4th Workshop on the Representation and Processing of Sign Languages:

Corpora and Sign Language Technologies

Saturday, May 22 & Sunday, May 23

ABSTRACTS

Workshop Organisers/Editors:

Philippe Dreuw, Eleni Efthimiou, Thomas Hanke, Trevor Johnston, Gregorio Martínez Ruiz, Adam Schembri

WORKSHOP PROGRAMME

Saturday, May 22
09:00-10:30 Oral Session A *General Corpus Issues*
10:30-11:00 Coffee Break
11:00-13:00 Poster Sessions B1 & B2 *General Corpus Issues & Elicitation*
13:00-14:30 Lunch Break
14:30-15:30 Panel Session C1 *Tools*
15:30-16:00 Oral Session C2 *Avatar Technology*
16:00-16:30 Coffee Break
16:30-19:00 Poster Sessions D1 & D2 *Tools & Avatar Technology*

Sunday, May 23
09:00-10:30 Oral Session E *Recognition & Translation*
10:30-11:00 Coffee Break
11:00-13:00 Poster Sessions F1 & F2 *Recognition & Translation*

A General Corpus Issues

Saturday, May 22, 09:00-10:30

Chairperson: **Thomas Hanke** Oral Session

Adding value to, and extracting of value from, a signed language corpus through secondary processing: implications for annotation schemas and corpus creation

Trevor Johnston

A basic signed language (SL) corpus is created through primary processing of video recordings using multi_media annotation software. Primary processing entails the tokenization and identification of SL units. For the purposes of linguistic research a corpus also needs secondary processing. Secondary processing entails appending tags for specific linguistic features to primary annotations. I draw on the experience from the Auslan corpus project to describe how primary and secondary processing can be used in corpus-based SL research. In particular, I show how the tier structure of ELAN can be used to tag SL units in a variety of ways, and how this information can be used to glean new information from the corpus which can then be added as new annotations to the corpus. Value-adding by principled and systematic primary and secondary processing of digital recordings is thus not only essential for corpus creation ('machine-readability'), it also enables further enriching of the corpus so that even more value can be extracted. I conclude by discussing the implications for annotation software and standardized annotation schemas used in the creation of SL corpora.

Adapting an Efficient Entry System for Sign Languages

Carlos R. Machado Oliveira

Building sign language written corpora may, combined with video corpora linguistics, provide richer sign language research frameworks. Tools that allow direct sign language writing could increase sign language corpora availability significantly. Here, the adaptation of a free efficient computer entry system to allow sign writing is presented.

Issues in creating annotation standards for sign language description

Adam Schembri, Onno Crasborn

In this paper, we discuss the need for a standardised system of annotation for sign language corpora. Although several tools exist for the annotation of video data (such as ELAN or iLex), and some existing projects have annotation guidelines (e.g., Crasborn et al., 2007; Johnston, 2010), a widely adopted standard is currently unavailable. First, we discuss the purpose of a set of unified annotation standards for sign languages: such standards would provide a shared set of conventions for the easy exchange of data across different sign language corpus projects and may increase consistency within corpora. Next, we discuss the properties that would define a good set of shared annotation conventions (Beckman et al., 2009). We examine some of the proposed annotation standards for spoken language description, such as the ToBI conventions for prosody and the Leipzig Glossing Rules for morpho-syntax. Lastly, we discuss the relationship between theory and description. Dryer (2006) pointed out that linguists often contrast 'theoretical linguistics' with 'descriptive' work. But if one accepts the argument that there is indeed no 'atheoretical description', then sign language linguists need to agree on a shared theory for basic sign language description, and how this translates into annotation practices.

B1 General Corpus Issues

Saturday, May 22, 11:00-13:00

Chairperson: **Trevor Johnston** Poster Session

An Arabic Sign Language Corpus for Instructional Language in School

Abdulaziz Almohimeed, Mike Wald, Robert Damper

Machine translation (MT) technology has made significant progress over the last decade and now offers the potential for Arabic sign language (ArSL) signers to access text published in Arabic. The dominant model of MT is now corpus based. In this model, the accuracy of translation correlates directly with size and coverage of the corpus. The corpus is a collection of translation examples constructed from existing documents such as books and newspapers; however, no written system for sign language (SL) comparable to that used for natural language has yet been developed. Hence, no SL documents exist, complicating the procedure for constructing an SL corpus. In countries such as Ireland and Germany, a number of corpora have already been developed from scratch and used for MT. There is no

ArSL corpus for MT, requiring the creation of a new ArSL corpus for language instruction. The goal of building this corpus is to develop an automatic translation system from Arabic text to ArSL.
This paper presents the ArSL corpus for instructional language constructed for use in schools, and the methodology used to create it. The corpus was collected at the College of Computer and Information Sciences at Imam Muhammad bin Saud University in Riyadh, Saudi Arabia. A group of interpreters and native signers with backgrounds in education were involved in this work.
The corpus was constructed by collecting instructional sentences used daily in schools for the deaf. The syntax and morphology of each sentence were then manually analysed. Each sentence was individually translated, recorded on video, and stored in MPEG format. The corpus contains video data from three native signers. The videos were then annotated using an ELAN annotation tool. The annotated video data contain isolated signs accompanied by detailed information, such as manual and non-manual features. The last procedure in constructing the corpus was to create a bilingual dictionary from the annotated videos.
The corpus comprises two main parts. The first part is the annotated video data, comprising isolated signs with detailed information, accompanied by manual and non-manual features. It also contains the Arabic translation script, including syntax and morphology details. The second part is the bilingual dictionary, delivered with the annotated videos.

Issues underlying a common Sign Language Corpora annotation scheme

Antonio Balvet

Corpus-based Sign Language linguistics has emerged as a new linguistic domain, and as a consequence large-scale and controlled video data repositories are under construction for different Sign Languages.
Nevertheless, as pointed by (Johnston, 2008) no unified annotation scheme is yet available, which compromises any chance of comparing or reusing corpora across research teams. Another related issue is the comparability of descriptions and formalizations between SL linguistics and mainstream linguistics. In this paper, we address the issue of the definition of a common annotation scheme for Sign Language corpora annotation, distribution, exchange and comparison. In section 2. we discuss the challenge of building inter-operable corpora for corpus-based linguistics. We also examine existing annotation schemes or strategies proposed for SL linguistics.
In section 3. we propose a small set of annotation tiers, based on Frame-Semantics, as a common annotation scheme. We also propose to add text-level as well as utterance-level metadata to this common annotation scheme, in order to broaden the range of future uses of SL corpora.

Why are you raising your eyebrows?

Genny Conte, Mirko Santoro, Carlo Geraci, Anna Cardinaletti

It is widely known that sign languages make an extensive use of non-manual markers (NMM) to transmit linguistic information. Some NMMs are specific to particular constructions (in several Sign Languages, furrowed eyebrows is mostly used to mark wh-questions, while headshake is used to mark negation), others may occur in several unrelated constructions (see eyebrow raising in American sign language). This study presents preliminary results of a quantitative investigation of the distribution of raised eyebrows (re-NMM) in Italian Sign Language (LIS). Re-NMM frequently occurs in spontaneous signing and is used to mark a variety of constructions; therefore re-NMM qualifies as a good candidate for a VARBRUL analysis. In particular, re-NMM may mark 8 different constructions in LIS: yes/no-questions, topics, if-clauses, correlative clauses, focus, contrastive focus, subordinate clauses, and the signer's attitude. Data come from a corpus of LIS and have been analyzed with the ELAN software. Results show an even distribution across the sample for most of the uses of re-NMM. Only two functions turned out to be significantly different: the use of re-NMM as a focus marker and the use of re-NMM as an attitude marker, which are sensitive to age.

Towards decoding Classifier function in GSL

Eleni Efthimiou, Stavroula-Evita Fotinea, Athanasia-Lida Dimou, Constandinos Kalimeris

Here we will present work based on a corpus specially designed and elicited in order to provide data for the study of Classifier function in Greek Sign Language (GSL).
Data elicitation was based on presentation to informants of a series of stimuli which lead to utterances entailing the set of Classifier functions met in the language.
The whole set of video recorded data were annotated in order to provide an appropriate corpus for the investigation of Classifier instantiations. Annotation work was complemented by the use of a search tool external to the ELAN environment, which was developed in order to handle the whole of annotated material. This search tool allows to create a data base of annotated video clips, by exploiting the set of classification features used to annotate the video recorded data. Among the attribute-value pairs forming the complete set of annotation features used, the following tiers of annotation were adopted: a) "Discourse Unit", indicating the sentence or utterance in which a classifier is met, b) "CP____" for marking the maximal classifier predicate, c) "CP_GLOSS" to describe the semantic content of classifiers, and d) "HS" with font symbols as values for specifying the handshape or handshapes involved in signing. These tiers provide the necessary information to group pieces of data as to the different classifiers and classifier functions met in GSL, i.e. [Discourse Unit: various types of tables], [CP_MAX: round tables of different sizes], [CP_GLOSS: ROUND, FLAT, SIZE], [HS: D, L, B...].
Theoretical analysis of the so created linguistic data supports formulation of a proposal for Classifier behaviour which differentiates among three distinguished major grammar functions. The key property that allows for a principled account of Classifier behaviour is that Classifiers are semantic markers which create semantic classes of objects recognised to share a set of common semantic features.

In this line, we will argue that Classifiers are morphemes articulated according to SL phonology primes. According to our proposal GSL utilises Classifier morphemes in three distinct ways:
i) To create new lexicon items: Classifier affixation adds specific semantic properties to an entity, making it part of the semantic class this specific Classifier identifies.
ii) To add qualitative/quantitative values: Classifiers function as modifiers adding qualitative/quantitative values to syntactic heads or maximal phrases.
iii) To serve co-indexing: In sign utterances, Classifiers may be used as pronominal elements, where co-indexing obligatorily involves an expanded set of agreement features which, apart from the standard features "Number" and "Gender", also includes the feature "Semantic Class".

The Icelandic sign language dictionary project: some theoretical issues

Nedelina Ivanova

There are approximately 300 deaf users of Icelandic Sign Language (Íslenskt táknmál, ITM). The first dictionary of ITM was published in 1976 and was last edited in 1988. The ITM dictionary is a wordlist consisting of illustrations of the signs, sometimes specially invented for the list's purpose, presenting an Icelandic word or an inflected form of a common Icelandic verb and loans from Swedish and Danish Sign Language because it was considered that the total number of signs was insufficient. In 2004 The Association of Parents and Benefit Society of Hard of Hearing children subsidized a compilation of signs which was published on the Internet under the name The sign bank. The novelty is that signs are shown by 'demo video clips'. Actual lexicographical work has not been done in this field in Iceland. These circumstances call for a compilation of an electronic dictionary of ITM based on linguistic principles and lexicographical methods.
The facts that dictionary compilation for SL is in general time-consuming, expensive and the limited number of potential users similarly to ITM make the work on a dictionary of ITM very difficult. The dictionary project for ITM has been more or less at a theoretical stage during the last two years, starting in 2008 with a M.A. thesis on lexicographical description for an electronic dictionary of ITM on the basis of linguistic principles and in 2009 with a description of a lexical bilingual database for the dictionary compilation. At the same time in 2009 a list of 6441 signs was compiled by Deaf and hearing researchers at the Communication Centre for The Deaf and Hard of Hearing. Today in 2010 the project is on hold due to financial reasons.
However, the electronic dictionary project of ITM is the first incisive research of ITM lexicon. The purpose with ITM dictionary with its 4000 entries, when published, is to give answers concerning sign's base form, meaning and appropriate usage.
This paper reports on the lexicographical description for construction of a dictionary for ITM. The author reviews briefly some theoretical issues regarding the dictionary's project: the languages in the dictionary and its potential users, sign's collection, evaluation and selection, the lemmatizing process, the dictionary entry, access structures, the dictionary article and two practical problems. The languages in the dictionary are Icelandic and ITM, with Icelandic as L1 and ITM as L2. Potential users of the dictionary include members of the general public interested in ITM; parents of Deaf children and their hearing friends, interpreters and hearing people teaching ITM, students in Sign Language studies, people who attend SL courses as well as the Deaf people themselves. Signs, thought to be every day vocabulary and used by most of the Deaf people, would to be found in the dictionary. The dictionary entry is a sign in its base form shown by 'demo video clip' and an Icelandic gloss. The lemma selection for lexical items with identical base form is influenced by mouth movements and mouth gestures as a lexicalized part of the lemma on the semantic level. The dictionary's access structure requires every sign's phonological description. With the potential users in mind, many access possibilities make the search for a sign easy and quick. It is possible to search in the dictionary after four criteria based on phonological structure of signs, after picture themes with illustrations and an Icelandic word. In the dictionary article phonological information is given by pictures which show sign's handshape and location; sign's meaning is given by Icelandic equivalent(s) or explanation(s); sign's modification for subject-object verb agreement is shown by example and sign's modification for plural is shown by link to the correspondent part in the explanatory grammar chapter in the dictionary. Information on the use of the entry is given by example which consists of 'demo video clip', sentence's gloss and Icelandic translation. Practical problems concern e.g. the presentation of classifier predicates in the dictionary article and the low reliability of hearing researcher moderating discussion sessions with Deaf informants.
In conclusion, the paper highlights the dictionary's importance for (1) documentation and basic research of ITM and (2) getting legal recognition of the language.

A comparison of two linguistic sign identification methods

Tommi Jantunen

This paper employs two linguistic sign identification methods – a manual one focusing on the dominant hand and a nonmanual one focusing on the mouth – and compares the kinds of sequences they classify as signs from a video containing continuous signing. The study is motivated by two projects, of which one investigates the ontological nature of the sign and the other aims to develop an automatic sign recognition tool. In the study, both methods were able to associate all the free semantic-functional elements in the data with signs. However, in the nonmanual method the overall number of identified signs was lower because the stretching of the mouth movement of the semantic element over the following pointing meant that the combinations of semantic elements and pointings were counted as single signs. Moreover, signs identified by the nonmanual method were longer than those identified by the manual method. The results from the nonmanual method agree with the claim that phrase internal sequences of semantic elements and pointings are lexical head plus clitic combinations. Consequently, it is suggested that pointings in such contexts do not need to be independently detected by the automatic sign recognition tool.

Glossing a multi-purpose sign language corpus

Ellen Ormel, Onno Crasborn, Els van der Kooij, Lianne van Dijken, Ellen Yassine Nauta, Jens Forster, Daniel Stein

This paper describes the strategies that have been developed for creating consistent gloss annotations in the latest update to the Corpus NGT. Although the project aims to embrace the plea for ID-glosses in Johnston (2008), there is no reference lexicon that could be used in the creation of the annotations. An idiosyncratic strategy was developed that involved the creation of a temporary 'glossing lexicon', which includes conventions for distinguishing regional and other variants, true and apparent homonymy, and other difficulties that are specifically related to the glossing of two-handed simultaneous constructions on different tiers.

You Get Out What You Put In: The Beginnings of Phonetic and Phonological Coding in the Signs of Ireland Digital Corpus

Gudny Bjork Thorvaldsdottir

This poster discusses a range of issues with respect to expanding the annotation of the Signs of Ireland (SOI) corpus to incorporate phonetic and phonological coding. This forms part of ongoing PHD research work that explores the phonology-morphology interface in Irish Sign Language (ISL).

The SOI corpus consists of over 40 narratives that have already been highly annotated: it contains glossed lexical signs, classifier constructions and non-manual features. Classifier handshapes have also been annotated. It is my intention to identify the phonemes and the allophones of ISL using the corpus and it is thus neccessary to incorporate a detailed annotation at the phonetic level.

In order to achieve this, a list of phonetic features for ISL must be identified. To date no research has been done in this area apart from basic work describing handshapes in ISL. Thus far, there is no agreement on the phonetic alphabet inventory for ISL: Ó'Baoill and Matthews (2000) identified 66 handshapes while Matthews (2005) identified 78. The issue of allophonic variation has not yet been tackled for this language.

For annotation purposes, challenges arise in terms of how handshapes are recorded: for example, of the 66 handshapes identified in Ó'Baoill and Matthews (2000), 28 are established as occurring as classifier handshapes also. These are annotated following ECHO project annotation norms (Nonhebel et al. 2004) where possible, with additional handshapes drawn from a list of 48 classifier handshapes described for BSL in Brennan (1992) using names like CL-B, CL-ISL-K etc. within the framework of the SOI corpus.

The other parameters that have traditionally been used to describe signs (i.e. location, movement and orientation) have not been researched in ISL at phonological or morphological level. All that currently exists is a vaguely phonetic level description of parameters respect to research on American Sign Languge (ASL) (See O'Baoill and Matthews 2000; Matthews 2005).

This poster outlines how, by drawing on Crasborn's (2001) and van der Kooij's (2002) work on Sign Language of the Netherlands (SLN), a list of phonetic features have been established for ISL and the changes to the original list of features that were required in order to accommodate ISL.

I also outline the factors influencing decisions regarding the coding and naming of handshapes at phonetic level. These include the question of whether already established naming conventions be maintained. For example, moving away from established protocols will result in inconsistencies within the annotations in the corpus. However, for the purposes of phonetic research a more elaborate coding might be necessary. Another challenge involves establishing what types of tiers are needed to accommodate the proposed research as well as future research at the phonetic and phonological level.

Sign Language in Europe

Mark Wheatley, Annika Pabsch

Sign languages across the globe are fully-fledged languages that differ between Deaf communities throughout Europe and the world. A recent survey by the European Union of the Deaf gathered that there are about 650,000 sign language users in the EU for whom using a sign language is the only way to communicate and have equal access. It is therefore crucial to legally recognise national sign languages. Being treated equally without prejudice also with regards to language is a basic Human Right as postulated in the UN Declaration of Human Rights. Other rights, such as the right to education and a fair trial can only be guaranteed if sign languages are recognised as distinct languages in order to provide sign language interpreters and education in sign language. At EU level, a number of documents and Resolutions have been adopted but so far only three European countries have recognised sign language at constitutional level: Austria, Finland and Portugal. Other countries, such as Hungary and Spain have taken other legal measures to protect their sign languages. Although Europe's sign languages enjoy some recognition, sign language users across Europe are still lacking legal protection at the same level as other minorities.

B2 Elicitation

Saturday, May 22, 11:00-13:00

Chairperson: **Trevor Johnston** Poster Session

On the creation and the annotation of a large-scale Italian-LIS parallel corpus

Nicola Bertoldi, Gabriele Tiotto, Paolo Prinetto, Elio Piccolo, Fabrizio Nunnari, Vincenzo Lombardo, Alessandro Mazzei, Rossana Damiano, Leonardo Lesmo, Andrea Del Principe

This paper presents the current development of the first large parallel corpus between Italian and Italian Sign Language (Lingua Italiana dei Segni, LIS). This initiative has been taken within the ATLAS project (Automatic Translation into Sign Languages), that aims at realizing a virtual interpreter, which automatically translates an Italian text into LIS.
The Italian-LIS virtual interpreter is implemented by means of two modules interfaced by the ATLAS Extended Written LIS (AEWLIS), which is a translation-oriented representation of LIS: the first module translates the source Italian text into AEWLIS; the second module transforms the AEWLIS content into a coherent LIS sequence, smoothly animated by a virtual character.
As no significant amount of electronic data are available for Italian and LIS, we have started building a parallel corpus from scratch in order to train and tune the Italian-AEWLIS translation system, and to compare the resulting virtual animations with human-performed LIS interpretations. The corpus, which will be freely available, actually presents a tri-lingual structure, with the Italian text, the AEWLIS sequence, and the signed LIS video.

Building a corpus for Italian Sign Language. Methodological issues and some preliminary results

Carlo Geraci, Robert Bayley, Chiara Branchini, Anna Cardinaletti, Carlo Cecchetto, Caterina Donati, Serena Giudice, Emiliano Mereghetti, Fabio Poletti, Mirko Santoro, Sandro Zucchi

The aim of this paper is to discuss some methodological issues that emerged during the creation of a corpus of data for Italian Sign Language, LIS. Data were collected from 10 cities spread across the country. 18 signers from each city have been recruited. They are native speakers of LIS or later-exposed to LIS and are divided into 3 age groups (19-38, 39-58, 59-78) of 6 signers each (3 males and 3 females). The methodology of data collection and transcription is similar to that used in previous studies of variation in American Sign Language (Lucas, Bayley & Valli 2001) and Australian Sign Language (Johnston & Schembri 2006), with some differences that we discuss. The corpus consists of various kinds of texts collected with different strategies: free conversation (45 minutes), elicited dialogues (about 5-10 minutes), narration (10 minutes) and a picture-naming task (42 items). For the transcription we adopted the ELAN software (Johnston & Crasborn 2006). Finally, a brief report on some preliminary results is presented.

DGS Corpus & Dicta-Sign: The Hamburg Studio Setup

Thomas Hanke, Lutz König, Sven Wagner, Silke Matthes

Not taking into account budget restrictions, the setup of a sign language studio always is a balancing act between high quality recordings on the one hand not to make the transcription process even more complicated than it is anyway and possibly to enable automatic processing of the recordings, and on the other hand an environment where the informants still feel comfortable enough so that the recording situation does not have too much impact on the signing. In the case of the DGS Corpus project, an additional constraint is that the studio is to be relocated twelve times over the course of two years as it was decided to make the recordings in the regions instead of inviting participants to one central place to avoid dialectal mixing. One of the implications of this approach is that the studio is operated by non-specialist deaf fieldworkers with limited time available for training.
Basically all tasks in the project involve two informants interacting in different ways with each other. A moderator (the fieldworker from the region) introduces the tasks and observes the conversation, but only interferes with the conversation if absolutely necessary.
The camera setup we finally ended up with consists of seven cameras altogether, three on each informant and one for the whole scene including the moderator. Two HD cameras on the informant provide frontal and birds-eye views while a stereo camera mounted on top of the frontal-view camera provides footage that helps automatic processing. The seventh camera is an HD camera as well.
In contrary to setups in earlier projects, we invite the two informants to sit down directly facing each other, with the frontal-view camera positioned above (and behind) the head of the other informant. Pre-tests revealed that with a distance of approximately three meters, the distorsion introduced by the elevated position of the camera does not negatively affect the transcription from video. Instead, this setting provides a front view of the informant similar to the addressee's, allowing to identify body shifts as well as direction of eye gaze more easily. At the same time, this constellation avoids informants targeting their signing back and forth between the addressee and the camera.
Elicitation material and instructions are presented to the informants on screens located on the floor between them. A custom software, “Session Director” allows the moderator to present slides to the informants by the click of a button, and to keep track of the time elapsed for each individual task as well as the whole session. Using pre-recorded instructions and elicitation materials not only reduces the burdens on the moderator, but also makes sure that all informants get exactly the same input.

Session Director keeps a log of all actions started by the moderator, allowing us to exactly reconstruct what task has been worked on when. This log is easily converted into tagging in our transcription environment, iLex. This not only allows automatic segmentation of tasks and pauses, but also introduces links from the transcript to the task and vice versa.
Task descriptions for Session Director are kept as XML files, making it easy to use this freely available tool for other projects as well.

Building a Database while Considering Research Ethics in Sign Language Communities

Julie A. Hochgesang, Pedro Pascual Villanueva, Gaurav Mathur, Diane Lillo-Martin

We are constructing an American Sign Language ID-gloss Database, which will enable sign language researchers and Deaf community members to access standard glosses for common signs. Since we are working with a language used by a community that has historically been marginalized during the research process, we feel the need to include an ethical framework for working with the Sign Language community as we consider best practices for developing sign language corpora. We will refer to the guidelines, Sign Language Communities' Terms of Reference (SLCTR), outlined in Harris, Holmes & Mertens (2009). Before making the database available to the ASL community, we plan to evaluate how members will use it and what they need from the research team to facilitate such use. This evaluation will go a long way towards ensuring that ownership of the research data lies with the ASL community. Such a reflexive evaluation of ethical practices is crucial from the beginning stages and throughout the research process. This means the ASL community is directly involved in the research process, is able to access aspects of the entire process, and can have a hand in the construction of knowledge about their own language, community and culture.

Elicitation tasks and materials designed for Dicta-Sign's multi-lingual corpus

Silke Matthes, Thomas Hanke, Jakob Storz, Eleni Efthimiou, Nassia Dimiou, Panagiotis Karioris, Annelies Braffort, Annick Choisier, Julia Pelhate, Eva Safar

Within the framework of the Dicta-Sign project, parallelised sign language corpora are being compiled for four European sign languages (BSL, DGS, GSL, and LSF). The aim for the data collection was to achieve as high a level of naturalness as can be achieved with semi-spontaneous utterances under lab conditions. Therefore, informants were filmed in pairs interacting with each other. With respect to parallelisability, elicitation tasks had to be designed that result in semantically close answers without predetermining the choice of vocabulary and grammar.
The domain selected for Dicta-Sign is 'Travel across Europe'. The tasks developed within the project cover different interaction formats ranging from monologues to sequences of very short turns, also with different levels of predictability. They include communication for transport by different means and contexts as well as related personal experiences. The elicitation materials are of different media formats and at various levels of complexity. They comprise of town and transportation network maps, pictures displaying a variety of places, items and situations linked to the target domain, as well as stories presented in sign language or as a picture story. In each session ten different tasks are to be performed, each of them planned to have a duration of about five to ten minutes, thereby switching roles between the informants several times during a recording session.
Taking into account cultural differences as well as language dependent issues regarding the different countries in the project, the material was designed in a way that only minor adjustments are needed that do not change the character of a task. The elicitation tasks and materials developed within the project as well as experiences gained adjusting and using the material for Dicta-Sign's different target languages are illustrated in this paper.

Elicitation methods in the DGS (German Sign Language) Corpus Project

Rie Nishio, Sung-Eun Hong, Susanne König, Reiner Konrad, Gabriele Langer, Thomas Hanke, Christian Rathmann

The DGS Corpus Project is a long-term project with two major aims: (i) to establish an extensive corpus of DGS and (ii) to develop a comprehensive dictionary of DGS-German based on the analysis of the corpus data. During the first three years the main focus is on data collection. Before setting up the corpus design we conducted a survey to get an overview on the existing elicitation materials. The design of our data collection contains a variety of different stimuli and tasks with the special attention to free conversation, dialogues and monologues. To this effect, a range of possible discourse modes were considered: narration and renarration, discussion, report and description. The stimuli include pictures, picture stories, non-verbal film clips (e.g. cartoons and realistic film clips) and signed movies. In order to minimize the influence of the surrounding spoken/written language, written German is not used if possible. Introduction and explanation of each task is provided in DGS in form of movie clips. All tasks were tested in a pilot phase to examine their feasibility and reliability. Some of the tasks tested needed to go through several rounds of modifications while others did not work at all and thus were excluded from the data collection. In this paper, we not only present the tasks for elicitation and stimuli, but also describe their development process. We also discuss reasons why some stimuli were adopted from other projects while others had to be developed specifically for the purpose of our project.

ATLAS Project: Forecast in Italian Sign Language and Annotation of Corpora

Mara Vendrame, Gabriele Tiotto

The paper presents the preliminary results of a research project focused on the creation and the annotation of one Italian Sign Language corpus concerning the weather forecasts domain. As a result of the annotation process, our annotations of signs sequences showed that

the semantics of the signed discourse cannot be grasped just through an annotation of single weather signs which exploits the five parameters handshape, movements, directions, locations and non-manual components. Rather, from the annotation process appears that, in order to grasp the discourse semantics, it is necessary to consider the extensive use of Highly Iconic Structures in order to specify the iconic properties of the different atmospherics phenomena. In particular, it often occurs that several signs are combined among themselves (see also Cuxac, 2000; Di Renzo, et al, 2006; Pizzuto et al., 2008; Pizzuto, Rossini & Russo, 2006). Thus, respect to single signs, our analysis of complex manual and non-manual units stored in our database suggests the necessity to better explore multidimensional aspects, in order to properly develop and train an automatic translator able to translate from Italian written text to Italian Sign Language.

C1 Tools

Saturday, May 22, 14:30-15:30

Chairperson: **Eleni Efthimiou** Panel Session

ELAN/iLex Comparisons

Penny Boyes Braem, Onno Crasborn, Thomas Hanke

Researchers beginning computerised sign language transcription projects are immediately confronted with the question of which transcription environment to use. Although there are numerous papers giving either overviews or describing specific aspects of particular transcription systems, no up-to-date comparison is available that would really help the researcher today in this initial decision process. This proposal addresses this problem by suggesting a panel with experts on two of the systems currently in use by many sign language transcription projects, ELAN and iLex. The moderator would present a number of key questions, and would give the proponent of each system the same amount of time to answer each question. Answer slots lengths ranging between 30 seconds and 5 minutes would be defined to ensure brief statements, but still allow the respondents to highlight the approach taken by their respective systems. Questions and time allocations would be negotiated beforehand between the moderator and the respondents, giving each party the chance to throw in their "key items". Respondents would be free to use slides or demos for their answers.
By addressing not only basic concepts but also overall views and plans for future developments, this panel will not only be of interest for people currently in the preparation phase of a corpus project, but also for people already having some knowledge of both systems discussed.
Suggested topics for discussion include:

- Ease of setup (installation, time needed to learn the system, hardware and software requirements)
- Future sustainability/maintenance of the system
- Support (manuals, video tutorials, helpdesks etc.)
- Group functionality
- Handling metadata
- Flexibility wrt video formats
- General usability vs. sign-language specific features
- Data exchange and security
- Lexicon integration
- Transcription quality assistance
- Flexibility wrt transcription conventions
- Tier dependencies
- Data mining, manual and automatic searching
- Corpus management

C2 Avatar Technology

Saturday, May 22, 15:30-16:00

Chairperson: **Eleni Efthimiou** Oral Session

Synthetic Corpora: A Synergy of Linguistics and Computer Animation

Jerry Schnepp, Rosalee Wolfe, John C. McDonald

Synthetic corpora are computer representations of linguistic phenomena. They enable the creation of computer-generated animations depicting sign languages and are the complement of corpora containing videotaped exemplars.
Synthetic corpora have the potential to serve multiple disciplines. They can aid in the automatic recognition of sign, because they contain the geometric data required for intelligent visual detection algorithms. Synthetic corpora can also provide visual depictions of abstract representations and act as a verification tool for data integrity and hypothesis testing.

Because the signs are synthesized, not retrieved, they can be modified as they are formed. This provides the flexibility to generate an endless variety of utterances not possible with recordings, thus opening possibilities for automatic translation efforts. While representing sign for this purpose is still an open question, a synthetic corpus has the potential to serve in this capacity. The flexibility of synthetically-generated sign is also useful for the development of interpreter training software and self-directed learning tools for deaf children.

By necessity, linguistics and computer animation must play a role in the creation of such corpora, as any corpus will need to serve both disciplines. At first glance, the goals of these disciplines would appear to be at cross purposes. Linguistics researchers often use corpora to form hypotheses through queries on linguistic features. Thus the corpora must encode such general abstractions as handshape, position, motion, palm orientation and non-manual signals. In contrast, creating computer animations of sign requires voluminous and detailed data, as the resulting animations must be realistic enough to pass the scrutiny of fluent signers.

In actuality, the fields of linguistics and computer animation create a mutually beneficial synergy. Having the detailed precision required for animation can facilitate the exploration of subtle interactions among linguistic phenomena. Likewise, animators need an abstract representation to organize, combine, and synthesize complex animation data.

Regardless of the animation technique, linguistic knowledge is necessary to produce any synthetic corpus. Animators who hand-transcribe need to work closely with linguists, so that the gloss is tagged correctly. Linguistic information guides the transcription artist's efforts to produce a natural exemplar that encapsulates the essential motions of a sign. With motion capture, the role of linguistics is no less central. Motion capture equipment generates massive amounts of data that must be cleaned to remove extraneous noise. The linguistic attributes of a sign give the cleanup artists precisely what they need to process and extract the desired motion.

Our work thus far has focused on the creation of detailed and accurate animations of sign. The data that drive these animations have similarities to abstractions created by linguists. Thus, linguistic research guides the creation of a framework for automated sign synthesis.

This presentation will examine several linguistic processes and discuss an approach to their representation in a synthetic corpus, including co-occurrence in nonmanual signals. This will include a demonstration of computer-generated American Sign Language.

D1 Tools

Saturday, May 22, 16:30-19:00

Chairperson: **Gregorio Martínez Ruiz** Poster Session

CopyCat: A Corpus for Verifying American Sign Language During Game Play by Deaf Children

Helene Brashear, Zahoor Zafrulla, Thad Starner, Harley Hamilton, Peter Presti, Seungyon Lee

The CopyCat project was designed to develop an interactive educational adventure game to help deaf children acquire language skills. The main goals of the project are to improve the language and memory abilities of deaf signing children, advance basic research in computer-based sign language recognition, and design an efficient language interaction model in order to assist in the language learning of deaf children. The CopyCat project was begun as a collaboration between Georgia Tech and the Atlanta Area School for the Deaf in 2004 and has been collecting ASL (American Sign Language) data since Spring of 2005. Since then we have collected 5829 signed phrases from over 30 children.

In this paper we describe the evolution of the CopyCat system design, data collection methodology, and resulting corpus, as well as challenges and successes throughout the process.

MobileASL: Overcoming the technical challenges of mobile video conversation in sign language

Anna Cavender, Neva Cherniavsky, Jaehong Chon, Richard Ladner, Eve Riskin, Rahul Vanam, Jacob Wobbrock

As part of the ongoing MobileASL project, we have built a system to compress, transmit, and decode sign language video in real-time on an off-the-shelf mobile phone. In this work, we review the challenges that arose in developing our system and the algorithms we implemented to address them. Separate parts of this research have been previously published.

Compression and transmission of sign language video presents unique difficulties. We must overcome weak processing power, limited bandwidth capacity, and low battery life. We also must ensure that the system is usable; that is, that the video is intelligible and the algorithms that we employ to save system resources do not irritate users.

We describe the evolution of the MobileASL system and the algorithms we utilize to achieve real-time video communication on mobile phones. We first review our initial user studies to test feasibility and interest in video sign language on mobile phones. We then detail our three main challenges and solutions. To address weak processing power, we optimize the encoder to work on mobile phones, adapting a fast algorithm for distortion-complexity optimization to choose the best parameters. To overcome limited bandwidth capacity, we utilize a dynamic skin-based region of interest, which encodes the face and hands at a higher bit rate at the expense of the rest of the image. To save battery life, we automatically detect periods of signing and lower the frame rate when the user is not signing.

We implement our system on off-the-shelf mobile phones and validate it through a user study. Fluent ASL signers participate in unconstrained conversations over the phones in a laboratory setting. They find the conversations with the dynamic skin-based region of interest more intelligible. The variable frame rate affects conversations negatively, but does not affect the users' perceived desire for the technology.

Ongoing work includes varying the spatial resolution instead of the temporal resolution, further optimization of rate-distortion-complexity, and a field study to determine usability over a long period of time in a realistic setting.

Distributed System Architecture for Assisted Annotation of Sign Language Video Corpora

Christophe Collet, Matilde Gonzalez, Fabien Milachon

This paper presents one component of Dicta-Sign, a three-year FP7 ICT project that aims to improve the state of web-based communication for Deaf people. A part of this project is the annotation of sign language corpora. To improve the annotation task in terms of reproducibility and time consuming, several plug-ins for sign language video processing are developed. The component presented in this paper aims to link several plug-ins to annotation software through the network. These plug-ins can be coded in different languages, operating systems and computers. For that, it uses the SOAP Web-service and a specific data-format in XML for the data exchange.

Using ELAN for annotating sign language corpora in a team setting

Onno Crasborn, Han Sloetjes

ELAN is a multimedia annotation tool that is employed in many sign language corpus projects. It is a standalone desktop application that, like many other desktop applications, principally is a single user, document oriented application. In many scenarios this is still perfectly satisfactory but in large-scale corpus projects, involving many collaborators who are working on the same documents, the problem arises of how to resolve edit conflicts and how to prevent undesirable modifications to parts of the document. The Corpus NGT project is such a project and this paper describes the challenges that arose in the process of its creation as well as in the exploitation of this large collection of annotation documents. It outlines recent and possible future development of ELAN and alternate solutions that have been explored and applied.

iLex: Handling Multi-Camera Recordings

Thomas Hanke, Jakob Storz, Sven Wagner

Until recently, sign language researchers were quite happy with just one or two views for each recording session. While ELAN allows the user to relate several media files to a transcript and to sync them, iLex just allows one single media container and relies on the container format, such as QuickTime, to group and sync several video streams into one container. In order to save screen real estate, iLex offers the user the possibility to switch on or off individual tracks within the media file. This works quite fine with two or three different views grouped, but fails to provide an adequate solution in multi-view projects such as Dicta-Sign or DGS Corpus with seven cameras altogether for a pair of informants. The advent of HD videos makes screen real estate really an issue: Even on very large screens, video competes with transcription space.
Here we present a user interface study that allows flexible switching between video layouts whenever the transcription focus changes. Switching (including zooming and cropping) may be initiated at any point of time by the user, or can be automated to depend on tagging such as tasks or turns. This user interface is backed up by a server infrastructure providing videos in different spatial resolutions as needed for optimal display while saving transfer bandwidth and local processing power which even nowadays becomes an issue when dealing with several HD videos in parallel.

Development of a Moodle VLE Plug-in to Support Simultaneous Visualisation of a Collection of Multi-Media Sign Language Objects

Markus Hofmann, Kyle Goslin, Brian Nolan, Lorraine Leeson, Haaris Sheikh

Using Virtual Learning Environments (VLE) to support blended learning is very common in educational institutes. Delivering learning material in a flexible and semi-structured manner to the learner transforms such systems into powerful eLearning tools. However, the presentation and visualisation of individual or multiple learning objects is mostly dictated by the system and cannot be altered easily.
This paper reports on a project between Trinity College Dublin (TCD) and the Institute of Technology Blanchardstown (ITB) that aims to improve the simultaneous visualisation of multiple multimedia objects for deaf learners of ISL. The project was implemented using the Open Source VLE Moodle. Moodle's nature of being Open Source and having the ability to code plug-ins qualified it to be the most suited vehicle to address the visualisation problem. Traditionally VLEs allow the viewing of one learning object at a time, which meant that deaf learners could either view a pre-recorded, signed in ISL, video lecture or concentrate on textual accompanying content but not both. The developed Moodle plug-in allows academics to group multiple videos into a 'lecture'. It further facilitates the addition of rich text content to each video. The learner can select and view one video from a possible sequence of many as well as view the text that belongs to the video. The paper further outlines detailed implementation and techniques applied.

Eliciting Spatial Reference for a Motion-Capture Corpus of American Sign Language Discourse

Matt Huenerfauth, Pengfei Lu

The goal of our research is to identify computational models of the referential use of signing space and of spatially inflected verb forms for use in American Sign Language (ASL) animations for accessibility applications for deaf users. This paper describes our collection and annotation of an ASL motion-capture corpus to be analyzed for our research. A study was conducted to compare alternative prompting strategies for eliciting single-signer multi-sentential ASL discourse that maximizes the use of pronominal spatial reference yet minimizes the use of classifier predicates, spatially complex ASL phenomena that are not the focus of our current research.

Towards semi-automatic annotation of video and audio corpora

Stefano Masneri, Oliver Schreer, Daniel Schneider, Sebastian Tschöpel, Rolf Bardeli, Stefan Bordag, Eric Auer, Han Sloetjes, Peter Wittenburg

AVATecH (Advancing Video/Audio Technology in Humanities Research) is a project in which two Fraunhofer Institutes and two Max Planck Institutes collaborate in order to promote the development and application of technology for semi-automatic annotation of digital audio and video recordings. One of the aims of the AVATecH project is to implement algorithms that allow for the automatic or semi-automatic creation of pre-annotations for the video corpora, hence reducing the time needed to perform the manual annotation task. Due to the huge size of the corpora, and the extreme variety of the video content, the algorithms developed need to be fast, efficient and robust. In this paper we will present some of the algorithms currently under development, the modifications applied in order to get them working with large video corpora and how the results of the annotations are stored, as well as how they can be integrated in ELAN annotation software.

Data organization in a collaborative sign language reference tool

Cedric Moreau, Bruno Mascret

The Acadmie Française institution is assigned and devoted to defending the French language and to making it a common heritage for all French speakers. The French Sign Language (LSF) has never had such a support. To face this situation, a reference tool has been created, supported by the French Ministry of Education and by the General Delegation to the French language and to languages of France. This tool is a collaborative website entirely bilingual French and LSF, and which proposes for each concept at least one definition and its associated descriptors in various knowledge fields. Before being spread on-line, the information given by users (text, picture, video, presentation) is examined by experts on form and content, and is validated or rejected by these experts with an explanation.
Considering regional and sociological differences, several signs may be proposed and validated for one concept. Our project does not wish to choose the ”ideal sign”, but wants to submit to our identified users all the proposals and to list their comments (have they come across this sign and if so, in which context). A set of information is thus collected for each sign and can be related to users profiles. The website is therefore an exchange platform, but can also be used as a linguistics observatory.
One of our main issues concerning the data organization was to manage to adjust users different viewpoints and different uses of the website. Indeed, our platforms goal is not to make a simple dictionary but to create a network of ontologies. Our other issue is now that we cannot use a rigid organization model, because our website must constantly evolve and include new concepts and new descriptions or functionalities, such as illustrations, homonyms, antonyms, etc. In this article we will first briefly describe our platforms goals, then present our specific data organization which allows for example several classifications to be used simultaneously. We will illustrate this approachs interest with a critic of Deweys classification, that we had at first implemented despite its limits (acceding to a precise concept is difficult, the organization is not intuitive, recent concepts or specific LSF concepts cannot be referenced, etc.). We will propose to replace it with classifications directly created by our users and corresponding to their expectations and needs. This way the tree diagram is built gradually and supervised by experts in each knowledge field.
Each content thus goes with descriptors and classifiers allowing it to play different parts depending on the context. Therefore a content can at the same time be a concept, a classification theme or sub-theme, or an illustration – the context will mobilize the appropriate contents depending on their descriptors and classifiers. We will finally present our current work on integrating direct resources in LSF through descriptors defining a sign’s spatial position and its moves (hands, body and face), in order to highlight our platforms great ability to evolve. We will also show that this data organization allows an easy conversion to other countries sign languages.

Development of E-Learning Service of Computer Assisted Sign Language Learning: Online Version of CASLL

Saori Tanaka, Yosuke Matsusaka, Kaoru Nakazono

In this study, we introduce the problems for realizing an e-learning system available online and outline some ethical issues behind these problems. The difficulties faced to us, when we were going to open Computer Assisted Sign language Learning (CASLL) system online, were one to expose the sign language movies to public with downloadable way, one for increasing the course materials, and one to enhance the collaboration between learners. The ethical discussions revealed that the reliability for the system and the collaborative work for expand the number of course materials were necessary for overcoming the difficulties. In order to realize the reliable and the collaborative e-learning system, we implemented CASLL within Moodle, an open-source Course Management System. For re-designing the system to actual use for sign language learners and teachers, we added new functions to Moodle; the protection function for the right of publicity, the wiki function to enable collaborative course editing and finally the Link function to enhance public relations. We are going to evaluate the system design from the viewpoint of the usability for teaching, the effectivity for learning, and the utility for collaboration.

D2 Avatar Technology

Saturday, May 22, 16:30-19:00

Chairperson: **Gregorio Martínez Ruiz** Poster Session

A Multilanguage Database for supporting Sign Language Translation and Synthesis

Roberto Borgotallo, Carmen Marino, Elio Piccolo, Paolo Prinetto, Gabriele Tiotto, Mauro Rossini

The design of a language database is an important task within projects targeting sign language research. In this paper is presented a database structure that supports both linguistic information and visualisation oriented data to assist a final publication of services for deaf people. The database has been designed within the Automatic Translation into sign LAnguageS (ATLAS) project that takes aim at getting the automatic translation from written Italian to Italian Sign Language (LIS). The final step of the overall process is the enrichment of the original video with a superimposed virtual character realised by 3D animated computer graphics. The top element within the database is the A_Product defined as the main primitive element managed by the ATLAS platform under which all the other data, from input sources to the final publication modalities and attributes lay. The A_Product includes the reference to the original content and all the intermediate elaborations results towards the final publication comprehensive of the virtual character animations. Among the others, the most important transformation is the automatic translation from a written Italian text to the intermediate language AEWLIS (ATLAS Extended Written LIS), formalized within the ATLAS project.

Towards Czech on-line sign language dictionary – technological overview and data collection

Pavel Campr, Marek Hrúz, Jiří Langer, Jakub Kanis, Miloš Železný, Luděk Müller

In this article we present the current state of our work on an on-line sign language dictionary. The aim is to create both an explanatory and a translation dictionary. It is primarily targeted (but not limited) to the Czech and Czech sign language. At first we describe technological aspects of the dictionary and then our data collection practices. The dictionary is an on-line application build with respect to the linguistic needs. We use written text to represent spoken languages and several representations are supported for sign languages: videos, images, HamNoSys, SignWriting and interactive 3D avatar. To decrease time required for data collection and publishing in the dictionary we use computer vision methods for video analysis to detect sign boundaries and analyze the manual component of performed sign for automatic categorization. The content will be created by linguists using both new and already existing data. Then, the dictionary will be opened to the public with possibility to add, modify and comment data. We expect that this possibility of on-line elicitation will increase the number of informants, cover more regions and makes the elicitation cheaper and the evaluation easier. Furthermore we prepare a mobile interface of the dictionary. The mobile interface will use different format of web pages and different video compression methods optimized for slower Internet connection. We also prepare an offline version of the dictionary which can be automatically generated from the online content and downloaded for offline usage.

Corpus Design for Signing Avatars

Kyle Duarte, Sylvie Gibet

The SignCom project uses motion capture (mocap) data to animate a virtual French Sign Language (LSF) signer. An important part of any signing avatar project is to ensure that a computer animation engine has a large quantity of interesting and on-topic signs from which to build novel signing sequences. In this article, we detail the process of selecting an adequate range of signs and situations to be included in our corpus: from controlling discourse topic to including signs that can accept modified movements or handshapes, we describe how an avatar corpus has a different motivation than traditional signed language corpora.

Towards the Integration of Synthetic SL Animation with Avatars into Corpus Annotation Tools

Ralph Elliott, Javier Bueno, Richard Kennaway, John Glauert

We outline the main features of our synthetic virtual human sign language system, JASigning. We describe how we have extended its input notation, SiGML, to allow explicit control of performance time, and we describe our initial steps on the path to integrating virtual human sign language performance into annotation tools, where it may be compared with video depicting the corresponding real human performance.

Combining constraint-based models for Sign Language synthesis

Michael Filhol, Maxime Delorme, Annelies Braffort

The framework is that of Sign Language synthesis by virtual signers. In this paper, we present a sign generation system using a variety of input layers, separated on two sides: an anatomical side and a linguistic side. In a first part we suggest a way of implementing the flexibility required by Sign Languages into the system by using combinations of necessary and suficient constraints. The anatomical side of the input specifies all morphological and articulatory constraints that model the behaviour of a human skeleton, while the linguistic input specifies language constraints (lexical, grammatical, iconic...) that must be applied to the signer's body to utter the correct sign sequence. A second part explains how to combine all these parts of the input in a conjunction of constraints for each time frame of the animation. A point is made that conflicting constraints may be given but that they need be prioritised in order still to decide on acceptable solutions. A first idea of a global priority order is given to illustrate this issue.

SiS-Builder: A Sign Synthesis Support Tool

Theodore Goulas, Stavroula-Evita Fotinea, Eleni Efthimiou, Michalis Pissaris

Here we will present research work performed in the framework of the DICTA-SIGN project, closely related to Sign Language Synthesis and Animation and especially intended to cover for the need of creating lexical resources and evaluating them when performed by a signing avatar. Along these lines, a tool has been created to automatically generate SiGML transcriptions of any given HamNoSys string, as well as the relevant transcription file, by providing the HamNoSys characters of a sign.
Users are allowed to create a phrase of up to 4 sign units, by introducing the corresponding HamNoSys notations in a sequence of appropriate fields. The related xml script is then automatically created, allowing also for on demand storage on the server.
The here reported tool is web based, accessible by everyone, and allows users to interact with it without any special installations on the client side. Users may register, but this is not mandatory for use of the tool. Registered members can save their created scripts on the server in contrary to the non-registered ones. Online and offline manuals are available to the users as well as a FAQ facility.
As regards further tool functionalities, users are also enabled to see the HamNoSys notation of a sign chosen from a validated list of lemmas or by entering the raw xml script in the proper field. Users are able to switch between SiGML data and HamNoSys notations on an instant by just selecting the wished function. This way, it is possible to test/ see the results of creation of a lexical item, either by consulting the HamNoSys sequence, for those familiar with the HamNoSys syntax, or by animating the results through the avatar with the use of the SiGML script.
Users can create HamNoSys sequences by choosing the proper selection on the menu. This is a new feature, enhancing the tool's functionalities, added -upon completion of an evaluation phase- to allow users to create HamNoSys strings on line and then proceed with automatic creation of the corresponding SiGML scripts.
Furthermore, along with the HamNoSys manual characters, users may add non manual characters to the creation of the SiGML script. If needed, users may add more than one movement of a particular body part, i.e. head or shoulders, to make animation look closer to natural signing.
The final step is creation of the script. The user is then able to copy and paste the script to the avatar page to visualise the results of the created sequence, the latter been performed by the avatar.
Registered users are able to maintain/modify the data created by them.
The tool is based on open source Internet technologies for easy access and compatibility. Technologies that have been used are mostly PHP and JavaScript.

Requirements for a Signing Avatar

Vince Jennings, Ralph Elliott, Richard Kennaway, John Glauert

We present the technical specification for an avatar that is compliant with Animgen, the synthetic signing engine used at the University of East Anglia for generating deaf signing animations. The specification will include both the basic definition required for any standard animating avatar, and the additional parameters that Animgen requires to generate signing. Avatars compatible with Animgen are created using the ARPToolkit, an application developed at UEA that has a plug-in architecture for tools that are used for rigging an avatar mesh for animation. The toolkit also generates the additional data needed by Animgen for each avatar.

New features in synthesis of sign language addressing non-manual component

Zdeněk Krňoul

A sign language synthesis system converts previously noted signs into the computer animation. The animation is created using a specially designed 3D model of the human figure and algorithms transferring the sign to movements of the model. In principle the sign language contains both the non-manual component (shape and movement of hands) and the non-manual component (facial movements, etc.). Notation of the non-manual component was not yet sufficiently explored in terms of an automatic conversion to the animation. In the article we describe both notation methodology of the non-manual component and technical aspects for conversion of symbols to movements of the animation model. In addition an appropriate animation method for the 3D shape of face is assumed. The result is an extended notation supplementing notation of the manual component with the non-manual component. The extended notation preserves the feasibility of an automatic conversion and keeps the original level of generality. In connection with the methodology we present the notations of the basic types of non-manual components of the Czech sign language.

Language Resources for Spanish - Spanish Sign Language (LSE) Translation

Rubén San-Segundo, Verónica López, Raquel Martín, David Sánchez, Adolfo García

This paper describes the first Spanish-Spanish Sign Language (LSE) parallel corpus for language processing research focused on specific domains. This corpus includes more than 4,000 Spanish sentences translated into LSE. For every sentence, there is a video with the sign language representation. These sentences are focused on two restricted domains (two personal services): the renewal of the Identity Document and Driver's License. This corpus has been obtained with the collaboration of Local Government Offices where these services are provided. Over several weeks, the most frequent explanations (from the government employees) and the most frequent questions (from the user) were taken down.

This corpus also contains more than 800 sign descriptions in several sign-writing specifications: in glosses, SEA (Sistema de Escritura Alfabética) (Herrero, 2004), HamNoSys (Prillwitz et al, 1989), and SIGML (Zwiterslood et al, 2004) (a link to a text file with the SIGML description necessary for representing the sign with the eSIGN avatar: http://www.sign-lang.uni-hamburg.de/esign/). These descriptions have been generated with a modified version of the eSign Editor. This new version includes a graph to phoneme system for Spanish and a SEA- HamNoSys converter.

These language resources have been very important for the research project developed by Universidad Politécnica de Madrid and Fundación CNSE (Spanish Association of Deaf People) during the last three years (www.traduccionvozlse.es). The main target of this project has been to develop an advanced communication system for Deaf including two translation modules: The first one is a Spanish into LSE translation module. This first module is made up of a speech recognizer (for decoding the spoken utterance into a word sequence), a natural language translator (for converting a word sequence into a sequence of signs belonging to the sign language), and a 3D avatar animation module (for playing back the signs). The second module is a Spanish generator from LSE. This module consists of a visual interface (where a deaf person can specify a sequence of signs in sign-writing), a language translator (for generating the sequence of words in Spanish), and finally, a text to speech converter. The visual interface allows a sign sequence to be defined using several sign-writing alternatives.

For language translation, three technological alternatives were integrated and combined: an example-based strategy, a rule-based translation method and a statistical translator.

E Recognition & Translation

Sunday, May 23, 09:00-10:30

Chairperson: **Philippe Dreuw** Oral Session

Large Lexicon Project: American Sign Language Video Corpus and Sign Language Indexing/Retrieval Algorithms

Vassilis Athitsos, Carol Neidle, Stan Sclaroff, Joan Nash, Alexandra Stefan, Ashwin Thangali, Haijing Wang, Quan Yuan

When we encounter a word that we do not understand in a written language, we can look it up in a dictionary. However, looking up the meaning of an unknown sign in American Sign Language (ASL) is not nearly as straightforward. This paper describes progress in an ongoing project aiming to build a computer vision system that helps users look up the meaning of an unknown ASL sign. When a user encounters an unknown ASL sign, the user submits a video of that sign as a query to the system. The system evaluates the similarity between the query and video examples of all signs in the known lexicon, and presents the most similar signs to the user. The user can then look at the retrieved signs and determine if any of them matches the query sign.

An important part of the project is building a video database containing examples of a large number of signs. So far we have recorded at least two video examples for almost all of the 3,000 signs contained in the Gallaudet dictionary. Each video sequence is captured simultaneously from four different cameras, providing two frontal views, a side view, and a view zoomed in on the face of the signer. Our entire video dataset is publicly available on the Web.

Automatic computer vision-based evaluation of similarity between signs is a challenging task. In order to improve accuracy, we manually annotate the hand locations in each frame of each sign in the database. While this is a time-consuming process, this process incurs a one-time preprocessing cost that is invisible to the end-user of the system. At runtime, once the user has submitted the query video, the current version of the system asks the user to specify hand locations in the first frame, and then the system automatically tracks the location of the hands in the rest of the query video. The user can review and correct the hand location results. Every correction that the user makes on a specific frame is used by the system to further improve the hand location estimates in other frames.

Once hand locations have been estimated for the query video, the system evaluates the similarity between the query video and every sign video in the database. Similarity is measured using the Dynamic Time Warping (DTW) algorithm, a well-known algorithm for comparing time series. The performance of the system has been evaluated in experiments where 933 signs from 921 distinct sign classes are used as the dataset of known signs, and 193 signs are used as a test set. In those experiments, only a single frontal view was used for all test and training examples. For 68% of the test signs, the correct sign is included in the 20 most similar signs retrieved by the system.

In ongoing work, we are manually annotating hand locations in the remainder of our collected videos, so as to gradually incorporate more signs into our system. We are also investigating better ways for measuring similarity between signs, and for making the system more automatic, reducing or eliminating the need for the user to manually provide information to the system about hand locations.

Computer-based recognition of facial expressions in ASL: from face tracking to linguistic interpretation

Nicholas Michael, Carol Neidle, Dimitris Metaxas

Most research in the field of sign language recognition has focused on the manual component of signing, despite the fact that there is critical grammatical information expressed through facial expressions and head gestures. We, therefore, propose a novel framework for robust tracking and analysis of nonmanual behaviors, with an application to sign language recognition.
Our method uses computer vision techniques to track facial expressions and head movements from video, in order to recognize such linguistically significant expressions. The methods described here have relied crucially on the use of a linguistically annotated video corpus that is being developed, as the annotated video examples have served for training and testing our machine learning models. We apply our framework to continuous recognition of three classes of grammatical expressions, namely wh-questions, negative expressions, and topics.
Our method is signer-independent, utilizing spatial pyramids and Hidden Markov Models (HMMs) to model the temporal variations of facial shape and appearance.

Exploiting signed TV broadcasts for automatic learning of British Sign Language

Patrick Buehler, Mark Everingham, Andrew Zisserman

In this work, we will present several contributions towards automatic recognition of BSL signs from continuous signing video sequences. Specifically, we will address 3 main points: (i) automatic detection and tracking of the hands using a generative model of the image; (ii) automatic learning of signs from TV broadcasts of single signers, using only the supervisory information available from subtitles; and (iii) discriminative signer-independent sign recognition using automatically extracted training data from a single signer.
Our source material consists of many hours of video with continuous signing and corresponding subtitles recorded from BBC digital television. This is very challenging material for a number of reasons, including self-occlusions of the signer, self-shadowing, blur due to the speed of motion, and in particular the changing background.
Knowledge of the hand position and hand shape is a pre-requisite for automatic sign language recognition. We cast the problem of detecting and tracking the hands as inference in a generative model of the image, and propose a complete model which accounts for the positions and self-occlusions of the arms. Reasonable configurations are obtained by efficiently sampling from a pictorial structure proposal distribution. The results using our method exceed the state-of-the-art for the length and stability of continuous limb tracking.
Previous research in sign language recognition has typically required manual training data to be generated for each sign, e.g. a signer performing each sign in controlled conditions - a time-consuming and expensive procedure. We show that for a given signer, a large number of BSL signs can be learned automatically from TV broadcasts using the supervisory information available from subtitles broadcast simultaneously with the signing. We achieve this by modelling the problem as one of multiple instance learning. In this way we are able to extract the sign of interest from hours of signing footage, despite the very weak and "noisy" supervision from the subtitles.
Lastly, we will show how the automatic recognition of signs can be extended to multiple signers. Using automatically extracted examples from a single signer we train discriminative classifiers and show that these can successfully recognize signs for unseen signers. This demonstrates that our features (hand trajectory and hand shape) generalise well across different signers, despite the significant inter-personal differences in signing.

F1 Recognition

Sunday, May 23, 11:00-13:00

Chairperson: **Adam Schembri** Poster Session

Sign Language Recognition using Linguistically Derived Sub-units

Helen Cooper, Richard Bowden

This work proposes to learn linguistically-derived sub-unit classifiers for sign language. The responses of these classifiers can be combined by Markov models, producing efficient sign-level recognition. Tracking is used to create vectors of hand positions per frame as inputs for sub-unit classifiers learnt using AdaBoost. Grid-like classifiers are built around specific elements of the tracking vector to model the placement of the hands. Comparative classifiers encode the positional relationship between the hands. Finally, binary-pattern classifiers are applied over the tracking vectors of multiple frames to describe the motion of the hands. Results for the sub-unit classifiers in isolation are presented, reaching averages over 90%. Using a simple Markov model to combine the sub-unit classifiers allows sign level classification giving an average of 63%, over a 164 sign lexicon, with no grammatical constraints.

SignSpeak - Understanding, Recognition, and Translation of Sign Languages

Philippe Dreuw, Jens Forster, Yannick Gweth, Daniel Stein, Hermann Ney, Gregorio Martinez, Jaume Verges Llahi, Onno Crasborn, Ellen Ormel, Wei Du, Thomas Hoyoux, Justus Piater, Jose Miguel Moya Lazaro, Mark Wheatley

The SignSpeak project will be the first step to approach sign language recognition and translation at a scientific level already reached in similar research fields such as automatic speech recognition or statistical machine translation of spoken languages. Deaf communities revolve around sign languages as they are their natural means of communication. Although deaf, hard of hearing and hearing signers can communicate without problems amongst themselves, there is a serious challenge for the deaf community in trying to integrate into educational, social and work environments. The overall goal of SignSpeak is to develop a new vision-based technology for recognizing and translating continuous sign language to text. New knowledge about the nature of sign language structure from the perspective of machine recognition of continuous sign language will allow a subsequent breakthrough in the development of a new vision-based technology for continuous sign language recognition and translation. Existing and new publicly available corpora will be used to evaluate the research progress throughout the whole project.

DICTA-SIGN: Sign Language Recognition, Generation and Modelling with application in Deaf Communication

Eleni Efthimiou, Stavroula-Evita Fotinea, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, François Goudenove

Here we present the components and objectives of the EU funded project DICTA-SIGN. Dicta-Sign (http://www.dictasign.eu) is a three-year research project that involves the Institute for Language and Speech Processing, the University of Hamburg, the University of East Anglia, the University of Surrey, LIMSI/CNRS, the Université Paul Sabatier, the National Technical University of Athens, and WebSourd. It aims to improve the state of web-based communication for Deaf people by allowing the use of sign language in various human-computer interaction scenarios. It researches and develops recognition and synthesis engines for signed languages, aiming at a level of detail necessary for recognizing and generating authentic signing. In this context, Dicta-Sign aims at developing several technologies demonstrated via a sign language-aware Web 2.0.
Dicta-Sign supports four European sign languages: Greek. British, German, and French Sign Language and differs from previous work in that it aims to integrate tightly recognition, animation, and machine translation. All these components are informed by appropriate linguistic models from the ground up, including phonology, grammar, and non-manual features.
Expected outputs of the project include:

- A parallel multi-lingual corpus for four national sign languages - German, British, French and Greek (DGS, BSL, LSF and GSL respectively),
- A substantial multilingual dictionary of at least 1000 signs for each represented sign language,
- A continuous sign language recognition system that achieves significant improvement in terms of coverage and accuracy of sign recognition in comparison with current technology; furthermore this system will research the novel directions of multimodal sign fusion and signer adaptation,
- A language generation and synthesis component, covering in detail the role of manual, non-manual and placement within signing space,
- Annotation tools which incorporate these technologies providing access to the corpus and whose long term utility can be judged by the up-take by other sign language researchers,
- Three bidirectional integrated prototype systems which show the utility of the system components beyond the annotation tools application,
- A showcase demonstrator which exhibits how integration of the different components can support user communication needs.

Best Practice for Sign Language Data Collections Regarding the Needs of Data-Driven Recognition and Translation

Jens Forster, Daniel Stein, Ellen Ormel, Onno Crasborn, Hermann Ney

We propose best practices for gloss annotation of sign languages taking into account the needs of data-driven approaches to recognition and translation of natural languages. Furthermore, we provide reference numbers for several technical aspects for the creation of new sign language data collections. Most available sign language data collections are of limited use to data-driven approaches, because they focus on rare sign language phenomena, or lack machine-readable annotation schemes. Using a natural language processing point of view, we briefly discuss several sign language data collection, propose best practices for gloss annotation stemming from experience gained using two large scale sign language data collections, and derive reference numbers for several technical aspects from standard benchmark data collections for speech recognition and translation.

Video Analysis for Continuous Sign Language Recognition

Justus Piater, Thomas Hoyoux, Wei Du

The recognition of continuous, natural signing is very challenging due to the multimodal nature of the visual cues (fingers, lips, facial expressions, body pose, etc.), as well as technical limitations such as spatial and temporal resolution and unreliable depth cues. On the other hand, signing gestures are designed to be robustly discernible. We therefore argue in favor of an integrative approach to sign language recognition that aims to extract sufficient aggregate information for robust sign language recognition, even if many of the individual cues are unreliable. Our strategy to implement such an integrated system currently rests on two modules, for which we will show initial results. The first module uses active appearance models for detailed face tracking, allowing the quantification of facial expressions such as mouth and eye aperture and eyebrow raise. The second module is dedicated to hand tracking using color and appearance. A third module will be concerned with tracking upper-body articulated pose, linking the face to the hands for increased overall robustness.

Data-Driven Sub-Units, Modeling Structure of Multiple Cues for Continuous Sign Language Recognition

Vassilis Pitsikalis, Stavros Theodorakis, Petros Maragos

We investigate the automatic phonetic modeling of sign language based on phonetic sub-units, which are data driven and without any prior phonetic information. Visual processing is based on a probabilistic skin color model and a framewise geodesic active contour segmentation; occlusions are handled by a forward-backward prediction component leading finally to simple and effective region-based visual features. For sign-language modeling we propose a modeling structure for data-driven sub-unit construction. This utilizes the cue that is considered crucial to segment the signal into parts; at the same time we also classify the segments by implicitly assigning labels of Dynamic or Static type. This segmentation and classification step disentangles Dynamic from Static parts and allows us to employ for each type of segment the appropriate cue, modeling and clustering approach. The constructed Dynamic segments are exploited at the model level via hidden Markov models (HMMs). The Static segments are exploited via k-means clustering. Each Dynamic or Static part, exploits the appropriate cue related to the movement. We propose that the movement cues are normalized in order to be translation and scale invariant. We apply the proposed modeling for further combination of the movement trajectory individual cues. The proposed approaches are evaluated in recognition experiments conducted on the continuous sign language corpus of Boston University (BU-400) showing promising preliminary results.

Automatic sign language recognition, translation: a social approach

Marina Serrano, Jesús Gumiel, José M. Moya

This paper reviews the social needs of the deaf community and describes the mechanisms and/or technologies which would improve the quality of life of this collective. The base of this project is a pilot of teleinterpretation developed in Andalusia (Spain), and as results of the interaction with the users, have been found two investigation lines, the telephone communication, and the e-learning. This activities have a clearly defined technology needs by hearing impaired, and the existing solutions do not fix completely the problem, so they are a good scenario to implement an automatic sing language recognition system. The aim of the paper is demonstrate how to thanks to this technology, social barriers can be torn down, allowing equal access to those services that today are restrictive for the collective of deaf people.

RWTH-Phoenix: Analysis of the German Sign Language Weather Forecast Corpus

Daniel Stein, Jens Forster, Uwe Zelle, Philippe Dreuw, Hermann Ney

In this work, the recent additions to the RWTH-Phoenix corpus, a data collection of interpreted news announcement, are analysed. The corpus features videos, gloss annotation of German Sign Language and transcriptions of spoken German. The annotation procedure is reported, and the corpus statistics are discussed. We present automatic machine translation results for both directions, and discuss syntactically motivated enhancements.

SIGNUM Database: Video Corpus for Signer-Independent Continuous Sign Language Recognition

Ulrich von Agris, Karl-Friedrich Kraiss

Research in the field of continuous sign language recognition has not yet addressed the problem of interpersonal variance in signing. Applied to signer-independent tasks, current recognition systems show poor performance as their training bases upon corpora with an insufficient number of signers. In contrast to speech recognition, there is actually no benchmark which meets the requirements for signer-independent continuous sign language recognition. Because of this absence we created a new sign language corpus based on a vocabulary of 450 basic signs in German Sign Language (DGS). The corpus comprises 780 sentences each performed by 25 native signers of different sexes and ages. This database is now available for all interested researchers.

About Recognition of Sign Language Gestures

Alexander Voskresenskiy, Sergey Ilyin

A motion capture technique for implementing sign language dictionary is described. Problems of perception and recognition of gestures of Russian sign language in system of the automated sign language translation are discussed. The new approach to morphology of gestures and a method for separate gestures in sign statements are offered. The working definition for "text understanding" is offered.

F2 Translation

Sunday, May 23, 11:00-13:00

Chairperson: **Adam Schembri** Poster Session

Dealing with Sign Language Morphemes in Statistical Machine Translation

Guillem Massó, Toni Badia

The aim of this research is to establish the role of linguistic information in data-scarce statistical machine translation for sign languages using freely available tools. The main challenge in statistical machine translation is the scarcity of suitable data, and this problem becomes more pronounced in sign languages. The available corpora are small, usually not domain-specific, and their annotation conventions can vary considerably. Elaborating our own corpus is a very time-consuming task and the amount of data that we can obtain is even more reduced. Under these conditions, morpho-syntactic information helps to improve statistical machine translation results, but there are not linguistic processing tools for sign languages. We have managed to improve translations from Catalan to Catalan Sign Language by using factored models in an open source translation system with basic linguistic information such as the lemma or an annotation tier tag. Furthermore, this allows us to deal with sign language morphemes in a more systematic way.

Building Sign Language Corpora for Use in Machine Translation

Sara Morrissey, Harold Somers, Robert Smith, Shane Gilchrist, Sandipan Dandapat

In recent years data-driven methods of machine translation (MT) have overtaken rule-based approaches as the predominant means of automatically translating between languages. A pre-requisite for such an approach is a parallel corpus of the source and target languages. Technological developments in sign language (SL) capturing, analysis and processing tools now mean that SL corpora are becoming increasingly available. With transcription and language analysis tools being mainly designed and used for linguistic purposes, we describe the process of creating a multimedia parallel corpus specifically for the purposes of English to Irish Sign Language (ISL) MT. As part of our larger project on localisation, our research is focussed on developing assistive technology for patients with limited English in the domain of healthcare.
Focussing on the first point of contact a patient has with a GPís office, the medical secretary, we sought to develop a corpus from the dialogue between the two parties when scheduling an appointment. Throughout the development process we have created one parallel corpus in six different modalities from this initial dialogue, namely English speech, English text, ISL videos, Bangla text, HamNoSys transcription and SiGML code. In this paper we discuss the multi-stage process of the development of this parallel corpus as individual and interdependent entities, both for our own MT purposes and their usefulness in the wider MT and SL research domains.

Sign Language HPSG

Eva Safar, John Glauert

We present an overview of some relevant aspects of sign language synthesis in the ViSiCAST project, which might serve as a possible basis for the Dicta-Sign project. Dicta-Sign is a 3-year EU-funded project, that undertakes parallel corpus collection in different Sign Languages (SLs) and fundamental research and development of sign recognition and generation techniques in order to open up new potential applications for sign language users. One of the aims in Dicta-Sign is to find a model that is suitable for both recognition and generation. In this paper we revisit the main aspects of the synthesis techniques implemented in ALE Prolog using a sign language specific HPSG with the view for future changes needed. We briefly describe the HPSG feature structure and the rules and principles of the grammar, which cover important SL phenomena like mode, prodrop, plurals, classifiers and signing space.