SignSpeak – Understanding, Recognition, and Translation of Sign Languages

Philippe Dreuw[1], Jens Forster[1], Yannick Gweth[1], Daniel Stein[1], Hermann Ney[1], Gregorio Martinez[2], Jaume Verges Llahi[2], Onno Crasborn[3], Ellen Ormel[3], Wei Du[4], Thomas Hoyoux[4], Justus Piater[4], Jose Miguel Moya[5], and Mark Wheatley[6]

[1]RWTH, Aachen, Germany
dreuw@cs.rwth-aachen.de
[2]CRIC, Barcelona, Spain
gregorio.martinez@cric.cat
[3]RUN, Nijmegen, The Netherlands
o.crasborn@let.ru.nl
[4]ULg, Liege, Belgium
justus.piater@ulg.ac.be
[5]TID, Granada, Spain
jmml@tid.es
[6]EUD, Brussels, Belgium
mark.wheatley@eud.eu

Abstract

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.

1. Introduction

The SignSpeak project[1] is one of the first EU funded projects that tackles the problem of automatic recognition and translation of continuous sign language.

The overall goal of the SignSpeak project is to develop a new vision-based technology for recognizing and translating continuous sign language (i.e. provide Video-to-Text technologies), in order to provide new e-Services to the deaf community and to improve their communication with the hearing people.

The current rapid development of sign language research is partly due to advances in technology, including of course the spread of Internet, but especially the advance of computer technology enabling the use of digital video (Crasborn et al., 2007). The main research goals are related to a better scientific understanding and vision-based technological development for continuous sign language recognition and translation:

- understanding sign language requires better linguistic knowledge
- large vocabulary recognition requires more robust feature extraction methods and a modeling of the signs at a sub-word unit level
- statistical machine translation requires large bilingual annotated corpora and a better linguistic knowledge for phrase-based modeling and alignment

Therefore, the SignSpeak project combines innovative scientific theory and vision-based technology development by gathering novel linguistic research and the most advanced techniques in image analysis, automatic speech recognition (ASR) and statistical machine translation (SMT) within a common framework.

1.1. Sign Languages in Europe

Signed languages vary like spoken languages do: they are not mutually understandable, and there is typically one or more signed language in each country.

Although sign languages are used by a significant number of people, only a few member states of the European Union (EU) have recognized their national sign language on a *constitutional* level: Finland (1995), Slovak Republic (1995), Portugal (1997), Czech Republic (1998 & 2008), Austria (2005), and Spain (2007). The European Union of the Deaf (EUD)[2], a non-research partner in the SignSpeak project, is a European non-profit making organization which aims at establishing and maintaining EU level dialogue with the "hearing world" in consultation and cooperation with its member National Deaf Associations. The EUD is the only organization representing the interests of Deaf Europeans at European Union level. The EUD has 30 full members (27 EU countries plus Norway, Iceland & Switzerland), and 6 affiliated members (Croatia, Serbia, Bosnia and Herzegovina, Macedonia, Turkey & Israel). Their main goals are the recognition of the right to use an indigenous sign language, the empowerment through communication and information, and the equality in education and employment. In 2008, the EUD estimated about 650,000 Sign Language users in Europe, with about 7,000 official sign language interpreters, resulting in approximately 93 sign language users to 1 sign language interpreter (EUD, 2008; Wheatley and Pabsch, 2010). However, the number of sign language users might be much higher, as it is difficult to estimate an exact number – e.g. late-deafened or hard of hearing people who need interpreter services are not always counted as deaf people in these statistics.

[1]www.signspeak.eu

[2]www.eud.eu

1.2. Linguistic Research in Sign Languages

Linguistic research on sign languages started in the 1950s, with initial studies of Tervoort (Tervoort, 1953) and Stokoe (Stokoe et al., 1960). In the USA, the wider recognition of sign languages as an important linguistic research object only started in the 1970s, with Europe following in the 1980s. Only since 1990, sign language research has become a truly world-wide enterprise, resulting in the foundation of the Sign Language Linguistics Society in 2004[3]. Linguistic research has targeted all areas of linguistics, from phonetics to discourse, from first language acquisition to language disorders.

Vision-based sign language recognition has only been attempted on the basis of small sets of elicited data (Corpora) recorded under lab conditions (only from one to three signers and under controlled colour and brightness ambient conditions), without the use of spontaneous signing. The same restriction holds for much linguistic research on sign languages. Due to the extremely time-consuming work of linguistic annotation, studying sign languages has necessarily been confined to small selections of data. Depending on their research strategy, researchers either choose to record small sets of spontaneous signing which will then be transcribed to be able to address the linguistic question at hand, or native signer intuitions about what forms a correct utterance.

1.3. Research and Challenges in Automatic Sign Language Recognition

In (Ong and Ranganath, 2005; Y. Wu, 1999) reviews on research in sign language and gesture recognition are presented. In the following we briefly discuss the most important topics to build up a large vocabulary sign language recognition system.

1.3.1. Languages and Available Resources

Almost all publicly available resources, which have been recorded under lab conditions for linguistic research purposes, have in common that the vocabulary size, the types/token ratio (TTR), and signer/speaker dependency are closely related to the recording and annotation costs. Data-driven approaches with systems being automatically trained on these corpora do not generalize very well, as the structure of the signed sentences has often been designed in advance (von Agris and Kraiss, 2007), or offer small variations only (Dreuw et al., 2008b; Bungeroth et al., 2008), resulting in probably over-fitted language models. Additionally, most self-recorded corpora consists only of a limited number of signers (Vogler and Metaxas, 2001; Bowden et al., 2004).

For automatic sign language recognition, promising results have been achieved for continuous sign language recognition under lab conditions (von Agris and Kraiss, 2007; Dreuw et al., 2007a). In the recently very active research area of sign language recognition, a new trend towards broadcast news or weather forecast news can be observed. The problem of aligning an American Sign Language (ASL) sign with an English text subtitle is considered in (Farhadi and Forsyth, 2006). In (Buehler et al., 2009; Cooper and Bowden, 2009), the goal is to automatically learn a large number of British Sign Language (BSL) signs from TV broadcasts. Due to limited preparation time of the interpreters, the grammatical differences between "real-life" sign language and the sign language used in TV broadcast (being more close to Signed Exact English (SEE)) are often significant. Even if the performances of the automatic learning approaches presented in those works are still quite low, they represent an interesting approach for further research.

[3] www.slls.eu

1.3.2. Environment Conditions and Feature Extraction

Further difficulties for such sign language recognition frameworks arise due to different environment assumptions. Most of the methods developed assume closed-world scenarios, e.g. simple backgrounds, special hardware like data gloves, limited sets of actions, and a limited number of signers, resulting in different problems in sign language feature extraction or modeling.

1.3.3. Modeling of the Signs

In continuous sign language recognition, as well as in speech recognition, coarticulation effects have to be considered. One of the challenges in the recognition of continuous sign language on large corpora is the definition and modelling of the basic building blocks of sign language. The use of whole-word models for the recognition of sign language with a large vocabulary is unsuitable, as there is usually not enough training material available to robustly train the parameters of the individual word models. A suitable definition of sub-word units for sign language recognition would probably alleviate the burden of insufficient data for model creation.

In ASR, words are modelled as a concatenated sub-word units. These sub-word units are shared among the different word-models and thus the available training material is distributed over all word-models. On the one hand, this leads to better statistical models for the sub-word units, and on the other hand it allows to recognize words which have never been seen in the training procedure using lexica. According to the *linguistic* work on sign language by Stokoe (Stokoe et al., 1960), a phonological model for sign language can be defined, dividing signs into their four constituent visemes, such as the hand shapes, hand orientations, types of hand movements, and body locations at which signs are executed. Additionally, non-manual components like facial expression and body posture are used. However, no suitable decomposition of words into sub-word units is currently known for the purposes of a large vocabulary sign language *recognition* system (e.g. a grapheme-to-phoneme like conversion and use of a pronunciation lexicon).

The most important of these problems are related to the lack of generalization and overfitting systems (von Agris and Kraiss, 2007), poor scaling (Buehler et al., 2009; Cooper and Bowden, 2009), and unsuitable databases for mostly data driven approaches (Dreuw et al., 2008b).

1.4. Research and Challenges in Statistical Machine Translation of Sign Languages

While the first papers on sign language translations only date back to roughly a decade (Veale et al., 1998) and typically employed rule-based systems, several research groups have recently focussed on data-driven approaches. In (Stein et al., 2006), a SMT system has been developed for German and German sign language in the domain weather reports. Their work describes the addition of pre- and post-processing steps to improve the translation for this language pairing. The authors of (Morrissey and Way, 2005) have explored example-based MT approaches for the language pair English and sign language of the Netherlands with further developments being made in the area of Irish sign language. In (Chiu et al., 2007), a system is presented for the language pair Chinese and Taiwanese sign language. The optimizing methodologies are shown to outperform a simple SMT model. In the work of (San-Segundo et al., 2006), some basic research is done on Spanish and Spanish sign language with a focus on a speech-to-gesture architecture.

2. Speech and Sign Language Recognition

Automatic speech recognition (ASR) is the conversion of an acoustic signal (sound) into a sequence of written words.

Due to the high variability of the speech signal, speech recognition – outside lab conditions – is known to be a hard problem. Most decisions in speech recognition are interdependent, as word and phoneme boundaries are not visible in the acoustic signal, and the speaking rate varies. Therefore, decisions cannot be drawn independently but have to be made within a certain context, leading to systems that recognize whole sentences rather than single words.

One of the keys idea in speech recognition is to put all ambiguities into probability distributions (so called stochastic knowledge sources, see Figure 1). Then, by a stochastic modelling of the phoneme and word models, a pronunciation lexicon and a language model, the free parameters of the speech recognition framework are optimized using a large training data set. Finally, all the interdependencies and ambiguities are considered jointly in a search process which tries to find the best textual representation of the captured audio signal. In contrast, rule-based approaches try to solve the problems more or less independently.

In order to design a speech recognition system, four crucial problems have to be solved:

1. preprocessing and feature extraction of the input,
2. specification of models and structures for the words to be recognized,
3. learning of the free model parameters from the training data, and
4. search of the maximum probability over all models during recognition (see Figure 1).

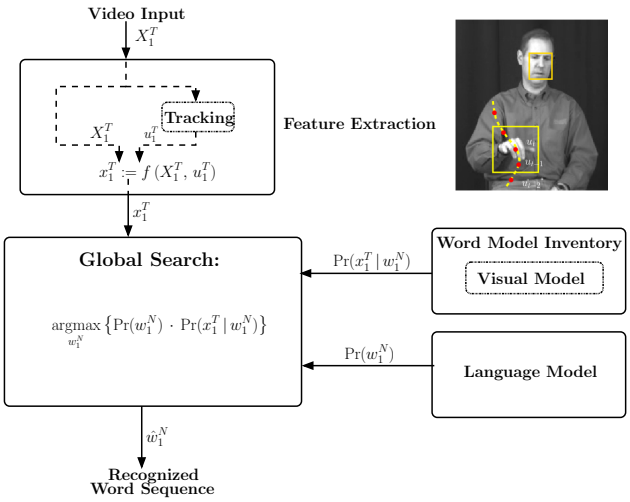

Figure 1: Sign language recognition system overview

2.1. Differences Between Spoken Language and Sign Language

Main differences between spoken language and sign language are due to linguistic characteristics such as simultaneous facial and hand expressions, references in the virtual signing space, and grammatical differences as explained more detailed in (Dreuw et al., 2008c):

Simultaneousness: Major issue in sign language recognition compared to speech recognition – a signer can use different communication channels (facial expression, hand movement, and body posture) in parallel.

Signing Space: Entities like persons or objects can be stored in a 3D body-centered space around the signer, by executing them at a certain location and later just referencing them by pointing to the space – the challenge is to define a model for spatial information handling.

Coarticulation and Epenthesis: In continuous sign language recognition, as well as in speech recognition, coarticulation effects have to be considered. Due to location changes in the 3D signing space, we also have to deal with the movement epenthesis problem (Vogler and Metaxas, 2001; Yang et al., 2007). Movement epenthesis refers to movements which occur regularly in natural sign language in order to move from the end state of one sign to the beginning of the next one. Movement epenthesis conveys no meaning in itself but contributes phonetic information to the perceiver.

Silence: opposed to automatic speech recognition, where the energy of the audio signal is usually used for the silence detection in the sentences, new spatial features and models will have to be defined for silence detection in sign language recognition. Silence cannot be detected by simply analyzing motion in the video, because words can be signed by just holding a particular posture in the signing space over time. Further, the rest position of the hand(s) may be somewhere in the signing space.

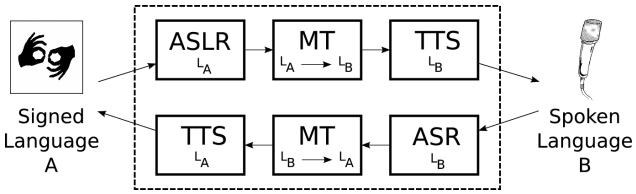

Figure 2: Complete six components-engine necessary to build a Sign-Language-to-Spoken-Language system (components: automatic sign language recognition (ASLR), automatic speech recognition (ASR), machine translation (MT), and text-to-speech/sign (TTS))

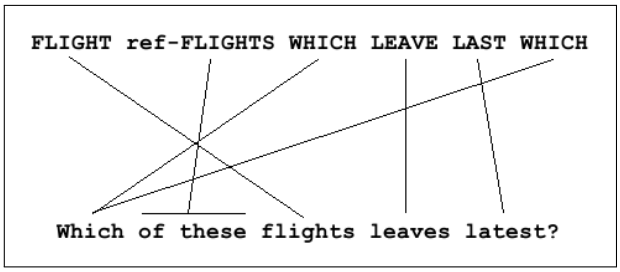

Figure 3: Reorderings are learned by a statistical machine translation system.

3. Towards a Sign-Language-to-Spoken-Language Translation System

The interpersonal communication problem between signer and hearing community could be resolved by building up a new communication bridge integrating components for sign-, speech-, and text-processing. To build a Sign-Language-to-Spoken-Language translator for a new language, a six component-engine must be integrated (see Figure 2), where each component is in principle language independent, but requires language dependent parameters/models. The models are usually automatically trained but require large annotated corpora.

The ASLR recognition is only the first step of a sign-language to spoken-language system. The intermediate representation of the recognized signs is further processed to create a spoken language translation. Statistical machine translation (MT) is a data-based translation method that was initially inspired by the so-called noisy-channel approach: the source language is interpreted as an encryption of the target language, and thus the translation algorithm is typically called a decoder. In practice, statistical machine translation often outperforms rule-based translation significantly on international translation challenges, given a sufficient amount of training data. As proposed in (Stein et al., 2007), a statistical machine translation system is used in SignSpeak to automatically transfer the meaning of a source language sentence into a target language sentence.

As mentioned above, statistical machine translation requires large bilingual annotated corpora. This is extremely important in order to train word reorderings or translate pointing references to the signing space (c.f. Figure 3). As reported in (Dreuw et al., 2007b), novel hand and face tracking features (see Figure 5) will be analyzed and integrated into the SignSpeak translation framework.

In SignSpeak, a theoretical study will be carried out about how the new communication bridge between deaf and hearing people could be built up by analyzing and adapting the ASLR and MT components technologies for sign language processing. The problems described in Section 2. will mainly be tackled by

- analysis of linguistic markers for sub-units and sentence boundaries,
- head and hand tracking of the dominant and non-dominant hand,
- facial expression and body posture analysis,
- analysis of linguistically- and data-driven sub-word units for sign modeling,
- analysis of spatio-temporal across-word modeling,
- signer independent recognition by pronunciation modeling, language model adaptation, and speaker adaptation techniques known from ASR
- contextual and multi-modal translation of sign language by an integration of tracking and recognition features into the translation process

Once the different modules are integrated within a common communication platform, the communication could be handled over 3G phones, media center TVs, or video telephone devices. The following sign language related application scenarios would be possible:

- e-learning of sign language
- automatic transcription of video e-mails, video documents, or video-SMS
- video subtitling

3.1. Impact on Other Industrial Applications

The novel features of such systems provide new ways for solving industrial problems. The technological breakthrough of SignSpeak will clearly have an impact on other applications fields:

Improving human-machine communication by gesture: vision-based systems are opening new paths and applications for human-machine communication by gesture, e.g. Play Station's EyeToy or Microsoft Xbox's Natal Project[4], which could be interesting for physically disabled individuals or even blind people as well.

Medical sector: new communication methods by gesture are being investigated to improve the communication between the medical staff, the computer, and other electronic equipments. Another application in this sector is related to web- or video-based *e-Care / e-Health* treatments, or an auto-rehabilitation system which makes the guidance process to a patient during the rehabilitation exercises easier.

[4] www.xbox.com/en-US/live/projectnatal/

Surveillance sector: person detection and recognition of body parts or dangerous objects, and their tracking within video sequences or in the context of quality control and inspection in manufacturing sectors.

4. Available and New Resources Within SignSpeak

All databases presented in this section are either freely available or can be purchased. Depending on the tasks and progress within the SignSpeak project, the focus will be shifted to one of the following databases briefly described in this section. Examples images showing the different recording conditions are shown for each database in Figure 4, where Table 1 gives an overview how the different corpora can be used for evaluation experiments.

4.1. CORPUS-NGT Database

The core of the SignSpeak data will come from the Corpus-NGT[5] database. This 72 hour corpus of Sign Language of the Netherlands is the first large open access corpus for sign linguistics in the world. It presently contains recordings from 92 different signers, mirroring both the age variation and the dialect variation present in the Dutch Deaf community (Crasborn et al., 2008).

For the SignSpeak project, the limited gloss annotations that were present in the first release of 2008 have been considerably expanded, and sentence-level translations have been added. Furthermore, more than 3000 frames will be annotated to evaluate hand and head tracking algorithms.

4.2. Boston Recordings

All databases presented in this section are freely available for further research in linguistics[6] and recognition[7]. The data were recorded by Boston University, the database subsets were defined at the RWTH Aachen University in order to build up benchmark databases (Dreuw et al., 2008b) that can be used for the automatic recognition of isolated and continuous sign language, respectively.

The RWTH-BOSTON-50 database was created for the task of isolated sign language recognition (Zahedi et al., 2006). It has been used for nearest-neighbor leaving-one-out evaluation of isolated sign language words.

The RWTH-BOSTON-104 has been used successfully for continuous sign language recognition experiments (Dreuw et al., 2007a). For the evaluation of hand tracking methods in sign language recognition systems, the database has been annotated with the signers' hand and head positions. More than 15.000 frames in total are annotated and are freely available[8].

For the task of sign language recognition and translation, promising results on the publicly available benchmark database RWTH-BOSTON-104 have been achieved for automatic sign language recognition (Dreuw et al., 2007a) and translation (Dreuw et al., 2008c; Dreuw et al., 2007b) that can be used as baseline reference for other researchers. However, the preliminary results on the larger RWTH-BOSTON-400 database show the limitations of the proposed framework and the need for better visual features, models, and corpora (Dreuw et al., 2008b).

4.3. Phoenix Weather Forecast Recordings

The RWTH-PHOENIX database with German sign language annotations of weather-forecast news has been first presented in (Stein et al., 2006) for the purpose of sign language translation (referred to as RWTH-PHOENIX-v1.0 in this work). It consists of about 2000 sentences, 9.000 running words, with a vocabulary size of about 1700 signs. Although the database is suitable for recognition experiments, the environment conditions in the first version cause problems in robust feature extraction such as hand tracking (see also Figure 4). During the SignSpeak project, a new release RWTH-PHOENIX-v2.0 will be recorded and annotated to meet the demands described in Section 5.. Due to the easier environment conditions in the RWTH-PHOENIX-v2.0 version (see also Figure 4), promising feature extraction and recognition results are expected.

4.4. The ATIS Sign Language Corpus

The ATIS Irish sign language database (ATIS-ISL) has been presented in (Bungeroth et al., 2008), and is suitable for recognition and translation experiments. The Irish sign language corpus formed the first translation into sign language of the original ATIS data. The sentences from the original ATIS corpus are given in written English as a transcription of the spoken sentences. The database as used in (Stein et al., 2007) contains 680 sentences with continuous sign language, has a vocabulary size of about 400 signs, and contains several speakers. For the SignSpeak project, about 600 frames have been annotated with hand and head positions to be used in tracking evaluations.

4.5. SIGNUM Database

The SIGNUM database[9] has been first presented in (von Agris and Kraiss, 2007) and contains both isolated and continuous utterances of various signers. This German sign language database is suitable for signer independent continuous sign language recognition tasks. It consists of about 33k sentences, 700 signs, and 25 speakers, which results in approximately 55 hours of video material.

5. Experimental Results and Requirements

In order to build a Sign-Language-to-Spoken-Language translator, reasonably sized corpora have to be created for the data-driven approaches. For a limited domain speech recognition task (Verbmobil II) as e.g. presented in (Kanthak et al., 2000), systems with a vocabulary size of up to 10k words have to trained with at least 700k words to obtain a reasonable performance, i.e. about 70 observations per vocabulary entry. Similar values must be obtained for a limited domain translation task (IWSLT) as e.g. presented in (Mauser et al., 2006).

[5] www.corpusngt.nl

[6] http://www.bu.edu/asllrp/

[7] http://www-i6.informatik.rwth-aachen.de/aslr/

[8] www-i6.informatik.rwth-aachen.de/~dreuw/database.php

[9] http://www.phonetik.uni-muenchen.de/forschung/Bas/SIGNUM/

Figure 4: Example images from different corpora used in SignSpeak (f.l.t.r.): Corpus-NGT, RWTH-BOSTON, RWTH-PHOENIX v1.0 and v2.0, ATIS-ISL, and SIGNUM

Table 1: Sign language corpora used within SignSpeak and their application areas

Corpus	Evaluation Area			
	Isolated Recognition	Continuous Recognition	Tracking	Translation
Corpus-NGT	✓	✓	✓	✓
RWTH-BOSTON-50	✓	✖	✖	✖
RWTH-BOSTON-104	✖	✓	✓	✖
RWTH-BOSTON-400	✖	✓	✖	✖
RWTH-PHOENIX-v1.0	✓	✓	✖	✓
RWTH-PHOENIX-v2.0	✖	✓	✖	✓
ATIS-ISL	✖	✓	✓	✓
SIGNUM	✓	✓	✖	✖

Similar corpora statistics can be observed for other ASR or MT tasks. The requirements for a sign language corpus suitable for recognition and translation can therefore be summarized as follows:

- annotations should be domain specific (i.e. broadcast news, or weather forecasts, etc.)
- for a vocabulary size smaller than 4k words, each word should be observed at least 20 times
- the singleton ratio should ideally stay below 40%

Existing corpora should be extended to achieve a good performance w.r.t. recognition and translation (Forster et al., 2010). During the SignSpeak project, the existing RWTH-PHOENIX corpus (Stein et al., 2006) and Corpus-NGT (Crasborn et al., 2008) will be extended to meet these demands (see Table 2). Novel facial features (Piater et al., 2010) developed within the SignSpeak project are shown in Figure 5 and will be analyzed for continuous sign language recognition.

6. Acknowledgements

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

7. References

R. Bowden, D. Windridge, T. Kadir, A. Zisserman, and M. Brady. 2004. A Linguistic Feature Vector for the Visual Interpretation of Sign Language. In *ECCV*, volume 1, pages 390–401, May.

Patrick Buehler, Mark Everingham, and Andrew Zisserman. 2009. Learning sign language by watching TV (using weakly aligned subtitles). In *IEEE CVPR*, Miami, FL, USA, June.

Jan Bungeroth, Daniel Stein, Philippe Dreuw, Hermann Ney, Sara Morrissey, Andy Way, and Lynette van Zijl. 2008. The ATIS Sign Language Corpus. In *LREC*, Marrakech, Morocco, May.

Y.-H. Chiu, C.-H. Wu, H.-Y. Su, and C.-J. Cheng. 2007. Joint Optimization of Word Alignment and Epenthesis Generation for Chinese to Taiwanese Sign Synthesis. *IEEE Trans. PAMI*, **29**(1):28–39.

Helen Cooper and Richard Bowden. 2009. Learning Signs from Subtitles: A Weakly Supervised Approach to Sign Language Recognition. In *IEEE CVPR*, Miami, FL, USA, June.

Onno Crasborn, Johanna Mesch, Dafydd Waters, Annika Nonhebel, Els van der Kooij, Bencie Woll, and Brita Bergman. 2007. Sharing sign langauge data online. Experiences from the ECHO project. *International Journal of Corpus Linguistics*, 12(4):537–564.

Onno Crasborn, Inge Zwitserlood, and Johan Ros. 2008. Corpus-NGT. An open access digital corpus of movies with annotations of Sign Language of the Netherlands. Technical report, Centre for Language Studies, Radboud University Nijmegen. `http://www.corpusngt.nl`.

P. Dreuw, D. Rybach, T. Deselaers, M. Zahedi, and H. Ney. 2007a. Speech Recognition Techniques for a Sign Language Recognition System. In *ICSLP*, Antwerp, Belgium, August. Best paper award.

P. Dreuw, D. Stein, and H. Ney. 2007b. Enhancing a Sign Language Translation System with Vision-Based Features. In *Intl. Workshop on Gesture in HCI and Simulation 2007*, pages 18–19, Lisbon, Portugal, May.

Philippe Dreuw, Jens Forster, Thomas Deselaers, and Hermann Ney. 2008a. Efficient Approximations to Model-based Joint Tracking and Recognition of Continuous Sign Language. In *IEEE International Conference Automatic Face and Gesture Recognition*, Amsterdam, The Netherlands, September.

Philippe Dreuw, Carol Neidle, Vassilis Athitsos, Stan

Table 2: Expected corpus annotation progress of the RWTH-PHOENIX and Corpus-NGT corpora in comparison to the limited domain speech (Verbmobil II) and translation (IWSLT) corpora.

	BOSTON-104	Phoenix		Corpus-NGT		Vermobil II	IWSLT
year	2007	2009	2011	2009	2011	2000	2006
recordings	201	78	400	116	300	-	-
running words	0.8k	10k	50k	30k	80k	700k	200k
vocabulary size	0.1k	0.6k	**2.5k**	3k	$>$ **5k**	10k	10k
T/T ratio	8	15	**20**	10	$<$ **20**	70	20
Performance	11% WER	-	-	-	-	15% WER	40% TER
	(Dreuw et al., 2008a)	-	-	-	-	(Kanthak et al., 2000)	(Mauser et al., 2006)

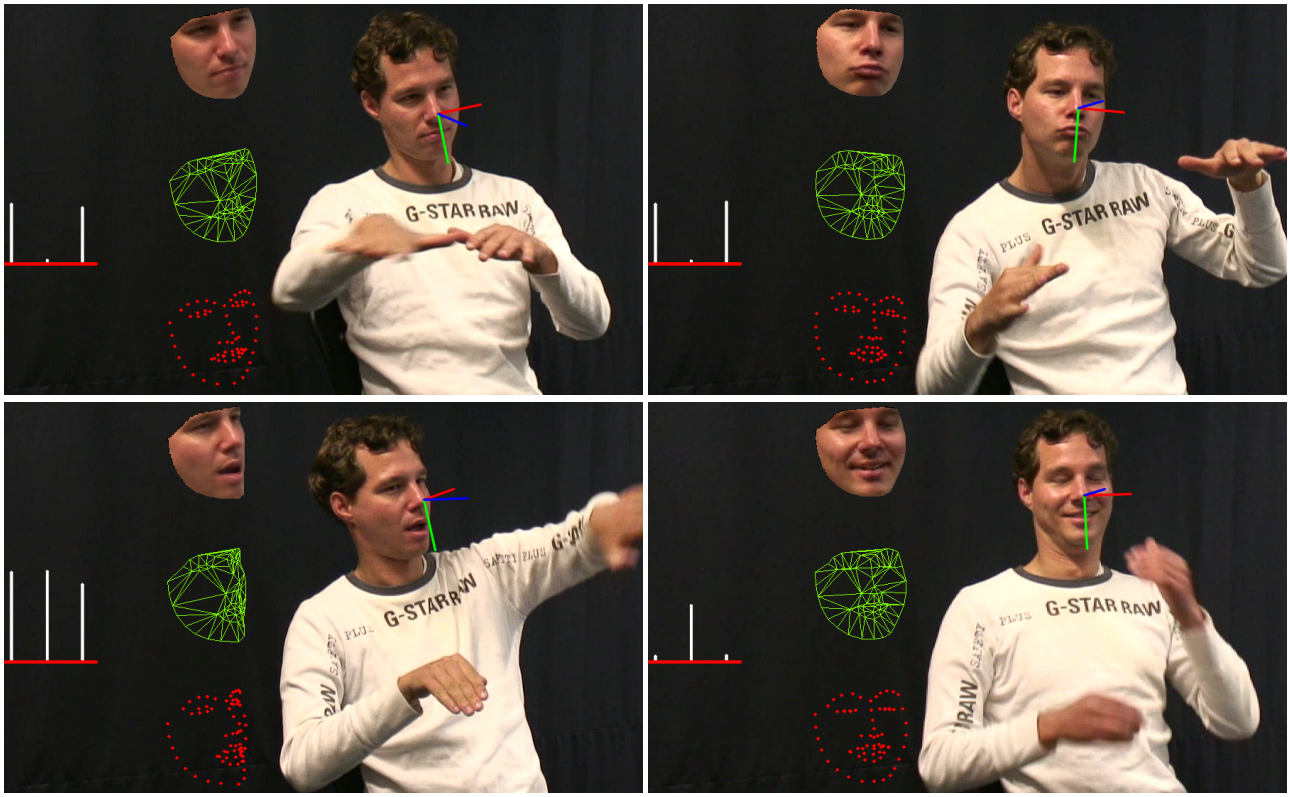

Figure 5: Novel facial feature extraction on the Corpus-NGT database (f.l.t.r.): three vertical lines quantify features like left eye aperture, mouth aperture, and right eye aperture; the extraction of these features is based on a fitted face model, where the orientation of this model is shown by three axis on the face: red is X, green is Y, blue is Z, origin is the nose tip.

Sclaroff, and Hermann Ney. 2008b. Benchmark Databases for Video-Based Automatic Sign Language Recognition. In *LREC*, Marrakech, Morocco, May.

Philippe Dreuw, Daniel Stein, Thomas Deselaers, David Rybach, Morteza Zahedi, Jan Bungeroth, and Hermann Ney. 2008c. Spoken Language Processing Techniques for Sign Language Recognition and Translation. *Technology and Dissability*, 20(2):121–133, June.

EUD. 2008. Survey about Sign Languages in Europe.

A. Farhadi and D. Forsyth. 2006. Aligning ASL for statistical translation using a discriminative word model. In *IEEE CVPR*, New York, USA, June.

Jens Forster, Daniel Stein, Ellen Ormel, Onno Crasborn, and Hermann Ney. 2010. Best Practice for Sign Language Data Collections Regarding the Needs of Data-Driven Recognition and Translation. In *4th LREC Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT)*, Malta, May.

Stephan Kanthak, Achim Sixtus, Sirko Molau, Ralf Schlüter, and Hermann Ney, 2000. *Fast Search for Large Vocabulary Speech Recognition*, chapter "From Speech Input to Augmented Word Lattices", pages 63–78. Springer Verlag, Berlin, Heidelberg, New York, July.

Arne Mauser, Richard Zens, Evgeny Matusov, Saša Hasan, and Hermann Ney. 2006. The RWTH Statistical Machine Translation System for the IWSLT 2006 Evaluation. In *IWSLT*, pages 103–110, Kyoto, Japan, November. Best Paper Award.

S. Morrissey and A. Way. 2005. An Example-based Approach to Translating Sign Language. In *Workshop in Example-Based Machine Translation (MT Summit X)*, pages 109–116, Phuket, Thailand, September.

S. Ong and S. Ranganath. 2005. Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning. *IEEE Trans. PAMI*, 27(6):873–891, June.

Justus Piater, Thomas Hoyoux, and Wei Du. 2010. Video Analysis for Continuous Sign Language Recognition. In

4th LREC Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT), Malta, May.

R. San-Segundo, R. Barra, L. F. D'Haro, J. M. Montero, R. Córdoba, and J. Ferreiros. 2006. A Spanish Speech to Sign Language Translation System for assisting deaf-mute people. In *ICSLP*, Pittsburgh, PA, September.

D. Stein, J. Bungeroth, and H. Ney. 2006. Morpho-Syntax Based Statistical Methods for Sign Language Translation. In *11th EAMT*, pages 169–177, Oslo, Norway, June.

D. Stein, P. Dreuw, H. Ney, S. Morrissey, and A. Way. 2007. Hand in Hand: Automatic Sign Language to Speech Translation. In *The 11th Conference on Theoretical and Methodological Issues in Machine Translation*, Skoevde, Sweden, September.

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

B. Tervoort. 1953. Structurele analyse van visueel taalgebruik binnen een groep dove kinderen.

T. Veale, A. Conway, and B. Collins. 1998. The Challenges of Cross-Modal Translation: English to Sign Language Translation in the ZARDOZ System. *Journal of Machine Translation*, 13, No. 1:81–106.

C. Vogler and D. Metaxas. 2001. A Framework for Recognizing the Simultaneous Aspects of American Sign Language. *CVIU*, 81(3):358–384, March.

U. von Agris and K.-F. Kraiss. 2007. Towards a Video Corpus for Signer-Independent Continuous Sign Language Recognition. In *Gesture in Human-Computer Interaction and Simulation*, Lisbon, Portugal, May.

Mark Wheatley and Annika Pabsch. 2010. Sign Language in Europe. In *4th LREC Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies (CSLT)*, Malta, May.

T.S. Huang Y. Wu. 1999. Vision-based gesture recognition: a review. In *Gesture Workshop*, volume 1739 of *LNCS*, pages 103–115, Gif-sur-Yvette, France, March.

Ruiduo Yang, Sudeep Sarkar, and Barbara Loeding. 2007. Enhanced Level Building Algorithm to the Movement Epenthesis Problem in Sign Language. In *CVPR*, MN, USA, June.

Morteza Zahedi, Philippe Dreuw, David Rybach, Thomas Deselaers, Jan Bungeroth, and Hermann Ney. 2006. Continuous Sign Language Recognition - Approaches from Speech Recognition and Available Data Resources. In *LREC Workshop on the Representation and Processing of Sign Languages: Lexicographic Matters and Didactic Scenarios*, pages 21–24, Genoa, Italy, May.