaps in the corpus, especially for everyday contents, can be searched for in the sign collections presumably with a high rate of success. Also, not all kinds of grammatical and lexical phenomena can be assumed to be contained in the corpus if not particularly asked for. Data from published sign collections, can be verified or discarded on grounds of corpus data to compare with.

## 5.3 Focus group

Additional elicitations later on in the project are costly and time consuming. For cases of doubt concerning, for example, the lexical status of a sign, matters of language change, different meanings or regional information, a focus group is installed. The focus group consists of approximately 20 deaf experts from different regions, which are trained and sensitised to linguistic questions. The members of the focus group are direct representatives of the language community and their active involvement is vital for the success of the project. Decisions such as those concerning lemma selection or the well-formedness of grammatical constructions will be continuously validated by the focus group. By falling back on the focus group, gaps in the corpus can be compensated for and additional surveys can generally be avoided. In some cases, it may still prove useful to conduct small additional surveys or elicitations to clarify certain questions or to supplement the data. Members of the focus group can then act as contact persons in their respective regions.

## 5.4 Public voting

A general voting web interface is established, which is open for all interested members of the deaf community. The focus group is very limited in terms of the number of members and, expectably, cannot answer all questions. The feedback from the focus group or partial results are still valuable as a starting point for public voting. Furthermore, the voting is useful for the validation of dialectal variants; variants that have not been included so far can be put in for discussion or for further voting by

deaf users, respectively. Signs which are only contained in the sign collections used for the basic vocabulary but not in the corpus can be validated by means of public voting as well.

Public voting allows a substantially larger part of the language community to directly participate in the project than possible through the informants, the focus group, and team members.

## 6. Analysis and composition of dictionary entries

The dictionary will be entirely based on the corpus with respect to the signs to be included as lemmas. Most examples of sign uses will be taken directly from the corpus. However, the information provided in the entries will decidedly exceed a conglomeration of just corpus references and examples. Rather, we will systematically abstract from the occurrences to obtain a generalized description of lexical items. This description will include following aspects:

- sign form (citation form)
- phonological variants
- use of space, morphology of sign (e.g. plural forms), modifiability of the sign
- word class, syntactic functions
- meaning: different readings; possible translations to German
- iconic value and visualisation technique; popular explanations as known in the sign community
- dialect information
- cross-references to related and similar sign, synonyms and antonyms
- examples to illustrate grammar, usage and different readings of the meaning: examples taken from the corpus or, where needed, invented examples, verified by the focus group.

After a first preliminary analysis, it is determined which additional information is needed for the final analysis and the composition of the entry. Additional detailed transcriptions are requested and material is given to the focus group and prepared for public voting. Once the detailed transcription of the requested sections is completed, large quantities of data can be compared and processed for the dictionary entry.

Not only is the structure of the lexicon analysed and described, but also grammatical phenomena are looked at more closely in order to compile a corpus-based dictionary grammar. To this aim, the filmed material from the tasks with different stimuli is used predominantly, but also grammatical constructions in fluent discourse (i.e. from the conversations) are examined.

## 7. Products

Several products will result from this project: a preliminary collection of basic DGS vocabulary, a large research corpus for DGS, a publicly accessible part of the corpus (supplied with English translation), the dictionary

grammar and a number of academic publications accompanying the work in progress.

### 7.1 Corpus-based electronic dictionary

The dictionary will be the first comprehensive corpus-based dictionary of German Sign Language. It is published in electronic form and primarily serves the following target groups:

- DGS learners who are native speakers of German, e.g. hearing people dealing with deaf people in work related contexts, parents and relatives of deaf children, students of Deaf Studies/sign language interpreting, and hearing impaired or late deafened adult learners of DGS as a second language,
- professional sign language interpreters for DGS ↔ German,
- native signers of DGS: deaf adults, children of deaf adults (CODAs),
- deaf children or pupils, acquiring DGS as native language,
- sign language teachers, linguists, language typologists and others concerned with sign language structure on a theoretical or practical level.

In order to be able to serve such diverse groups, the dictionary has to combine different types of dictionary functions to allow various kinds of uses. Lemma selection and lemma articles clearly focus on the DGS part. Like in a learner's dictionary, example sentences are included to illustrate sign uses and meanings. The German part first of all provides access to DGS for hearing users via their native language. Additionally, deaf users get basic information about German words and example sentences. It thus serves as a first starting point for the consultation of further references. The dictionary is going to contain an estimated number of 6000 sign entries.

### 7.2 Dictionary grammar

An important part of a comprehensive dictionary is the dictionary grammar. Regular grammatical properties of lexical items and formal paradigms do not have to be listed time and again in the entries. References to the dictionary grammar render the entry shorter, more compact and therefore make it more clearly. Learners also greatly benefit from a solid and well-founded grammar of the target language written in easily understandable terms. The dictionary grammar will be based on the corpus; issues, which cannot be resolved by solely looking at the corpus data, are given to the focus group or small additional elicitations have to be carried out. It will contain a general overview of the most important grammatical features of DGS, supplemented by examples taken from the corpus.

### 7.3 Annotated corpus

To comply with international corpus-linguistic objectives, a representative selection of about 50 hours of corpus material will be made publicly available online. To make

the data accessible for researchers who do not understand German, an English version of the existing German translation is provided for this part of the corpus. All annotated signs are additionally endowed with an English gloss. iLex provides formats for the interchange of transcripts and metadata without loss.

In the context of co-operations, the complete corpus will be available for linguists and PhD candidates, working on special issues in exchange for supplying additional annotations which can, in turn, be used by the project team.

### 7.4 Preliminary collection of basic vocabulary

In the fifth year of the project, a collection of basic DGS vocabulary will be published in an electronic version. This dictionary is preliminary since it is not based on corpus data. It is planned as a bilingual dictionary for DGS and German and should include the basic vocabulary of both languages. To compile this dictionary, already published sign collections will be compared and evaluated. Signs that are listed several times in different sources are likely to be included in the basic vocabulary, signs that have only one or few listings have to run through a verification process (by the focus group and the public voting) to avoid a listing of artificial signs and artefacts. A basic vocabulary list for German will be included, and missing sign equivalents will be provided by different methods such as small elicitations, public voting and input from the focus group. At the end of the project this product is replaced by the larger, fully corpus-based dictionary.

## 8. Co-operation

Besides elicitation settings, tasks and technical equipment mentioned above, we consider it essential to push standardisation or at least compatibility of annotation and transcription conventions to reach comparability of results across projects in cross-linguistic research. To this aim, we have arranged co-operations with other national corpus projects and look forward to co-operating with more projects currently in preparation. Topics like quality assurance or transcription efficiency will be best discussed in cross-project workshops.

## 9. References

- 777 *Gebärden 1-3. Alle 3 Folgen auf einer DVD* (2002). Version 2.0. Guxhagen: Manual Audio Devices. (DVD-ROM).
- Crasborn, O., Hanke, T. (2003b). *Additions to the IMDI metadata set for sign language corpora*. Agreements at an ECHO workshop, May 8–9, 2003, Nijmegen University. URL: [http://www.let.ru.nl/sign-lang/echo/docs/SignMetadata\\_Oct2003.pdf](http://www.let.ru.nl/sign-lang/echo/docs/SignMetadata_Oct2003.pdf) (last accessed March 2008).
- Grundgebärden 1. Für Einsteiger*. (1999). Hamburg: Verl. hörgeschädigte kinder. (CD-ROM).
- Grundgebärden 2*. (2000). Hamburg: Verl. hörgeschädigte kinder. (CD-ROM).

- IMDI (ISLE Metadata Initiative; 2003). *IMDI Metadata Elements for Session Descriptions, Version 3.0.4*, MPI Nijmegen. URL: [http://www.mpi.nl/IMDI/documents/Proposals/IMDI\\_MetaData\\_3.0.4.pdf](http://www.mpi.nl/IMDI/documents/Proposals/IMDI_MetaData_3.0.4.pdf) (last accessed March 2008).
- König, S. Konrad, R., Langer, G. (in preparation). Lexikon: Der Wortschatz der DGS. In H. Eichmann, J. Hessmann. *Einführung in die Gebärdensprachlinguistik*. Hamburg: Signum.
- Langer, G. (2005). Bilderzeugungstechniken in der Deutschen Gebärdensprache. In *Das Zeichen* 70, pp. 254--270.
- Lucas, C., Bayley, R., Valli, C. (2001). *Sociolinguistic Variation in American Sign Language*. Washington, DC: Gallaudet Univ. Press.
- Metzger, C., Schulmeister, R., Zienert, H. (2000). *Die Firma. Deutsche Gebärdensprache do it yourself*. Hamburg: Signum. (CD-ROM).
- Metzger, C., Schulmeister, R., Zienert, H. (2003). *Die Firma 2. Deutsche Gebärdensprache interaktiv. Aufbaukurs in Deutscher Gebärdensprache – Schwerpunkt Raumnutzung*. Hamburg: Signum. (CD-ROM).
- Schembri, A., Johnston, T. (2004). Sociolinguistic variation in Auslan (Australian Sign Language). A research project in progress. In *Deaf Worlds* 20 (1), 78-90.

## British Sign Language Corpus Project: Open Access Archives and the Observer's Paradox

Adam Schembri

Deafness Cognition and Language Research Centre, University College London  
49 Gordon Square, London, WC1H 0PD, United Kingdom  
E-mail: a.schembri@ucl.ac.uk

### Abstract

The British Sign Language Corpus Project is a new three-year project (2008-2010) that aims to create a machine-readable digital corpus of spontaneous and elicited British Sign Language (BSL) collected from deaf native signers and early learners across the United Kingdom. In the field of sign language studies, it represents a unique combination of methodology from variationist sociolinguistics and corpus linguistics. The project aims to conduct a studies of sociolinguistic variation, language change and language contact simultaneously with the creation of a corpus. As such the nature of the dataset to be collected will be guided by the need to create a judgement sample of the deaf community rather than a strictly representative sample. Although the recruitment of participants will be balanced for gender and age, it will focus only on signers exposed to BSL before the age of 7 years, and adult deaf native signers will be disproportionately represented. Signers will also be filmed in 8 key regions across the United Kingdom, with a minimum of 30 participants from each region. Furthermore, participant recruitment will rely on deaf community fieldworkers in each region, using a technique of 'network sampling' in which the local community member begins by recruiting people he or she knows, and asks these individuals to recommend other individuals matching the project criteria. Moreover, the data will be limited in terms of situational varieties, focusing mainly on conversational and interview data, together with narratives and some elicitation tasks. Unlike previous large-scale sociolinguistic projects, however, the dataset will be partly annotated and tagged using ELAN software, given metadata descriptions using IMDI tools, and will be archived and made accessible and searchable on-line. As such, we hope that it will become a standard reference and core data source for all researchers investigating BSL structure and use. This means, however, that, unlike previous sociolinguistic projects on ASL and Auslan, participants must consent to having the video data of their sign language use made public. This seems to put at risk the authenticity of the data collected, as signers may monitor their production more carefully than might otherwise occur. As the aim of variationist sociolinguistics is to study the vernacular variety (i.e., the variety adopted by speakers/signers when they are monitoring their style least closely), open-access archives thus may not always provide the best data source. While recognising that this concept of the vernacular represents an abstraction, we discuss the possibility of overcoming this problem by making some of the conversational data password protected for use by academic researchers only, while making other parts of the corpus publicly available as part of a dual access archive of BSL.

### Introduction

The British Sign Language Corpus Project (BSLCP) is a new three-year project (2008-2010) funded by the British Economic and Social Research Council that aims to create a machine-readable digital corpus of spontaneous and elicited British Sign Language (BSL) collected from deaf native signers and early childhood learners across the United Kingdom. Researchers at University College London are leading the project, with co-investigators based at Bangor University (Wales), Heriot-Watt University (Scotland), Queens University Belfast (Northern Ireland) and the University of Bristol (England). In the field of sign language studies, the BSLCP represents a unique combination of methodology from variationist sociolinguistics and corpus linguistics. The project aims to conduct a studies of sociolinguistic variation, language change and language contact simultaneously with the creation of a corpus. Unlike previous large-scale sociolinguistic projects, however, the dataset will be partly annotated and tagged using ELAN

software, given metadata descriptions using IMDI tools, and will be archived and made accessible and searchable on-line. As such, we hope that it will become a standard reference and core data source for all researchers investigating BSL structure and use.

### 2. BSLCP research questions

In order to exemplify the kinds of research questions that can be explored using a corpus-based approach to the study of BSL, we will undertake five specific studies. Of these, four studies will investigate sociolinguistic variation and change in (1) a phonological variable (e.g., variation in handshape, location, movement or the presence of the non-dominant hand in a specific subset of BSL signs) (2) a grammatical variable (e.g., variation and grammaticalisation in the use of agreeing/indicating verbs or in the morphosyntactic marking of tense, aspect and modality); (3) a set of lexical variables (we will focus on fifty vocabulary items known to exhibit regional variation or undergoing language change in BSL); and (4) bilingualism and language contact (e.g., variation in the

use of English mouthing in conversational BSL or possibly also variation in the use of contact signing). Lastly, there will be a study of lexical frequency in BSL. Based on an annotated subset of 100,000 lexical items, we will investigate which are the most frequent signs in BSL conversation, adapting the methodology successfully employed for a study of lexical frequency in New Zealand Sign Language (NZSL) by McKee and Kennedy (2006).

In the sociolinguistic studies, our aim will be to investigate how the variation in three target variables correlates with linguistic (e.g., the preceding or following segment in the case of phonological variable; the degree of obligatoriness in the case of the grammatical variable) and social factors (e.g., the signer's region, gender, age, language background and possibly the strength of their social network ties, socio-economic class and ethnicity). The specific target variables will be selected after a literature review and an initial viewing of the BSL data collected (the latter is necessary because the frequency of particular linguistic variables in BSL is unknown, and we will need to focus only on those variables for which our data will yield many examples). An additional aim of our project relates to cross-linguistic differences in sociolinguistic variation and language change, with specific comparative studies planned with both a closely related sign language (Australian Sign Language, or Auslan) and an unrelated sign language (American Sign Language, or ASL).

## 2. BSLCP Methodology

The studies of lexical frequency, sociolinguistic variation and language change in BSL (and all future studies which will draw upon this corpus) require that we collect and analyse data from a large representative sample of the British deaf community. In order to make possible cross-linguistic comparison, the methodology is similar to related studies undertaken on ASL (Lucas, Bayley & Valli, 2001), Auslan (Schembri, Johnston & Goswell, 2006; Schembri & Johnston, 2007) and NZSL (McKee, McKee & Major, 2006). Using a previously implemented research design will also allow us to identify and propose solutions to any potential problems in the project methodology.

### 2.1 Site selection

To obtain a representative sample of regional variation in BSL, at least eight sites will be necessary (sites in south-east, south-west, north-east, north-west and central England as well as at one site each in Northern Ireland, Wales and Scotland). These sites will need to be carefully selected so that they are as representative of the major regional varieties of BSL as possible. In addition to this, it is most likely that the site selection will involve a focus on large urban centres where the largest number of deaf individuals is concentrated and where large residential schools for deaf children are or were previously located. This will ensure that there are sufficient numbers of deaf people to make it possible to collect a sample that is balanced for gender, age, language background, and potentially other social factors such as strength of social

network ties, socio-economic class and ethnicity.

### 2.2 Participant selection

For this project, as has become standard in sociolinguistics research (Tagliamonte, 2006), we will recruit using a judgement sample of individuals from the British deaf community. Our aim is to recruit deaf native or near-native BSL signers who have lived in their local community (i.e., in the sites selected above) for at least ten years, and have thus had sufficient length of exposure to the local dialect of BSL (preference will be given to life long residents where possible). Thus, the participants will be both deaf individuals from deaf families who learned to sign natively in the home as well as deaf individuals who were exposed to sign language before age 7 by mixing with deaf peers in school. Similar numbers of participants will be recruited in four different age groups: (1) 18-35 years of age, (2) 35-50 years, (3) 51-70 years, and (4) 71 years or over. The division of participants into these age groups is partly motivated by changes in language policy in deaf education during the twentieth century (c.f., Lucas, Bayley & Valli, 2001).

The selection of our participants will also take gender, social class and possibly ethnicity into account. Gender and class differences in language use have been a major focus of research on sociolinguistic variation in spoken languages (e.g., Coates, 1986, Labov, 1990) and signed languages (Lucas, Bayley & Valli, 2001; Schembri, Johnston & Goswell, 2006). We will recruit equal numbers of male and female participants from both working and middle class backgrounds in order to determine whether gender and social class are important sociolinguistic factors in the British deaf community. We define 'working class' individuals as those who were employed in unskilled, semi-skilled or skilled manual jobs (e.g., labourer, factory worker, or plumber) or as semi-skilled non-manual workers (e.g., clerk). 'Middle class' participants are those, possibly with a university education, who worked in skilled non-manual jobs (e.g., BSL tutor) or in professional and/or managerial positions (e.g., manager of an interpreting service). Due to the fact that university education has only become widely accessible to deaf people in the UK following disability discrimination legislation enacted since the 1980s, we would not always rely on university qualifications as a defining part of our social class classification (this was a key criterion used in the study of sociolinguistic variation in ASL, see Lucas, Bayley & Valli, 2001). Strength of social network ties is another factor at work in systematic sociolinguistic variation. Research reported in Milroy (1987) showed the degree and nature of social contact between individuals influenced their language use. This is bound to be of great relevance to deaf communities where social networks are comparatively dense and multiplex. We will also investigate the possibility that a signer's ethnic background (e.g., White British, South Asian or Afro-Caribbean background) may be a relevant factor.

### 2.3 Recruitment

The participants will be recruited in each of the sites by deaf community fieldworkers who will all be deaf individuals living in the 8 target sites with knowledge of the local deaf community. They will be responsible for identifying and recruiting native or near-native BSL users who have lived in the community for at least ten years. Initially, participants already signed up to the DCAL participant database who match our selection criteria will be approached, but additional recruitment will need to be carried out in order to reach our target number of 30 participants at each site. A technique of ‘network sampling’ will be used, in which the fieldworker begins by recruiting people he or she knows, and then asks these individuals to recommend other individuals matching the project criteria (Milroy & Gordon, 2003). In this way, we hope to ensure that participants will be filmed in pairs consisting of two individuals who will already know each other. All participants will be paid for their time.

Based on our experience collecting data for the Auslan corpus project and on the success of a similar American project (Lucas, Bayley & Valli, 2001), we do not anticipate any problems with finding sufficient number of participants. Exact figures for the number of people in the British deaf community are not available, but it is likely to be somewhere between 20-60,000 people. If only 10% of this population are native signers or early childhood learners of BSL, then even the lowest estimate gives a total population of 2000 deaf signers who meet our criteria. As a result, finding 240 individuals to participate in our project should not be a problem.

It is likely, however, that we may not be able to attract sufficiently balanced numbers of participants from working class and middle class backgrounds from the four age groups listed above, or from different ethnic groups. For social class and ethnicity, we would either run analyses on smaller samples of the participant pool in which the numbers were balanced, or simply not include these factor groups in our analysis and concentrate our analysis on the other social factors. For age, we could simplify the age bands into two groups, younger (e.g., 18-45 years of age) versus older (e.g., 46-90 years of age), rather than analyse the data using the smaller subdivisions into four age groups (i.e., 18-35, 35-50, 51-70, 71 and over).

## 2.4 Data collection

The data to be collected will be of four types: (1) a set of elicited lexical data, (2) a set of elicited narrative and grammatical data, (3) thirty to forty-five minutes of free conversation, and (4) an interview.

The data in type (3) will be collected first. This will consist of at least thirty minutes of free conversation between the members of a particular dyad in order to collect data that is as naturalistic as possible. To ensure that the deaf participants do not adjust their signing to match the preferences of the researchers (who will not, in most cases, be members of their local deaf community) and to minimise influences from contact with spoken English (Lucas & Valli, 1992), none of the researchers

(hearing or deaf) will be present.

After this, data of type (4) will be collected. The field worker will interview the participants about their background, patterns of language use, degree of bilingualism in BSL and English, and attitudes to sign language and deafness.

Next, the data collection will focus on type (1) data. In this case, the field worker will show the participants a set of 50 flashcards to elicit their signs for selected concepts. The set of signs to be elicited will be in part based on earlier work on lexical variation in BSL (Woll, 1991; Brien, 1992). In previous projects, it has been shown that lexical variation was significant in particular semantic fields, such as signs for colour terms, days of the week, and numerals (Deuchar, 1981; Sutton-Spence & Woll, 1999). Our current research design will build on this work by exploring in more detail the correlation between this lexical variation and social factors such as age and region (as has been demonstrated for the use of fingerspelling in BSL, see Sutton-Spence, Woll & Allsop, 1990). In most cases, we will use pictures to elicit the target lexical items (e.g., a flashcard that is coloured green to elicit the sign for this concept), but where this is not possible, the signs will be elicited by the use of flashcards showing a picture together with the sign’s closest English equivalent (e.g., a card with the written English word ‘people’ on it in order to elicit the sign for this concept).

In separate data collection sessions, data of type (2) will be collected. We will use a mixture of tasks to elicit narratives and some of the key grammatical features of the language (e.g., agreement/indicating verbs, constituent order, and non-manual features), but do so indirectly, as the nature of the tasks will involve the participant describing aspects of the stimulus material to another signer, rather than the language features themselves. Some of the tasks will be selected from those used in the Auslan and Sign Language of the Netherlands (NGT) corpus projects (Johnston & Schembri, 2006). Using materials successfully employed on previous projects will ensure that we are able to elicit the kind of linguistic data we need for the purposes of this study, as well as facilitate cross-linguistic comparison.

For all data collection tasks, we will use up to four high definition digital video cameras on tripods so as to provide body length views of the individuals in each group as well as views from above of their use of signing space. Both the chairs and video cameras will be infixed positions so as to ensure that participants are in the centre of the frame at all times. In some cases as well (e.g., where natural light is insufficient), it may be necessary to use studio lighting to ensure the best images of the participant’s signed communication are captured.

## 3. Open-access corpus data and sociolinguistic variation

The BSL Corpus will represent the largest corpus of its kind when it is complete, involving data from 240 participants. As such, the corpus will lend itself naturally to large-scale investigations into sociolinguistic variation

and change in BSL, despite the fact that the data will be limited in terms of situational varieties (i.e., it includes mainly conversational and interview data, together with narratives and some elicitation tasks). Unlike previous large-scale sociolinguistic projects (e.g., Lucas, Bayley & Valli, 2001; Schembri, Johnston & Goswell, 2006; McKee, McKee & Major, 2006), however, the dataset will be archived and made accessible and searchable on-line. As such, we hope that it will become a standard reference and core data source for all researchers investigating BSL structure and use. This means, however, that, unlike previous sociolinguistic projects on ASL, Auslan and NZSL, participants must consent to having the video data of their sign language use made public. This seems to put at risk the authenticity of the data collected, as signers may monitor their production more carefully than might otherwise occur. As Tagliamonte (2006) explained, a specific aim of variationist sociolinguistics is to study the vernacular variety. Labov (1972) defined this as the variety adopted by speakers when they are monitoring their style least closely. The focus on the vernacular reflects the belief among sociolinguists that it is the most systematic variety, as it is assumed to be the variety that was acquired first and is the most free from self-conscious style-shifting or hypercorrection (Tagliamonte, 2006).

Thus, combining sociolinguistic methodology with corpus linguistics objectives creates a unique form of the observer's paradox. Although ideally we want to observe how deaf people use BSL with each other when they are not being observed, participants will in fact be filmed using four cameras, perhaps with lighting equipment, after they have filled in consent forms that make them fully aware that the aim of the project is to create an open-access on-line BSL corpus using the data collected. Thus, participants will not only be aware that their signing will be seen by researchers, but also potentially by anyone with a computer that has access to the internet! It is not yet clear how much this will cause signers to shift away from casual usage.

Some sociolinguists, however, argue that the concept of the vernacular represents an abstraction, claiming that all varieties of speech vary considerably in response to situational contexts. As such, "...the concept of an entirely natural speech event (or an entirely unnatural one) is untenable" (Milroy & Gordon, 2003: 50). Despite this, clearly we need to consider how best to adapt the variety of techniques sociolinguists use to overcome the observer's paradox, or at least to reduce its effects. As in other projects, we can ensure that participants are comfortable with each other and filmed in the most familiar surroundings possible (e.g., deaf clubs, offices of the British Deaf Association etc). We also will allow the participants to chat with each other about any topic in the free conversation session for a minimum of thirty minutes. This should allow time for the participants to relax in the presence of the cameras. We also need to investigate the possibility of making access to the conversational data, for example, password protected and for use by academic researchers only, while making other parts of the corpus

publicly available as part of a dual access archive of BSL. Participants would be informed at the outset that the conversational data will not be part of the open-access archive, and that researchers who wish to see it will have to fill in an online registration form that includes a confidentiality agreement. This may help to make participants feel more relaxed, and thus more likely to produce examples of vernacular BSL.

## Conclusion

Corpus-based approaches represent new territory for sociolinguists of both signed and spoken languages. As such, advances in digital video technology open up new possibilities for data sharing and research collaboration, but they also present new challenges. In this paper, I have outlined the methodology to be employed in the BSLCP, and attempted to anticipate some of the problems it may encounter, as well as suggest some possible solutions.

## 1. Acknowledgements

I would like to thank my co-investigators Kearsy Cormier, Margaret Deuchar, Frances Elton, Donall O Baoill, Rachel Sutton-Spence, Graham Turner and Bencie Woll for assistance in writing the grant proposal from which much of the content of this paper is drawn.

## 2. References

- Brien, D. (Ed.). (1992). *Dictionary of British Sign Language/English*. London: Faber & Faber.
- Coates, J. (1986). *Women, men and language*. London: Longman.
- Deuchar, M. (1981). Variation in British Sign Language. In B. Woll, J. G. Kyle & M. Deuchar (Eds.), *Perspectives on British Sign Language and Deafness* (pp. 109-119). London: Croom Helm.
- Johnston, T. & Schembri, A. (2006). Issues in the creation of a digital archive of a signed language. In: Barwick, L. & Thieberger, N. (Eds.), *Sustainable data from digital fieldwork*. (pp. 7-16). Sydney: University of Sydney Press.
- Labov, W. (1972). *Language in the Inner City*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1990). The intersection of sex and social class in the course of language change. *Language Variation and Change*, 2, 205-254.
- Lucas, C., & Valli, C. (1992). *Language contact in the American deaf community*. San Diego: Academic Press.
- Lucas, C., Bayley, R., & Valli, C. (2001). *Sociolinguistic Variation in American Sign Language*. Washington, DC: Gallaudet University Press.
- McKee, D., & Kennedy, G. (2006). The distribution of signs in New Zealand Sign Language. *Sign Language Studies*, 6(4), 372-390.
- McKee, D., McKee, R. & Major, G. 2007. Sociolinguistic variation in New Zealand Sign Language numerals. Paper presented at the *Ninth International Conference on Theoretical Issues in Sign Language Research*. Florianopolis, Brazil, December 6-9.
- Milroy, L. (1987). *Language and social networks*. Oxford:

Blackwell.

- Milroy, L., & Gordon, M. (2003). *Sociolinguistics: Method and interpretation*. Oxford: Blackwell.
- Schembri, A., & Johnston, T. (2007). Sociolinguistic Variation in the Use of Fingerspelling in Australian Sign Language (Auslan): A Pilot Study. *Sign Language Studies*, 7(3).
- Schembri, A., Johnston, T., & Goswell, D. (2006). NAME dropping: Location Variation in Australian Sign Language. In C. Lucas (Ed.), *Multilingualism and sign languages: From the great plains to Australia* (Vol. 12). Washington, DC: Gallaudet University Press.
- Sutton-Spence, R., & Woll, B. (1999). *The linguistics of British Sign Language: an introduction*. Cambridge, U.K.: Cambridge University Press.
- Sutton-Spence, R., Woll, B., & Allsop, L. (1990). Variation and recent change in fingerspelling in British Sign Language. *Language Variation and Change*, 2, 313-330.
- Tagliamonte, S. (2006). *Analysing sociolinguistic variation*. Cambridge: Cambridge University Press.
- Woll, B. (1991). Historical and comparative aspects of British Sign Language. In S. Gregory & G. M. Hartley (Eds.), *Constructing deafness*. London: Printer Publishers.

## **Tactile sign language corpora: capture and annotation issues.**

**Sandrine SCHWARTZ**

Paris 8 University, UMR 7023 CNRS

Université Paris 8

2, rue de la Liberté

93526 Saint-Denis cedex 02

FRANCE

E-mail: sandrine.schwartz@gmail.com

### **Abstract**

Sign language, being a visual-gestural language, can also be used tactually among or with deaf people who become blind. When this language is shared between people who are totally blind, non-manual features of signs are totally neutralised, resulting into a purely kinesthetic-gestural variant of sign language. This tactile modality of reception leads to adjustments impacting sign language pragmatics, as well as sign order and to a lesser extent, the way signs are formed. We aim to explore these phenomena by carrying out a systematic analysis of tactile sign language corpora.

Such a corpus has been filmed in 2006, involving six French deafblind informants, all of them using tactile sign language as their primary means of communication. A total of 14 hours of spontaneous discussions, free conversations or elicited data were captured by up to three digital cameras.

In order thoroughly to analyse our corpus, we need the help of a reliable annotation tool. After trying a couple of them, we decided to select Anvil, for its visual layout and flexibility, as well as its temporal granularity. We need a partition annotation system which allows us to create, rename or reorder tracks freely even while annotating. The first steps of our annotation will take us on the lanes of conversational analysis, using a mix of glosses and pragmatic occurrences, eventually to lead us on the more sinuous paths of a syntactic micro-analysis.

## **1. Tactile sign language**

### **1.1. The Deafblind community and its sign language**

Sign languages being visual-gestural languages, it obviously requires sight to be perceived. Nevertheless, some deaf signers may undergo a partial or total loss of sight, leading to an inability to use sign language efficiently. Most of these people suffer from Usher Syndrome Type 1 (a congenital profound deafness combined with an evolutive visual impairment due to retinitis pigmentosa), which means that their visual field gradually becomes tubular and eventually shrinks totally in the most severe cases. Other conditions may lead to deafblindness, which explains the heterogeneity of the people who use sign language tactually. By becoming blind, they do not lose their expressive skills, but need to adapt their reception modality of signs by placing their hands on the signers' hands in order to feel the manual characteristics of the signs. Touch doesn't allow to receive the same amount of information as sight does. It is a proximo-sensibility and it cannot convey as much information simultaneously. The tactile broadband, or rather "narrowband" requires a sequential and slow exploration of objects and above all a mental integration work to reconstruct the object image from all its properties gathered via tactile and kinesthetic inputs

(Hatwell, 2000). The specificity of the tactile perceptive field has some significant consequences when it comes to receiving sign language. Firstly, conversation in tactile sign language can only occur between two people (very rarely three, trilogues being quite awkward). Secondly, unlikely visual sign language whose expression and reception use two different channels, in tactile sign language, hands have to carry out both functions. And lastly, all non-manual features such as eye gaze and facial expressions become totally irrelevant when sign language is received by touch. Informal and professional interactions with tactile signers showed us that their communication is as efficient as the one we may notice among deaf-sighted signers, which means that deafblind people developed a set of adaptive strategies. Their tactile modality of reception leads to adjustments affecting sign language pragmatics, as well as sign order and to a lesser extent, the way signs are formed. We aim to explore these phenomena by carrying out a systematic analysis of sign language corpora.

### **1.2. State of the art**

Only very few documents describing tactile sign language are available. Two major linguistic studies were conducted in Sweden and in the USA, and we found a more in fields which are more or less related to linguistics. Johanna Mesch achieved a PhD dissertation about turn-taking and questions in conversations of deafblind

signers. For the purpose of her research, she captured dialogues in Swedish and Finnish Sign Languages involving deaf and deafblind signers and only deafblind signers, whom she asked to converse freely. She focused on reception styles (how hands are placed on the signers'), the way turn-taking is ruled and back-channel feedback on one hand, and on the other hand she analysed every question occurrences and their contexts (Mesch, 2000). Steven Collins and Karen Petronio led a contrastive study between visual and tactile forms of American Sign Language. They filmed free conversations between deafblind people and analysed the phonological, morphological and syntactic changes caused by the tactile receptive modality (Collins & Petronio, 1998)

## 2. Corpus collection

### 2.1. "Pre-corpus"

Only very scarce video of tactile French Sign Language being available, and none of them being suitable for a linguistic study, we had to test our first analysis on an original corpus. We seized the opportunity of a national gathering of deafblind people in Paris to film tactile communication involving deafblind signers. We aimed to film totally free conversation, in order to capture spontaneous strategies. Using only one shoulder camera, we could hardly master the environment while filming, which led to a forty-minute capture, but barely analysable for several reasons. Light conditions were very poor and exiguity of the places did not allow us to film the signers correctly. We tried to film everything instead of focusing on relevant moments, which resulted into a mixture of unfinished conversations, often interrupted by other people. Although we planned only to film dialogues between deafblind people, we ended up with most of our corpus involving deafblind and deaf-sighted people. We faced another unexpected difficulty when we started to analyse our data: while it is possible to show a video to deaf-sighted informants when we need their feedback regarding signs we are uncertain about, it is totally impossible with deafblind informants. This situation leaves us with data we hope to clarify with the help of sighted signers who are used to interact with the deafblind community. This unsuccessful experience provided us with great lessons to apply when filming the corpus currently use for our PhD dissertation and which I will now describe more deeply.

### 2.2. Corpus general settings

In 2006, six deafblind informants gathered for a week end of activities and socialisation in Poitiers, thanks to the support of CRESAM (Experimental Resources Centre for Children and Adults who are Deaf-Partially Sighted and Deafblind) who helped us with funding as well as a team

of 8 volunteers (professional educators and interpreters used to work with deafblind people). We organised a program of visits for the participants in order to motivate them and be able to express themselves about shared and not shared experiences. We split the group into two sub-groups, one visiting a brewery while the other one was remaining at the centre, discussing for the purpose of our research, the first group came back to the centre while the other group went to visit a nut oil manufacture. Then they told each other about their visit. The day after, the whole group was taken to the village market place, where they bought a few souvenirs and food. Once back to the centre, they discussed about what their experience at the market. The remaining of the corpus consists in free conversation, either in the laboratory, or during informal moments, after the meals.

### 2.3. Deafblind informants

Two women and four men, aged from 27 to 70 years old took part in our study. They are all legally deafblind, though two of them have some light and shapes perception. They all use tactile sign language everyday as their primary means of communication. They were all born profoundly deaf or very hard of hearing and gradually lost their sight between ages 12 and 37. While in the previous studies, nearly all the informants had Usher Syndrome type one, we deliberately chose to film deafblind people with other conditions (congenitally deaf and partially sighted, congenitally deaf and glaucoma in childhood), in order to respect the heterogeneity of profiles in the deafblind community. We know it may be a supplementary difficulty in our analysis, but we did really want to stick to the reality of tactile sign language users. Regarding communication, our informants learnt LSF (French Sign Language) between ages 4 and 37, without formal teaching, by socialising with deafblind peers or with hearing and deaf professionals. Among our informants, one person learnt tactile LSF directly, at age 37, without any prior knowledge of LSF in its visual form. Since they became deafblind, they use tactile LSF as their preferred face-to-face means of communication. For their written way of communicating, two are very comfortable with French that read Braille grade 1 and 2, one uses French in Braille grade 1, one can still read French in very big and bold fonts, and one doesn't know any French but uses pictograms based on LSF. Two of our informants are able to use speech when interacting with hearing people, one of them being used to speak and sign simultaneously, even when addressing a deafblind person.

### 2.4. Data capture

For conversational purposes, signers ideally sit or stand facing each other with the receiver's hands covering the signers', thereby hindering capture and visibility. After many attempts, we preferred to film in-vitro dyads with

three cameras (one focusing on each signer and one filming from above) as well as a few triads (one camera focusing on each signer). Nevertheless, we filmed part of our corpus with only one or two camera, depending on how many deafblind informants were conversing simultaneously, especially during informal moments and free conversation sessions.



Figure 1 : One camera on two signers



Figure 2 : Two cameras, one on each signer



Figure 3 : Two cameras, one front view and one rear view



Figure 4 : Triad, three cameras, one on each signer

Mesch (2000) considers that the ideal capture setting involves 3 cameras, yet, we used several configurations in

order to test their efficiency.



Figure 5: Optimal cameras position: one on each signer, one filming from above

While video seems to be the ideal medium to store a visual language, we wondered about the storage of a tactile form of sign language. Should we use digital gloves, motion capture device or even functional magnetic resonance imagery? Even though such modern technologies would be available and affordable, it would be too intrusive, if not invasive to be used on people who need their full sensory broadband, already narrowed by deafblindness, in order to communicate efficiently.

### 2.5. Type of data and annotation issues

At the end of the week end, we managed to collect a total of about 14 hours of videotaped conversations:

- Free conversations: in-vitro, imposed dyads (or triads), sitting face to face, time being controlled or not. (7h40min)
- Elicited data: in-vitro, imposed dyads, imposed subjects of conversation (talking about what they respectively visited or what they both experienced), time being controlled. (3h45min)
- Spontaneous conversations: in-vivo, sitting or standing, free choice of the dyads, no imposed subject of conversation, no control of time (3h50min)

We decided to capture a mix of in-vivo and in-vitro conversations in order to observe spontaneous conversational strategies in the first case and gathering more analysable data in the latter, thanks to a good mastering of our filming environment in the laboratory.



Figure 6: In-vivo conversation, after a meal, filmed with one camera



Figure 7: In-vitro data, filmed with 1,2 or 3 cameras in the laboratory

In order thoroughly to analyse our corpus, we need the help of a reliable annotation tool. After trying to understand and to use a couple of them, we decided to chose Anvil, for its visual layout and flexibility, as well as its granularity regarding time slots. I immediately knew that we would not need a parametric transcription as signs rarely change from visual to tactile sign language. Nevertheless, we chose a partition annotation system which allows us to create, rename or reorder tracks freely even while annotating. We are currently testing Anvil for our corpus annotation, and hope to have chose the appropriate tool. We are still open to other systems if they prove to offer more advantages than Anvil for our corpus.

### 3. Macro and micro analysis or tactile LSF

#### 3.1. Conversational analysis

One of tactile sign language's specificity is that it can only be used between two persons and very rarely three, but leading to an awkward and unbalanced communication. This dialogical constraint leads to conversational styles that are quite unique. While in any visual sign languages turn-taking is managed by visual clues, they become irrelevant with deafblind people and must be replaced by clues given by signers' hands vertical and horizontal positions and in the signing space. Regarding backchannel feedback, it does also exist among deafblind signers and is very important for the consistence of the conversation. It consists in an ingenuous system of fingers' tapping and pressing on the signers' hands as well as a kind of tactile nod. Another element we want to study through our corpus is the way deafblind people "listen" to tactile sign language: one hand, two hands symmetrically or asymmetrically? We haven't been able to find any research about conversational description applied to LSF, and only a very few concerning other national sign languages, which is why we will apply the Kerbrat-Orrechioni (1996) model of conversational analysis, designed for vocal languages.

#### 3.2. Signing without non-manual features?

Another dimension of tactile sign language we wish to explore is the way tactile modality impacts the language structures where non-manual features are necessary. When deafblind people talk, these non-manual features become totally irrelevant as their reception of sign language is purely tactile, ore more precisely kinesthetic. We aim to deal with this issue through contrastive methods: by comparing occurrences in tactile LSF and with the way they would be expressed by sighted signers. We will mainly focus on three topics: clauses types, space and iconicity. We already noticed that when a non-manual feature is necessary in visual LSF, there is always an alternative strategy in its tactile form, be it lexical or not. For example, the addition of other signs can lift up ambiguity, or of a tactile component like an increased muscular tension. Regarding iconicity and role shifting, according to Christian Cuxac's theory (2000), they are the core of sign language and they can be spotted by changes in eye gaze. As eye gaze is irrelevant for deafblind signers, how can they mark these structures boundaries? Do they still use these structures even if they cannot use eye gaze and facial expression?

## 4. Perspectives

Tactile LSF has to be rather linear because it cannot use multi-channel signs and it is received by a sequential sense of touch. But it doesn't make tactile LSF become a kind of signed French. By this research, we aim to show that deafblind signers developed a very clever variant of their national sign language, adapting its structure to the constraints of touch and inability to use non-manual features. Our corpus will be primordial to deal with these issues : that is why we need to opt for a powerful and reliable annotation system to be able to analyse tactile LSF with the best granularity.

## 5. References

- COLINS Steven & PETRONIO Karen (1998), *What happens in tactile ASL ?* in Pinky Extension & Eye Gaze – Language Use in Deaf Community, editor Ceil Lucas, Gallaudet University Press, Washington DC, pp 18-37.
- CUXAC Christian (2000), *La Langue des Signes Française (LSF) : les voies de l'iconicité* in Faits de Langues n°15/16, editions Ophrys, Paris
- HATWELL Yvette (2000), *Toucher pour Connaître, Psychologie Cognitive de la Perception Tactile Manuelle*, PUF, Paris.
- KERBRAT-ORECCHIONI Catherine (1996), *La Conversation*, Seuil, Paris.
- MESCH Johanna (2000), *Tactile Sign Language, Turn Taking in Signed Conversations of Deaf-Blind People*, International studies on sign language and communication of the deaf, volume 38.

## Building 3D French Sign Language lexicon

J r mie Segouat<sup>(1)(2)</sup>, Annelies Braffort<sup>(1)</sup>, Laurence Bolot<sup>(1)</sup>,  
Annick Choisier<sup>(1)</sup>, Michael Filhol<sup>(1)</sup>, Cyril Verrecchia<sup>(1)</sup>

<sup>(1)</sup> LIMSI-CNRS

BP 133, 91 403 ORSAY Cedex, France

<sup>(2)</sup> WebSourd

99, Route d'Espagne, B t. A, 31 100 Toulouse, France

E-mail: {forename.name}@limsi.fr

### Abstract

We present a methodology to build 3D French sign language (LSF) corpus of lexicon. These signs will be used in various software dealing with signing avatar. One software we are developing is a display information system in a railway station, to provide as information to deaf travellers as the hearing can get. Another one is a Web dictionary of LSF lexicon with particular entry ways. The third software is a LSF translation on our laboratory's Web site to provide accessibility for deaf users. Our aim is to build a set of signs so that one sign can be used in each software listed. In order to create this 3D LSF corpus lexicon, we set up a new methodology. We propose four steps to ensure the quality of the result: selection of the participants, meetings to elaborate the lexicon, elaboration of the video corpus that will be used for the 3D conception, creation of the 3D corpus by means of 3D software. We suggest the participation of the end-users from the corpus conception, the participation of experts and specialists for the corpus content, and a multi-level evaluation (on technical, use, ergonomic and linguistic sides) of the result. Finally we present which step we have achieved in each of our software project

### 1. Introduction

Our study is in the field of automatic French Sign Language (LSF) generation and 3D virtual animation of a signing avatar. Our aim is to generate LSF utterances by concatenation of isolated signs, which would be performed by the signing avatar. We want to set up a 3D LSF sign corpus creation methodology together with its evaluation within various software.

The interest of this work is to propose a 3D LSF animation methodology which could be generic whatever the software that will use it: these 3D signs could be re-used regardless of the end-software. We are not saying that our methodology is generic for corpus creation, we just focus on corpora whose aim is to generate LSF.

With our approach, one sign can be used in different software, to produce a large amount of utterances (we use LSF lexicon signs, but eventually we could generate productive signs thanks to: Bolot (2006), Ch telat-Pel  (2007), Filhol (2008).

In the next section (2) we introduce the notions of video and 3D SL corpora creation and how we plan to use the last one. Then, in section 3, we detail the four steps of our methodology: selection of the people who will participate to the corpus creation, conception of the corpus, building the video corpus, building the 3D corpus. In section 4 we present an evaluation methodology for the SL video and 3D corpus. Finally we present which step of our methodology we achieve for each software we are developing.

### 2. Video corpus and 3D corpus

There are two types of corpora: video and 3D. We elaborate the first one by filming a deaf person, following technical and linguistic specific criteria. These criteria

differ from one goal to another. The second corpus is created by using 3D animation software. The animations are created by using a video reference of each sign.

Video corpora<sup>1</sup> are built for research studies, especially by researchers in linguistics. Because those corpora are created with a specific goal, it has consequences on the type (in vivo or in vitro), the technical characteristics (how many cameras, which shot, etc.), the selection of the panel (how many signers, which level of language, etc.), the linguistic content (isolated signs or full utterances, narration or dialog, etc.) of the corpus. There is no generic methodology for creating SL video corpora. There is a standardisation (IMDI project) but it applies to the metadata: Crasborn, Hanke (2003). In this project, he structure of the information about the corpus is normalized but not the conception.

Concerning 3D LSF corpus creation, as far as we know there is neither a corpus nor a methodology to create one.

At LIMSI, we have created several video corpora for previous projects with linguists, for example in: ARC-LSF and LS-COLIN<sup>2</sup> (Braffort & al., 2001) projects, and TALS<sup>3</sup> (2005). More recently, we have built 3D LSF corpora to be used with a signing avatar: the goal of the first one is to create a system providing information in a railway station, and another one is used to experiment concatenation of isolated signs to generate a LSF sequence giving the date of the day (Figure 1).

The sign corpora (video and 3D) we just talked about were created with different technical characteristics, goals and people in the film. Those creations give us experience

<sup>1</sup><http://www.sign-lang.uni-hamburg.de/BibWeb/>

<http://www.bsllcorpuproject.org>

<http://www.bu.edu/asllrp/cslgr/>

<sup>2</sup><http://www.irit.fr/LS-COLIN>

<sup>3</sup><http://tals.limsi.fr/>

better to choose our methodology depending on the end-use of the corpus.



Figure 1: Video and 3D corpus

We intend to use our 3D LSF sign corpus in three types of different software:

- a railway station information system. As hearing people have access to the information given by a voice generated system, we aim to display a signing avatar on a screen to inform deaf people about general events (keeping an eye on luggage, etc.) and more specific ones (the delay of a train or a change in the platform number where a train will arrive, etc.);

- a bilingual LSF / French dictionary on the Web. We would like to provide Web users with a 3D sign dictionary to avoid the video drawbacks (non anonymous signer, impossible modifications, etc.) and to structure their display with several ways that are not yet implemented in other Web LSF dictionary (Moreau, 2007): by parameters, or by more complex properties like symmetry, re-localisation, etc. (Filhol, 2007);

- an ECA (Embodied Conversational Agents) on the Web site of our laboratory. It will interactively translate or explain in a LSF way specific words and concepts of our laboratory's research fields.

The software listed above is of different types in terms of

graphic user interface (isolated sign or utterances), target users (in a station, on a web site), and cognitive context: the goal of the users is different if they want information about a train, or about lexicon, or if they want a translation or an explanation on a word or a concept. Nevertheless, we want to use the same 3D signs corpus (numbers, for example, would be used in the three projects). In order to achieve this goal, we propose a methodology detailed in the next section.

### 3. Methodology

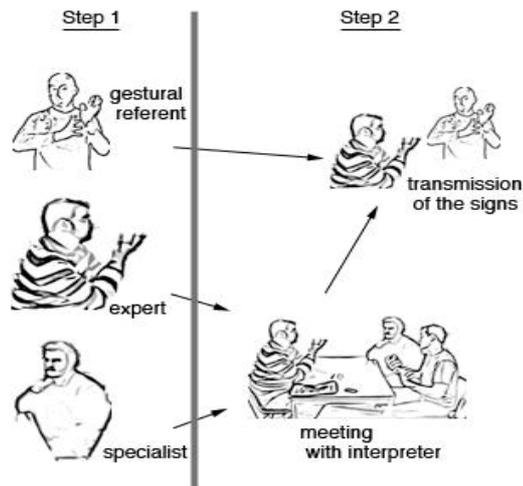


Figure 2: methodology 1/2

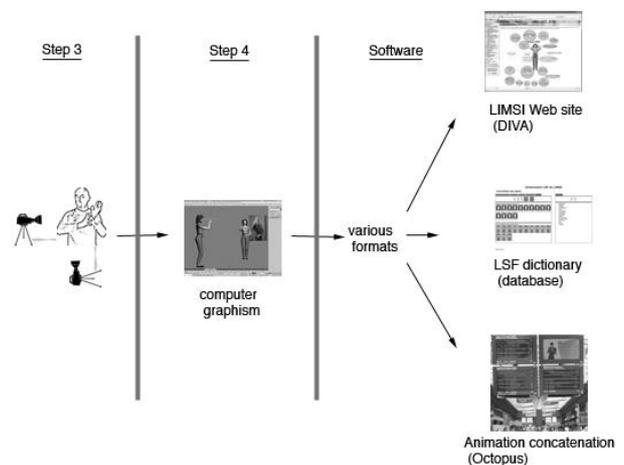


Figure 3: methodology 2/2

Our goal is to have a generic 3D LSF sign corpus. To build it, we need a video reference of each sign. The whole process is composed of four steps, which are represented in a graphical form in the figures 2 (steps 1 and 2) and 3 (steps 3 and 4).

1) The first step is to define who will be our « gestural referent ». In other words, how we select the person who will be filmed signing every LSF sign we need. This person must be a deaf and skilful user of LSF, expressing clearly and easily. The person is proposed by our team and validated by our computer graphist. The team's choice is

made with deaf and fluent people in LSF, and the computer graphist validates the choice depending on how it matches with the work he will have to do with the video. At this moment, we have two gestural referents, so we have to keep in mind that we will have to evaluate combining 3D signs build with video from both referent in terms of quality perception (and comprehension) of the message by the end-user. There remains a little difference in style even if the computer graphist, when creating 3D signs, makes some smoothening (see step 4). During a previous project where information messages were displayed in a railway station, end-users (travellers) easily recognized the person who was chosen to be our gestural referent and whose video signs were used to create the 3D animation. Comparing this previous project and the present one, the difference is that we now build a corpus of isolated signs, and not an utterance corpus. Combining isolated signs, even if they were created from a unique referent, with context and co-articulation influences, should erase this phenomenon. We'll see, with the first evaluation results if this hypothesis is confirmed.

At the same time, we contact « specialists » of each field we want to get lexicon of.

For the railway station project, lead by the SNCF (French national railway company), we have a partnership with WebSourd. We have to create messages similar to the audio announcements. For instance, « TGV number 1234 with destination Rennes, stopping in Le Mans and Laval, leaves at 8h56, from platform number 19 », or « due to an incident at level crossing, train number 35750, normally arriving at 13h01, will be delayed about 15 minutes ». SNCF provide all sentences that have to be displayed with the LSF 3D signing avatar. Then WebSourd provide a translation in LSF. We then dispose of the video corpus. In this project, the filmed gestural referent's is a LSF/French translator. Therefore, he is an expert of the two languages, working with his team to find the most accurate translation. He is used to be filmed because of his occupation, so he fully verifies our criteria to be a gestural referent. The video corpus is built by filming each sign isolated and in a sentence in step 3.

For the dictionary and the laboratory's Web site, in addition to the previous criteria, we needed a person with a significant general knowledge to express signs from a paper dictionary (IVT, containing signs of the everyday life) and a simple lexicon of our laboratory's Web pages (for the specific lexicon of the pages, we call on an « expert »). Consequently, we work with a deaf actor.

Still at the same time, we look for « experts » : deaf persons, skilful in a field for which we want some translations and/or explanations in LSF (if possible, all fields of our laboratory) or from connected fields (because it is not easy to find deaf people working in the same field as our laboratory).

2) We then begin the second step which is the conception of the video corpus. If the gestural referent has enough skills in the knowledge field he is going to express signs, we shoot directly, in step 3 (of course, the referent prepare the signs before being shot). Otherwise, we call the expert

and the specialist. The expert meets the specialist better to understand all the words or concepts he is not sure of. This meeting takes place with a LSF - French interpreter, so there could be an influence in some ways. If there is a discussion on a particular French word but there is no word-to-sign translation, the first solution used by the interpreter will be a circumlocution based on choices due to the translation principles. This influence could be erased partially if there is a discussion between the expert and the specialist: the expert can reformulate what he has understood and the specialist can validate or be more accurate rephrase to. Even is the influence is minimal, it exists and we have to keep it in mind while evaluating the users' answers on the evaluation time. The meeting is recorded on video tapes, stored for the expert, but also to take into account the influence of the interpreter's translation while evaluating the end-users' answers about the quality of the translation. Lastly, the expert meets the gestural referent to give him every sign that will be shot for the video corpus.

3) The third step is to film the video corpus. Following the computer graphist recommendations, we chose a double shot: side and front medium shots. Those two shots are edited together in a single video file given to our computer graphist.

4) He conducts the fourth and last step which is the 3D signs corpus creation based on the videos corpus. The creation is made by copying key frames of the video manually, with 3DSmax™. Those 3D animations are then converted in various formats, depending on which software will use them (Octopus<sup>4</sup>, DIVA<sup>5</sup>, database). We finally have our 3D LSF sign corpus.

#### 4. Evaluation

As far as we know, there are few studies on evaluation of video and 3D corpora evaluation. Huenerfauth (2007) proposes an ASL generation evaluation methodology, but his work is based upon classifier predicates (CP) generation, while ours at present focus on generating standard lexicon. CP's generation is evaluated through the user's choice of an animated sequence that translates what was displayed in ASL. The evaluation that Hunerfauth suggests is not immediately usable for our work, because his goal was to evaluate the rightness of the CP's choice and not lexicon animation. Nevertheless we will follow the generally accepted principles (like Huenerfauth does) of using questionnaires to get the users' opinions.

During the evaluation, we will gather information about the software and the 3D corpus. The evaluations will be different depending on the software's characteristics. For the information system in railway stations, we will propose a questionnaire in the station (a face to face interview in LSF) and a video questionnaire on a Web site. The answers will give us critics on ergonomic, technical, linguistic and use aspects. The panel of the users should be

<sup>4</sup><http://www.limsi.fr/Individu/bolot/octopus.html>

<sup>5</sup><http://www.limsi.fr/Individu/jps/online/diva/divahome/index.html>

composed of deaf users of the railway network and deaf specialists of LSF and information and communication technology (ICT). For the dictionary and the accessibility of our Web site, users will give their opinion by written or video email, and a video questionnaire will be proposed on the two Web sites. Questions will focus on ergonomic, relevance of the proposal entries and correctness of the signs (and of the translation), and, in particular for the Web site accessibility project, on the comprehensibility of the signing avatar's LSF and its overall ergonomic quality. We will ask users via a questionnaire on the Web sites and we will invite LSF and ICT specialists to test the system.

## 5. Conclusion

We want to dispose of a LSF 3D isolated signs corpus to be used within various software. We suggest a 3D sign creation methodology from a video corpus. The video corpus itself is built following a particular methodology: selection of a deaf gestural referent, possible selection of a deaf expert and a specialist, and selection of signs by the gestural referent or the expert after meeting with the specialist. This process is interesting because we will lastly have a set of 3D signs that every software can use according to the needs. At the beginning we will also have to propose validated signs by the gestural referent which rely on an existing paper dictionary, or by the expert. The validation will be iterated: signs will be displayed to the public, it will evaluate it, and we will modify the signs according to the public feedbacks, and will re-propose them to the public, and so on). This validation process will be the same for the signs and for the 3D creation methodology proposed.

The methodology we set up should guarantee us a significant relevance, thanks to the discussion between deaf experts and specialists of the field, and a good quality of the video signs, and 3D signs too, thanks to the selection criteria of the gestural referent. The evaluation step should guarantee maximal feedback about our methodology validation all along the four steps. What will be evaluated is the gestural referent selection (his style) for the first step, translation choices by the expert for the second, the quality of the video corpus for the third, and the quality of the 3D corpus for the last step. Of course, we will also evaluate the overall software.

We are at the first step of our methodology concerning the information system, at the second step for the accessibility of our laboratory's Web site, and we already have passed the third step for our dictionary project. The beginning evaluation step will confirm our methodological choices, and will give us ways to make it better.

## 6. References

Bolot, L. Braffort, A. Filhol, M. (2006). Elsi balbutie ! Vers une plateforme d'animation d'avatar signant. *Actes du 2ème Workshop Francophone sur les Agents Conversationnels Animés 2006 : WACA 2006*, Toulouse.

Braffort, A. Choisier, A. Collet, C. Cuxac, C. Dalle, P. Fuselier, I. Gherbi, R. Jausions, G. Jirou, G. Lejeune, F.

Lenseigne, B. Monteillard, N. Risler, A. Sallandre, M.-A. (2001). Projet LS-COLIN. Quel outil de notation pour quelle analyse de la LS ?. *Actes de Journée Recherches sur la langues des signes*, Toulouse.

Chételat-Pelé, E. Braffort, A. Véronis, J. (2007). Mise en place d'une méthodologie pour l'annotation de gestes non manuels. *Actes de TALS 2007. Traitement Automatique des Langues des Signes 2007 : atelier de Traitement Automatique des Langues Naturelles 2007*, Toulouse.

Crassborn, O. Hanke, T. (2003). Additions to the IMDI metadata set for sign language corpora. ECHO Project KUN, Taalwetenschap.

Filhol, M. Braffort, A. (2008). Description lexicale des signes. Intérêts linguistiques d'un modèle géométrique à dépendances. *Revue Traitement Automatique des Langues*, Volume 48 Numéro 2 : *Modélisation et traitement des langues des signes* (à paraître).

Huenerfauth, M. Zhou, L. Gu, E. Allbeck, J. (2007). Evaluating American Sign Language Generation Through the Participation of Native ASL Signers. *Ninth International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS-2007*, Tempe, Arizona, USA.

Moreau, C. Tranchant, C. and al., (2007). Modèle de recherche de signes standard en langue des signes française, *Traitement Automatique des Langues des Signes TALS 2007, TALN'07*, Toulouse, 5-8 juin 2007.

# Interface Development for Computer Assisted Sign Language Learning: Compact Version of CASLL

Saori Tanaka<sup>†</sup>, Yosuke Matsusaka<sup>‡</sup>, Kaoru Nakazono<sup>††</sup>

<sup>†</sup>MID  
Nonoshita, Nagareyama, Chiba,  
270-0135, Japan  
tanaka@mid-japan.com

<sup>‡</sup>National Institute of Advanced Industrial  
Science and Technology  
Umezono, Tsukuba, Ibaraki 305-8568, Japan  
yosuke.matsusaka@aist.go.jp

<sup>††</sup>NTT Network Innovation Laboratories  
3-9-11, Midori-cho, Musashino-shi, Tokyo,  
180-0012, Japan  
nakazono@core.ecl.net

## Abstract

In this paper, we introduce e-learning system called CASLL and demonstrate the small interface that is applicable for the different types of screen including the laptops and the other types of mobile movie players. In the existing learning program, users learn sign words and then try to select the appropriate Japanese translations in a natural conversation expressed by two native signers. In the proposed program, users try to segment each word from a stream of signing by manipulating a control knob on the bottom of a movie screen, and then do the same tasks in the existing learning model. Ten Japanese learners participated in the evaluation experiments. The mean accuracy rate of the proposed program was higher than that of the existing program. The result has indicated that focusing on transitional movements has an effect for learning JSL as a second-language. Although the segmentation learning method has been shown as an effective learning method, there were some technical problems. Some learners answered that they could not see each JSL movies at once by using their own laptops to conduct the learning programs. Therefore, we modified the interface to match the smallest interface as possible. This paper ends with a future plan for the evaluation of new interface.

## 1. Introduction

In order to present some contents of native sign language by using the technology of computer networks, a lot of research has been done all over the world in the field of educational engineering. In these works, there are two types for their aims; one is to support the deaf people themselves in order to fill the social gaps between the deaf people and the hearing people, and the other is to disseminate the knowledge of sign language in a society so that the circumstances around the deaf people are improved.

In the case of first group, the research of animation generation by natural language processing technology has been progressing in Greek Sign Language(Efthimiou et al., 2004). In Japan, based on the knowledge that the speed for playing JSL images depends on the level of proficiency for JSL, a system for playing JSL images in five speed level has been developed(Isono et al., 2006).

Meanwhile, in the case of the second type of research, a remote communications system to connect a class of American Sign Language to the terminal of students has been developed(Lehman and Conceicao, 2001). In Japan, the self learning system of finger spelling with the function of feedback(Tabata et al., 2001), the JSL database with search function based on the linguistic knowledge of native Japanese signers(Fukuda, 2005) and the teachware of JSL have been developed(for JSL Learning, ).

As one of our series of previous studies for developing human interface by using Japanese Sign Language (JSL) contents(Tanaka et al., 2007b; Nakazono and Tanaka, 2008; Tanaka et al., 2008), we proposed a new learning program and compare it with the existing learning program implemented in the Computer Assisted Sign Language Learning (CASLL) system(Tanaka et al., 2007a). Those studies can be categorised in the second type of research.

## 2. Goal of the paper

In this paper first, we introduce a more effective learning method compared to the existing learning program implemented in the CASLL system. Secondly, we define the size of movie screen as small as possible, and develop user-friendly interface that is applicable for the different sizes of screen.

## 3. Design of segmentation learning

In this section, we will sort out the problems of the existing learning style compared to the characteristics of continuous sign in natural discourse.

In natural discourse by native signers, it is observed that hand-shape and position changes under the influence of those in the proceeding word, while it didn't occur when each word is carefully pronounced (Fukuda et al., 1961)<sup>1</sup>. It is also reported that hand-shape slightly changes its feature when two words form a phrase (Ichida, 2005). When a signer produces number "1" as a single word, he/she holds the other four fingers tight except for the index finger. However, when it forms a phrase with the word which five fingers open softly, the other four fingers except the index finger are closed lightly. This fact is caused by the fact that the hand shape with five fingers open effects the anteroposterior word, and then the feature of "open" is added to the four fingers except the index finger. This phenomenon seems to be equivalent to the phoneme assimilation in spoken language, the term of "assimilation" is used in sign language research(Ichida, 2005). The phenomenon is related to efficient articulation and is often observed in natural speed discourse in native signers.

Meanwhile in the learning situation of beginners, there are many students who face difficulty reading words in natu-

<sup>1</sup>In American Sign Language, it is also observed that the position of word is getting closer to the proceeding word position when the speed of articulation is increased(Claude, 2003)

ral discourse, even though they have already learned each word. It is not uncommon to see the number of students are decreasing by more than half in a class, as the levels of a course advances. One of the reasons for this problem seems to be the general learning method used in classes or groups of JSL. In some classes, students are often required to memorize all the vocabulary for greetings, introducing their family, and numbers. After the lesson, they start to practice to read and express the words with students in pairs. By such kinds of learning style, it is difficult to read sentences with many phoneme assimilations in natural discourse.

Based on these problems, we defined three points as necessary functions for the design of proposed learning program; (1) students can check the images as many times as they want to, (2) students can learn a word expressed carefully by comparing it to the same word expressed with other words continuously and (3) students can be aware of the segment boundaries including transitional movement.

We also implement the word-learning program that users mainly study JSL vocabulary on CASLL, and then evaluate how much each user can recognize the words in natural discourse. Based on the results, we verify the effectiveness of proposed program including the segment task before the vocabulary task.

Fig.1 shows the relations in each leaning program and continuous sign images in natural discourse. In the word learning that is more conventional in JSL education, users mainly focuses on the lexical movement without transitional movement in the segment: the short arrows in fig.1. Compared with this, we design the segmentation learning which makes users focus on the long arrow in fig.1. We expect that the users can study sign production including the natural assimilations between words, and it will improve their skill for reading as a result. We will describe the details in each learning program.

#### 4. Two learning programs and system

Table.1 show the process in each learning program. The texts in boldface are common tasks in each one. In the segmentation learning, there are three different tasks before the common tasks. Those three tasks are compartments of the segmentation learning, the proposed method in this study. In the word learning, users study only lexical movements without paying attention to transitional movements, and then memorize the Japanese translations. After that, users challenge the reading task to answer the meaning of word in each question. On the other hand, in the segmentation learning, the users count the stroke of movements in each example of natural discourse. This task is aimed at making them focus attention in each component of movement, and they should become used to the natural speed with many assimilations. After the counting task, they try to correct the order of word images which are randomly-aligned in the order of natural discourse. In the “word images”, a signer produces each word separately. At the upper side of the word image, there is discourse image in which a native signer is signing (Fig.2). The re-alignment task is aimed at relating the words and the sequential signs. The users should be conscious of the difference between the words

Table 1: Learning Processes of Two Learning Programs

Word Learning Program		Segmentation Learning Program	
1	<b>Word Learning Task</b>	1	Counting Task
2	<b>Word Test</b>	2	Realignment Task
3	<b>Reading Task</b>	3	Segment Task
4	Realignment Task	4	Word Learning Task
5	Translation Test	5	Word Test
		6	Reading Task
		7	Translation Test



Figure 2: User's Screen for Reorder Task

and the segments including transitional movement and lexical movement. After the re-realignment task, users perform the segment task in which they try to match the knob at each segment in the short piece of discourse image (Fig.3). We designed these three tasks so that users can be conscious of each segment including transitional movement. After these three tasks, they perform the tasks for the word learning: the word learning task, and the examination for checking how many word they correctly memorized. In the following section, we will explain about the detail of each learning program.

#### 4.1. Outline of two learning programs

##### • Word Learning Program

WRONG ANSWER (The first and second control knob are in wrong positions)

CASLL - 15/247

Please control the slider and move the knobs at the end of words



Figure 3: User's Screen for Segmentation Task

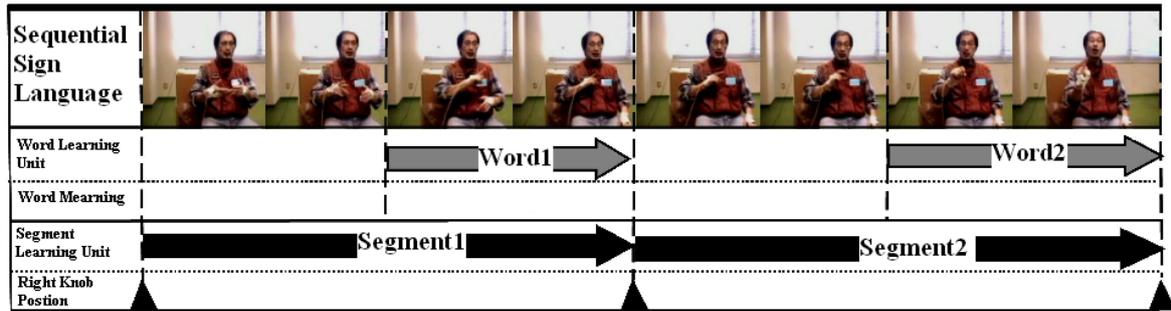


Figure 1: Learning Units for Segmentation and Word Learning Programs

### 1. Word learning task

Users are required to check each word image in a example sentence, and follow those movements. Each word image has a short expository text for the meaning and the movement.

### 2. Word test

Users are required to select the correct Japanese translation for each which is randomly-presented word. This test is aimed at testing whether they can remember the meaning of each word presented in the word learning task.

### 3. Reading task

Users are required to read each word presented in the image of natural discourse which length of about 5-10 seconds and select the meaning of it.

### 4. Realignment task

Users are required to correct the order of some word images that are randomly-aligned from the order of a sentence in natural discourse. Each image has Japanese subtitles.

### 5. Translation test

Users are required to select a correct translation that is appropriate for the example image by a native signer.

## • Segmentation Learning Program

### 1. Counting task

Users are required to count the stroke of movements in each example of natural discourse. They can select the number of stroke from choices.

### 2. Realignment task

Users are required to correct the order of some word images that are randomly-aligned from the order of a sentence in natural discourse. In order to match the order of word images to the sentence in discourse image, they need to manipulate the buttons for switching the images without Japanese subtitles. At this point, they have not learned the meaning of each word, so they can only check how each stroke structures the word in natural discourse.

### 3. Segment task

Users are required to select the length of each segment in a short piece of image cut from natural discourse. As in Fig.3, they need to manipulate the knobs on the slider with referring to the discourse image, and set the position of knob at at the last frame when the hands start to change the shape and translate to the next word. If the user doesn't set it at the right position, the arrow message is shown up saying "you have wrong answer (the second knob is at the wrong position)". After the user corrects the position, he/she can move to the next task<sup>2</sup>.

### 4. Word learning task

Users are required to check each word image in a example sentence, and follow those movements. Each word image has a short expository text for the meaning and the movement.

### 5. Vocabulary test

Users are required to select the correct Japanese translation for each word which is randomly-presented. This test is aimed at testing whether they can remember the meaning of each word presented in the word learning task.

### 6. Reading task

Users are required to read each word presented in the image of natural discourse that length is about 5-10 seconds and select the meaning of it.

### 7. Translation test

Users are required to select a correct translation that is appropriate for the example image by a native signer.

## 5. System overview

Fig.4 shows the system overview of CASLL.

Once the scripts with the procedure of two learning programs has been sent to the server, the server finishes the preparation for judging the answers; generation of hints; and generation of the forms of buttons, knobs, and images.

<sup>2</sup>The correct answer was set with the 2 frames margin in the vicinity.

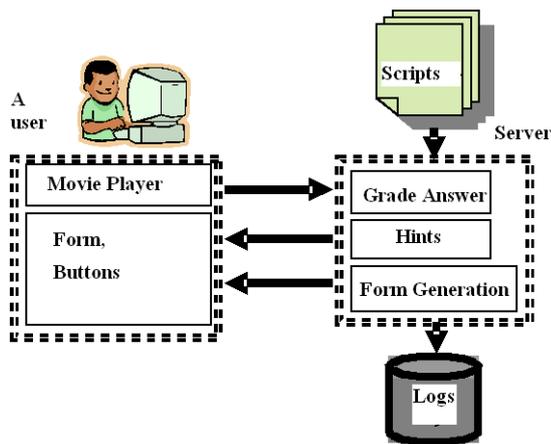


Figure 4: System Overview

Once the user finished each task with displaying images from the server, each answer will sent to the server. The server will display hints according to the user's answer, and present the form of next task when the user's answer is correct. The hints include the information about the number of correct answers per question, and the wrong position number of the knobs in the segment task. The user's record will be stored in the server in order to evaluate for the next analysis. We designed CASLL based on the assumption that users have Internet Explore 6.0+ (Microsoft), QuickTime 7+ (Apple), and ADSL or cable or large-scale network at universities. From the experiments that we have done before, the users can perform the tasks in CASLL with these environments.

## 6. Evaluation experiment

### 6.1. Procedure

The aim of this experiment is to present a self-learning program by which a user can be thoughtful about his/her choice of tasks. For this aim, we compare the scores of two subject groups; the word learning group and the segment learning group. We evaluate the effect of the segment learning for the reading tasks in CASLL. The users can see the question images as much as they want and perform the tasks until they reach each correct answer. In this situation, if the learning effects of two learning programs for the reading task were not so different, we could not say that there was a special learning effect in the segment learning program. If so, the effect of the reading task would be only the result of the word learning. Therefore, for this experiments, we will analyze the proficiencies of vocabulary in two programs. If we could confirm that the proficiencies are high enough in both two programs, we would use it as a baseline and then compare the scores of reading task in both programs. For this comparison, we would use the percentage of number of questions that the users get correct at the first trial, and the number of trials for reaching each correct answer.

### 6.2. Subjects

The subjects were 10 beginners who have less than two years learning experience at some JSL learning clubs. In

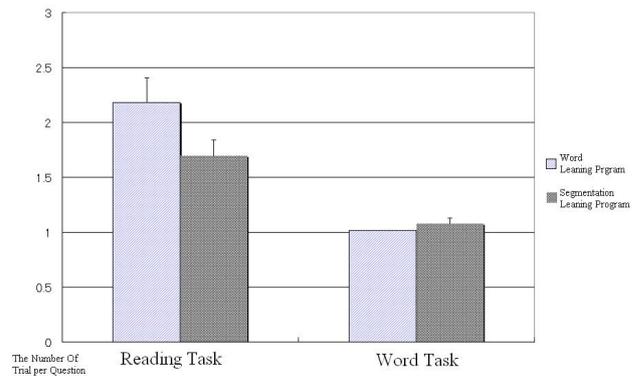


Figure 5: Mean Number of Trials for Reading Task and Word Task

these subjects, 5 beginners performed the word learning program and the other 5 beginners performed the segment learning program. The range of their age was from 20 to 38 years old, and the male-female ratio was 3:2 respectively.

### 6.3. Learning program

Although the learning program at the server can be restructured by writing the script, we used some images cut from a image of natural discourse. The length of each image was about 7 seconds. The total number of images were 6, and those included 53 basic words that were used in a natural discourse. The length of whole image was 49 seconds and the average speed of signing was 1.54 [word/sec]. That speed was quite fast for the users based on the results of Questionnaire that all users answered after the experiment. For those 6 images, we made the script for the two learning programs in Table.1, and implemented the tasks by which each user perform the 6 trials. We also added a function by which the users could answer each question 5 times and after 5 times, they could select "give up" and go next task.

### 6.4. Result and discussion

#### 6.4.1. Level of proficiency for vocabulary

$$\left( \frac{\text{the number of question} - \text{the number of wrong answer}}{\text{the number of question}} \right) \times 100 \quad (1)$$

From the expression (4.1), we calculated the proficiencies of 53 words in both learning programs. These were 98 in both learning programs that were close to perfect. As a result, it was confirmed that the proficiencies of the vocabulary in both learning groups were similarly high.

#### 6.4.2. Analysis for score in reading task

In order to calculate the score of the reading task, we substituted the number of reading questions and the number of wrong answer in the reading task to (4.1). As a result, the average in the word learning group was 67.0 and that in the segment learning group was 54.8. The former average was significantly higher than the later one ( $t(8)=1.97, p<.05$ ).

## 6.5. Analysis for number of trial

Although this system allows the users to skip the question with the giving up function after the fifth trials, they can also answer the questions as many as they want without the giving up function. Therefore, at least in principle, they can continue the trials to answer a question until they reach the correct answer. If he/she answers correctly at the first trial, he/she can have higher score, so we analyzed the number of trials they had in both the reading task and the word task. If the segmentation learning had more effect for the reading task than the word learning, the average number of trials of segmentation users would be less than that of the word users. The result was shown in Fig.5.

In Fig.5, the two bars on the left side shows the result of reading task. The mean trial number of reading task in the segmentation users was less than that in the word users. This means that the segmentation users reached correct answers with less trials than those by the word users. The mean trial numbers of reading task were 2.18 in the word learning group, and 1.69 in the segmentation learning group; SD were 0.49 and 0.32 respectively ( $t(8)1.84$ ,  $p=.051$ ). The less trial numbers needed by the segmentation users seems to reflect the higher score in the reading task.

In Fig.5, the two bars on the left side show the result of word task. Both groups of users reached the correct answers mostly at each first trial. The mean trial numbers of word task were 1.02 in the word learning group, and 1.07 in the segmentation learning group; SD were 0 and 0.11 respectively. From this result, we found that all user in both learning groups answered almost perfectly for the random questions in the word task.

The difference of learning program in both learning groups were the following; the segmentation users had the counting task, the realignment task, the segment task, the word learning task, the vocabulary test and then the reading task. The first three tasks are the original component of the segmentation learning. On the other hand, the word users had the word learning task, the vocabulary test and then the reading task. The word users seems to find it difficult to read the meaning of each word in natural discourse immediately after the word learning task and the vocabulary test. Additionally, the word users had a perfect vocabulary skill in 53 words, so even if they had more trials for the word learning task, the score of the reading task would be same. Compared with them, the segmentation users saw the natural discourse image over and over again, and had the counting task, the realignment task and the segment task, so the users seems to find it easy to recognize the relation between movement and meaning in the early stages. From the fact that the levels of proficiency for vocabulary are highly enough in both learning programs, the segmentation learning showed the effectiveness in reading the signing words in natural discourse.

### 6.5.1. Questionnaire

From the above analyses, we found the effectiveness of the segmentation learning, but there might be room for improvement in the program which was a first edition. We studied the possible way for improvement from the analysis

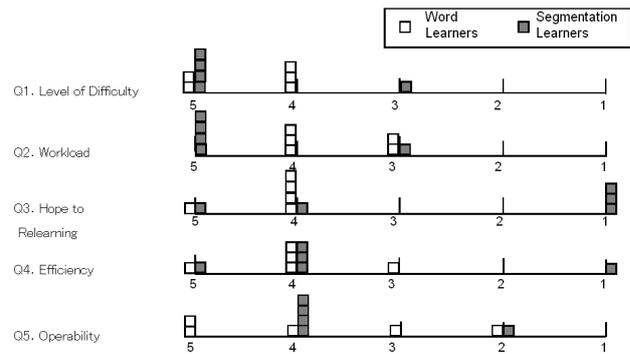


Figure 6: Results of Questionnaire

of questionnaires. Fig.6 shows the result of questionnaires that all user answered after their learning programs. They were asked to rate their impression based on five questions from Q.1 to Q.2 on the grade of five scale each. Q.1 was “Were the questions difficult for you?” and almost all user rated more than 4 (difficult). It indicated that the challenge levels were high for all user in both programs. Q.2 was “How did you feel about your workload?”. For this question, two word users rated 3 (right quantity) and three rated 4 (it was heavy). On the other hand, one segmentation users rated 3 and four rated 5(it was too heavy). The extra duration of the segmentation leaning seems to make the users feel workload to be heavier. Q.3 was “Do you want to learn by this same learning program again?”. Four word users rated 4 (yes, when I have time), and one rated 5 (definitely yes), while three segmentation users rated 1 (no, never), one rated 4 (yes, when I have time) and one rated 5(definitely yes). The evaluation result varied greatly in the segmentation learning. Q.4 was “How was the efficiency of this program compared to other learning methods by which you had learned before”. One segmentation user rated 1 (highly inefficient) but more than half of users answered “efficient” and “highly efficient”. Q.5 was “How was the operability of your computer for this program?”, and more than half users rated “good” and “very handy”. The user who answered “a little bit difficult to operate” described her impressions that her computer screen was too small to see all word image when she attempted the realignment task. From the above results of operability, we found that the problems for the interface would be improved easily if we changed it to the adjustable display system with user’s screen size. Additionally, we found some room for improvement in the curriculum of the segmentation learning. For the purpose of comparative experiments, we needed to design the segmentation learning program with a number of questions, but the evaluation of Q.3 seems to be improved if we control the number of trials that the segmentation users study at once. Furthermore, presenting a kinder explanation of each of task and meaning at the beginning of the program would be help to reduce the user’s mental burden for the number of questions.

## 7. Interface for Small Screen

### 7.1. Requirements

As we described in the previous section, our interface for *realignment task* has the problem to overview the words. In order to realize learning with mobile device which has small screen display, we have to squeeze the interface to more smaller size. Although, at the same time, because we have to keep learning efficiency and usability, the interface have to satisfy the following requirements.

**The interface has to be in small size:** As in our main objective.

**The user is easy to overview visual words at a glance:** Words in sign language is a visual information, and visual information requires certain size on the screen to tell correct information to the user. On designing the interface, we have to care not to make overlaps between the videos nor show the video in too small size to recognize sign language.

**The user is easy to pickup words using small interaction devices:** Because the user have to interact with the interface using through small and unprecise interaction devices (e.g. touch panel of the mobile phone), the word selection has to be done without any precise control.

### 7.2. Design

After a concern on the above requirements, we have made following 3 prototype designs.

**Type 1:** Original interface with a scroll bar. Same as original interface described in previous study with a scroll bar to scroll the screen to browse the words went over the screen (Fig. 2).

**Type 2:** Overlapped words. Overlap the 1/2 size of the words with the other words. When the mouse overlaps a word, the word will be zoomed and the user can see the word without overlapping (Fig. 7).

**Type 3:** Slanted words. The words are slanted and shown in small view without overlapping. When the mouse overlaps a word, the word will be zoomed and justified, and the user can see the word without slanting (Fig. 8).

### 7.3. Implementation

The movie is 320 x 240 pixels and encoded in H.264 codec. The interface is implemented using Adobe Flex3 SDK<sup>3</sup> with papervision3d library<sup>4</sup>.

Because the interface runs on *flash* platform, user can open the interface on their any preferred web browser.

### 7.4. Evaluation Plan

Alghouth we have just started to develop the new interface and classify the visual styles, we are going to evaluate the usability of the new types of interface from the view point of mental workload. As the one of subjective workload assesment methods, NASA-TLX has been developed by NASA (National Aeronautics and Space Administration) Ames Research Center(Hart and Bittner, 1988).

The evaluation procedure consisted of the following three steps.

<sup>3</sup>Adobe Flex3 SDK:<http://www.adobe.com/products/flex/sdk>

<sup>4</sup>papervision3d library: <http://papervision3d.org/>



Figure 7: Overlap the 1/2 size of the words with the other words. When the mouse overlaps a word, the word will be zoomed and the user can see the word without overlapping(Type2).



Figure 8: The words are slanted and shown in small view without overlapping. When the mouse overlaps a word, the word will be zoomed and justified, and the user can see the word without slanting(Type3).

1. Evaluation of workloads on six subscales: Evaluate the workload on six subscales using values from 0 to 100
2. Comparison of subscale weights: Evaluate the weights of the subscales by the pair comparison method
3. Unification of workloads: Calculate the weighted workload (WWL) by taking the average of the weighted values of the subscales

The six subscales are:

1. Mental Demand
2. Physical Demand
3. Temporal Demand

4. Performance
5. Effort
6. Frustration

The one of the authors has developed the JSL version of NASA-TLX (Nakazono et al., 2006). The original NASA-TLX, the explanations of the procedure and evaluation items are written in English. Descriptions in the original NASA-TLX were so stiff and technical that it was assumed to be hard to understand them completely and carry out fair assessments for ordinary people. With this JSL version of NASA-TLX, communication quality of intermediated on-line conversation by interpreter (Nakazono and Tanaka, 2008) have been already evaluated. We will use this methodology for the evaluation of new interface.

## 8. Conclusion

In this paper, we introduced an effective learning method implemented in the e-learning system called CASLL. We also demonstrated the small types of interface which are applicable for the mobile movie players. We will evaluate the usability of the new types of interface by JSL version of NASA-TLX. The most usable interface would be applicable for the website that assumed users have different sizes of screen, cell-phones and other types of mobile movie players.

## 9. References

- M. Claude. 2003. *Undershoot in two modalities: Evidence from fast speech and fast signing*. Ph.D. thesis, Univ. of Texas.
- E. Efthimiou, G. Sapountzaki, K. Karpouzis, and S-E. Fotinea, 2004. *Developing an e-Learning platform for the Greek Sign Language*, pages 1107–1113. Springer, 3118 edition.
- Compass: Video for JSL Learning. <http://www.inform-inc.co.jp/video.htm>.
- Y. Fukuda, H. Kimura, Y. Ichida, K. Fukushima, Y. Izumi, T. Sekine, K. Suzuki, N. Nakajima, A. Kazugai, and M. Norimatsu. 1961. Expression through hand and finger actions used in " words in continuous signing " : A study in reference to the dialogue in Japanese Sign Language by the deaf persons (in Japanese). *IEICE technical report. Speech*, 97(587):pp.81–85.
- F. Fukuda. 2005. Compilations of the electronic dictionary of Japanese Sign Language (a second edition) and its instruction manual (in Japanese). *IEICE technical report. Welfare Information technology*, (66):pp.39–44.
- J. Hart and A. Bittner. 1988. Development of NASA-TLX: Results of empirical and theoretical research in human mental workload. *British Journal of Social Psychology*, pages pp.139–183.
- Y. Ichida. 2005. Phonology and numeral system in jsl (in Japanese). *Sign Language Communication*, (57):pp.8–16.
- H. Isono, Y. Takiguchi, M. Katsumata, and M. Nakama. 2006. Preferred reproduction speed of sign language image and audiovisual system via the internet (in Japanese). pages pp.5–8.
- R. Lehman and S. Conceicao. 2001. Involving the deaf community in distance learning, using blended technologies and learning objects. *IEEE Electronic Journal*, pages pp.3–4.
- K. Nakazono and S. Tanaka. 2008. Study of spatial configurations of equipment for online sign interpretation service. *IEICE - Transactions on Information and Systems archive*, (6).
- K. Nakazono, Y. Takeda, M. Terauchi, and Y. Nagashima. 2006. Evaluation of mental workload of delayed visual communication. *TR of IEICE WIT2006-49*, 106(285):95.
- K. Tabata, T. Kurota, M. Murakami, Y. Manabe, and K. Chihara. 2001. Prototype design for sign language education system (in Japanese). In *Proceedings of Japanese Association of Sign Linguistics*, number 27, pages pp.34–35.
- S. Tanaka, Y. Matsusaka, and K. Uehara. 2007a. Segmentation learning method as a proposal for sign language e-learning (in Japanese). *Human Interface*, 9(2):61–70.
- S. Tanaka, K. Nakazono, M. Nishida, Y. Horiuchi, and A. Ichikawa. 2007b. Skill-nms for an indicator of qualitative skill in the interpreters of japanese sign language. *Proceedings of International Symposium on Skill Science (ISSS)*, pages 178–180.
- S. Tanaka, K. Nakazono, M. Nishida, Y. Horiuchi, and A. Ichikawa. 2008. Evaluating interpreter's skill by measurement of prosody recognition. *Transactions of the Japanese Society for Artificial Intelligence*, 23(3):117–126.

## Annotation of Sign and Gesture Cross-linguistically

Inge Zwitterlood<sup>1,2</sup>, Asli Özyürek<sup>1,2</sup>, Pamela Perniss<sup>1,3</sup>

<sup>1</sup> Max Planck Institute for Psycholinguistics, Nijmegen

PO Box 310, NL-6500 AH Nijmegen, The Netherlands

<sup>2</sup> Department of Linguistics, Radboud University Nijmegen

<sup>3</sup> Deafness Cognition and Language (DCAL) Research Centre, University College London

E-mail: i.zwitterlood@mpi.nl, asli.ozyurek@mpi.nl, pamela.perniss@mpi.nl

### Abstract

This paper discusses the construction of a cross-linguistic, bimodal corpus containing three modes of expression: expressions from two sign languages, speech and gestural expressions in two spoken languages and pantomimic expressions by users of two spoken languages who are requested to convey information without speaking. We discuss some problems and tentative solutions for the annotation of utterances expressing spatial information about referents in these three modes, suggesting a set of comparable codes for the description of both sign and gesture. Furthermore, we discuss the processing of entered annotations in ELAN, e.g. relating descriptive annotations to analytic annotations in all three modes and performing relational searches across annotations on different tiers.

### 1. Introduction

In a five-year project, we compare expressions in the spatial domain, particularly related to establishing and maintaining reference, between two unrelated sign languages (German Sign language and Turkish Sign Language; henceforth DGS and TID), the co-speech gestures accompanying two spoken languages (German and Turkish), and the pantomime-like structures used by hearing non-signers in Germany and Turkey when asked to convey information without speaking. This comparison aims to discover the similarities and differences in the way in which information pertaining to the identity, location, motion, and action of is expressed between the sign languages, between the co-speech gesture expressions in the spoken languages, and between the signing, co-speech gesture and no-speech pantomime modes. To this end, a large video corpus of task-related discourse data is being constructed. The aim is to record 90 minutes of useable data per participant, with 15 participants per condition (resulting in 135 hours of useable data).

The data will be described using the IMDI metadata standards and entered in the Browsable Corpus based at the MPI for Psycholinguistics. Parts of the data will be made accessible for other researchers and for educational purposes at the end of the project. The video data are annotated using the ELAN (Eudico Linguistic Annotator) annotation tool developed at the same institute.

In this paper we report on the development of the annotation conventions we use in this project, discuss their advantages and shortcomings, and suggest further improvements. Furthermore, we

will explain how we devised our annotation templates in order to enable relational searches between tiers and across annotations after they have been entered in ELAN.

### 2. Data Collection

#### 2.1 Stimuli

In order to facilitate comparison between the languages and the communication modes, the same stimulus materials were used in all three conditions. We used (and where necessary, adapted) existing materials that have been used successfully in the past, but also created new materials to target specific domains of our research questions. Future cross-linguistic research into several aspects of the sign languages and modes was also taken into account in the choice of these materials.

The stimuli include animated movies from "Die Sendung mit der Maus" (as used by Perniss 2007), the Canary Row episodes (used in various sign and gesture language research projects, e.g. McNeill 1992), and selected scenes from Charlie Chaplin movies (So *et al.*, 2005). These, as well as a subset of Zwitterlood's classifier pictures (Zwitterlood 2003) and the Balloon Story pictures (used e.g. by Kuntay 2002), were used to elicit first and subsequent mentions of referents in various types of locative and motion constructions. Furthermore, a subset of the pictures used by Volterra *et al.* (1984) as well as newly constructed movies were included to elicit verbs expressing events of *giving* and *taking* and their arguments. Finally, a large set of photographs was compiled for elicitation of quantified expressions of location of single and multiple referents. In the data collection sessions, the sets of stimuli are presented to the

speakers/signers on a laptop computer, and are worked through at the participants' own pace.

In addition to elicited data, (semi-)spontaneous data are recorded by asking participants to describe their family and their home, and to tell one or more personal narratives of their own choice. These data are very important as a control for the frequency bias of particular grammatical structures resulting from the use of elicitation materials. Targeted elicitation is necessary to ensure the presence of the phenomena under investigation in the data. The inclusion of spontaneous data is important to confirm the occurrence of these phenomena in more natural discourse.

## 2.2 Recording Procedure

Each recording session requires two native (or near-native) speakers or signers. One person is the main signer/speaker; the other is the addressee. First, this creates a more natural conversation. Second, in some tasks, the addressee is asked to perform a task (such as to point to a picture out of four that matches the signer's description), so we can test whether the message was understood.

The participants are seated opposite each other. Each participant is recorded individually with a Sony DV camera. Both participants are also recorded together from above by a Sony DV camera with a wide angle lens, mounted on a tall tripod. Previous research on motion and location in sign language (Zwitserslood 2003) has shown that a top view, in which the relation of the hands to the body in terms of distance and direction can be seen clearly, is indispensable when investigating the use of space by language users.

The tasks are explained by a native speaker/signer, who coordinates the recording session and helps out where necessary.

## 2.3 Data Processing

The data are recorded on mini-DV tapes with standard DV recorders. The content of the tapes is captured on Apple Macintosh computers and processed using Final Cut Pro version 4. The video signals from the three recordings (i.e. front view of speaker/signer, front view of addressee, top view) are synchronized on the basis of an audio peak, resulting from three claps at the start of a recording session (see also Crasborn & Zwitserslood, this volume). The segments containing relevant data are exported as separate files and compressed to MPEG1 format.

## 3. Annotation

The data are made accessible and searchable by providing annotations. The annotation program

used is ELAN, displaying, as needed, movies of the speaker/signer, addressee, and/or top view. In order to be able to compare annotations of data from the different languages and modes, the coding templates for each mode contain the same or similar tiers, with a comparable coding scheme for making annotations in each mode.

### 3.1 Existing Annotation Conventions

The annotation of sign language, gesture, and pantomime is relatively new and to date there are no clear, standard conventions. The process is extremely time consuming, especially when there are so many aspects that could be of importance. Due to time limitations and particular research targets, researchers necessarily make choices about which aspects need to be annotated and how, in order to be able to answer their specific research questions. This is unavoidable, but hampers comparison to data and annotations of other researches. We have studied annotation conventions of previous projects, in order to learn from their experiences and to use (subsets of) these annotation conventions. Regarding sign language annotation, there are various annotation systems, some even quite extensive (e.g. the Berkeley Transcription System (Slobin *et al.* 2001), conventions as used in the ECHO project (Nonhebel *et al.* 2004a,b) and the Auslan corpus project (Johnston & De Beuzeville, 2007). For gesture annotation, only one coding system is reported Kita *et al.* (1997). We combined the methods developed for gesture coding and sign coding and extended it for our purposes.

Current annotation systems and conventions for sign languages often directly analyze parts of the sign stream (e.g. by providing a sign with a gloss) or combine parts of that which is observable (e.g. the form of the sign or a particular location in space) with an interpretation or analysis (e.g. the annotation PT:PRO indicates that a sign is a pointing sign [PT] and that this pointing sign is a pronoun [PRO]). Since the structures we are interested in still need a lot of study and it is known that sign language annotation tends to involve (sometimes undesired) interpretation, at a (too) early stage (see also Leeson & Nolan, this volume), we wanted to make a clear division between annotations on a mere *descriptive* level and annotations on an *analytic* level. Annotations on descriptive level tiers describe signs/expressions in terms of their phonetic/phonological<sup>1</sup> form only, while annotations on the analytic level tiers provide

<sup>1</sup> We use the term "phonetic/phonological", since it is still unclear in many cases if a particular sign component can be analysed as a phoneme or should be considered a particular pronunciation of a phoneme.

an interpretation and/or analysis. With both types of tiers, analytic annotations can be based on and linked to descriptive annotations, or can be independently re-analyzed if this proves to be necessary, and mismatches between descriptive and analytic annotations can easily be found and adapted. (This will be described in more detail in sections 3.3 and 3.4 below.)

However, to date, the transcription of non-oral utterances has been severely impeded by the lack of an orthographic or a phonetic/phonological notation system. Some phonetic or phonological systems have been developed for the notation of sign languages, such as the Stokoe transcription system and later developments (Stokoe *et al.* 1965 and later documentation) and HamNoSys (see Prillwitz *et al.* 2001 and later documentation). However, either such a system cannot be used in ELAN (since a system needs to have an accepted Unicode font in order to be implemented in ELAN; this is not the case for some systems); or a system is not transparent (enough), using regular fonts (letters, punctuation marks, etc.) that have no relation to what they describe. In the end, we selected and combined conventions from several systems.

### 3.2 Annotation of Different Modes

Some differences in annotation between the different modes are unavoidable, since in the co-speech gesture mode it is possible to annotate the spoken words, using the commonly used orthography. This is not possible in the sign/gesture mode, because of the aforementioned lack of an orthography or a clear and transparent system for phonetic/phonological annotation of gesture/sign. As a result, in the co-speech gesture there is a separate (descriptive) tier for the annotation of the German or Turkish speech. Along with this tier, there is an (analytic) tier containing the English translation, for easy access, quick reference, and comparison.

In the sign language mode, every sign/gesture is annotated by means of glosses in German or Turkish by native signers who are bilingual in DGS and German, or in TID and Turkish). Three tiers are involved: a tier for each hand and a tier for signs in which both hands are acting in unison (e.g. when both hands are clasped together). One-handed signs are annotated on the appropriate Left or Right Hand Tier, two-handed signs on both the Left and Right Hand Tier, and the special cases in which both hands act together are on the Both Hand Tier. Separate tiers (linked to the German or Turkish tiers) contain English translations of these glosses.

Annotation is a laborious, time consuming process. In view of the large amounts of data and the time allotted to the project it would be impossible to give such detailed annotations for each utterance. Therefore, besides annotating each word in the co-speech gesture mode and each sign/gesture in the sign language and no-speech pantomime modes, we focus our efforts on the utterances that are of particular interest to us. In the chosen utterances, in all modes, positions and movements of the hand(s), face, eyes, and body are annotated in a way as similar as possible.

### 3.3 Annotation at the Descriptive Level

#### 3.3.1 Manual Elements

In all three modes, for the selected utterances, position, action and shape of each hand is annotated on the appropriate Left Hand, Right Hand or Both Hands tier. Descriptions of the hand configuration (handshape and orientation combined) is described in one annotation; location or movement is described in an annotation on a separate tier.

Handshape is described using the handshape table from HamNoSys version 2, with additions as described in Van der Kooij (2002). Since the HamNoSys font is not available in ELAN and because of a lack of a generally accepted set of handshape labels, we used the solution by Kita *et al.* (1997), who assigned letters to the rows in the handshape table and numbers to the columns. For example, a  handshape, placed in the first row and in the third column of the table, is coded as "A3".

Finger and palm orientation of each hand are coded together, in the vein of the HamNoSys codings. For palm orientation, we use a subset of the HamNoSys codings, labeled with small letters (e.g. "d" for "down", "r" for "right"). We diverge slightly from HamNoSys in the interpretation of the palm orientation: In handshapes where the fingers are extended we code the orientation of the *inside of the fingers* rather than the orientation of the palm of the hand. This results in the same orientation for handshapes with straight and bent fingers and is therefore, in our opinion a better description of the palm orientation. Our interpretation of finger orientation is also slightly different from that prescribed in HamNoSys: We code the direction in which *the fingers* are actually pointing instead of the direction in which the fingers would point if they extended straight from the hand ("extended finger orientation"). This way of coding finger orientation gives a clearer indication of the orientation than the original one, because often handshapes with bent extended fingers are pronunciation variants of handshapes with

extended fingers. The finger orientations of HamNoSys are labeled with capital letters (e.g. “U” for “Up”, “LD” for “Left Down”).

An example is “E6tD”, describing the hand configuration in Figure 1.

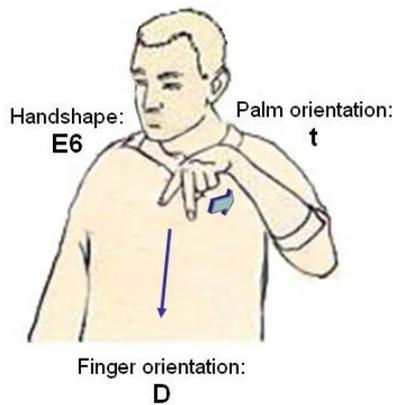


Figure 1: Hand configuration codes

In case a sign contains a handshape change and/or an orientation change, this is coded with the initial and final handshapes/orientations, separated by an arrow, e.g. “A3Ou->C10Oau” describes a handshape and orientation change.

Description of locations and movements of the hand(s) (and of non-manual elements) is, in existing systems, either not possible or too crude, e.g. “left” is not detailed enough in contexts where there may be several referents located to the left of the signer. Also, height may need to be taken into account. Therefore, we devised a 3-dimensional grid with combination codes, to which horizontal and vertical locations in signing space can be assigned. The vertical codes are shown in Figure 2, the horizontal codes in Figure 3. A combination of these codes is used within single annotations.

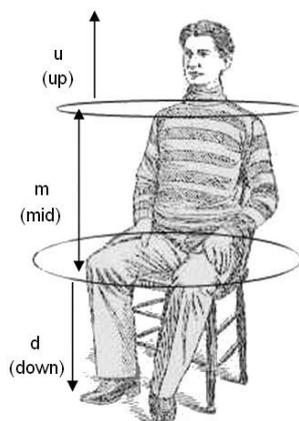


Figure 2: Vertical part of 3D location grid

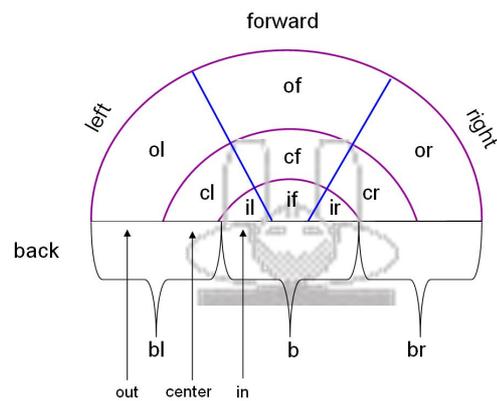


Figure 3: Horizontal part of 3D location grid

An example is “clm” (center left of signer, mid level).

In signs in which the hand moves from one location to another, this is described by the codes for the initial and the final locations, separated by an arrow, e.g. clm->clu.

### 3.3.2 Non-manual Elements

In communication, many non-manual elements can be used to convey information, e.g. about particular referents involved in the event that is being described. Body position, eye gaze, facial expression are well-known for this. To some extent, they are also used for referent indication in spoken languages. Therefore, we also code these elements in the utterances we select. For eye gaze, we use a separate tier, using (if possible) the codes from the 3-dimensional grid. There are also tiers for body position and head position. These are described using codes that express dynamic and static tilts, bends, and turns of the head and body, and head nods and shakes. To describe these, we selected a subset of the options described in HamNoSys (Hanke *et al.* 2001) and (as yet unpublished) in the annotation conventions used in a research project on prosody in the Sign Language of the Netherlands (Van der Kooij, p.c.). We use codes such as “sLF” and “dLL” to describe that the signer’s body shows a static turn to the left and a dynamic movement, leaning leftwards, respectively, and “tiltL” and “SNodU” to describe a head tilt to the left and a single upward nod of the signer’s head, respectively.

### 3.4 Annotation at the Analytic Level

Besides giving a description of the forms we see in a given discourse, we need an interpretation of the signs/gestures and other, non-verbal information. For example, we code whether a sign contains a classifier and the type of classifier. We are especially interested in coreference mechanisms in

the discourse, that is, the ways in which referents receive first and subsequent mentions. In sign languages, this can be done manually, by pointing or signing at particular locations in signing space, or by using classifier handshapes. Non-manually, it can be done by body or head shifts towards particular locations in signing space and/or by facial expression. In co-speech gesture, it is argued that similar ways of referring to referents are available. We indicate all referents that are referred to in the sign/gesture/speech signals in annotations on a separate tier, and we try to connect them to annotations on descriptive tiers. In that way, we hope to find systematicity in the expression of referents on three possible levels: language-specifically, cross-linguistically, as well as cross-modally.

#### 4. Further Use of the Annotations

What is the next step if one has finished a set of annotations? ELAN is a powerful annotation tool with search functionality, but that functionality is, so far, restricted. It is possible to find particular annotations in one or more files and to restrict one's searches (e.g. to a particular time interval or to a subset of tiers). However, it is not possible to enter relational searches, i.e. searches where the annotations one is looking for on one tier are related to annotations on another tier. It is important to realize this before one starts to enter annotations, because the use one wants to make of the annotations influences the structure of one's ELAN templates. In our case, we wanted to be able to list annotations linked to particular annotations on other tiers, e.g. we wanted to be able to see all handshapes and locations that are used to refer to a particular referent (in all modes).

Although such relational searches cannot be done in ELAN, there are possibilities to do such searches outside the tool, in data that are exported from ELAN to another application that does have those facilities (in our case: Microsoft Excel). In order for this to work, the relations between annotations on different tiers should already be made in the ELAN template; the exported annotations then include these relations. It is possible to link annotations on parent tiers (independent annotations) with annotations on child tiers (dependent annotations). These annotations can be exported to Excel in a schematic structure, that can then easily be used for several searches.



Figure 4: Screenshot of annotation of a German narrative in ELAN

#### 5. Concluding Remarks

In this project, we have extensively considered the possibilities and intricacies of making comparable annotations of similar types of information expressed in several modes. The first, real challenge is to find a means to describe the non-verbal expressions in such a comparable way, especially since there are no clear-cut, interpretation-neutral conventions for the annotation of non-verbal expressions. The second challenge is to find a way to relate the different types of annotations that are entered in ELAN and to be able to make easy comparisons on the basis of those annotations.

The second challenge is answered by using annotation templates, in which the relations between annotations on different tiers that we are interested in are already established, so that the relations can be viewed in another application (i.e. in Excel).

With respect to the first challenge, we use particular annotation conventions to circumvent the problems of mixing or missing information concerning the form and the interpretation of non-verbal expressions by distinguishing *descriptive level* and *analytical level* annotations, and by using non-analytic codes in annotations at the descriptive level. However, the codes we use are a combination of existing codes, adapted where these codes

appeared not to be clear (enough) and extended with extra codes, and thus, they do not form a conventional system. Furthermore, a real problem is the fact that many codes in our system are still not very transparent, as they are based on common fonts used for the description of spoken languages. We would like to encourage the linguistic community (especially that part of that community that is involved in non-verbal communication) to work on an (accepted) orthography for sign language *and* transparent phonetic and phonological annotation systems for non-verbal communication, that can and must be implemented in software applications for the annotation and processing of such communication. That way, over- and misinterpretation as often caused by mere gloss annotations and annotations that combine descriptions and analyses can be avoided in the future. Furthermore, easier and better comparison of data and analyses is facilitated.

## 6. Acknowledgments

The project is funded by the Netherlands Organization for Scientific Research (NWO), in the framework of the Vernieuwingsimpuls (VIDI grant (no: 276-70-009) awarded to Asli Özyürek, and also by the MPI for Psycholinguistics, The Netherlands.

## 7. References

- Kita, S., Van Gijn, I. & Van der Hulst, H. (1997). Movement Phase in Signs and Co-Speech Gestures, and Their Transcriptions by Human Coders. In Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction 1371, pp. 23-35.
- Hanke, T., Marshall, I., Safar, E., & Schmaling, C. (2001). "Interface Definitions", Deliverable D5-1 from the earlier ViSiCAST project - a technical description of the languages defined to interface between various components used in eSIGN work.
- Johnston, T. & De Beuzeville, L. (2007). Auslan Corpus Annotation Guidelines. Ms., Macquarie University & SOAS, University of London.
- Kuntay, A. (2002). Development of expression of indefiniteness: Presenting new referents in Turkish. In *Discourse Processes* 33(1), pp. 77-101.
- Nonhebel, A., Crasborn, O. & Van der Kooij, E. (2004). Sign language transcription conventions for the ECHO project. Version 9, 20 January 2004. Ms., Radboud University Nijmegen.
- Nonhebel, A., Crasborn, O. & Van der Kooij, E. (2004). Sign language transcription conventions for the ECHO project. BSL and NGT mouth annotations. Ms., Radboud University Nijmegen.
- Perniss, P. (2007). Space and Iconicity in German Sign Language (DGS). Nijmegen, MPI Series in Psycholinguistics 45.
- Prillwitz, S., Leven, R., Zienert, H., Hanke, T., Henning, J., (1989). Hamburg Notation System for Sign Languages: An Introductory Guide. Hamburg, International Studies on Sign Language and the Communication of the Deaf 5.
- Slobin, D. I., Hoiting, N., Anthony, M., Biederman, Y., Kuntze, M., Lindert, R., Pyers, J., Thumann, H., Weinberg, A. (2001). Sign language transcription at the level of meaning components: The Berkeley Transcription System (BTS). *Sign Language & Linguistics* 4, pp. 63-96.
- So, W.C., Coppola, M., Licciardello, V., & Goldin-Meadow, S. (2005) The seeds of spatial grammar in the manual modality. *Cognitive Science* 29, pp. 1029-1043.
- Stokoe W.C., Casterline, D.C. & Croneberg, C.G. (1965). *A Dictionary of American Sign Language Based on Linguistic Principles*. Silver Spring, Md.: Linstok.
- Van der Kooij, E. (2002) Reducing phonological categories in Sign language of the Netherlands. Phonetic implementation and iconic motivation. Utrecht: LOT Dissertation Series 55.
- Volterra, V., Laudanna, A. Corazza, S., Radutsky, E. & Natale, F. (1984). Italian Sign Language: the order of elements in the declarative sentence. In Loncke, F. *et al.* (Eds.) *Recent Research on European Sign Languages*, pp. 19-48.
- Zwitserlood, I. (2003). *Classifying Hand Configurations in Nederlandse Gebarentaal (Sign Language of the Netherlands)*. Utrecht, LOT Dissertation Series 78.