Extensions of the Sign Language Recognition and Translation Corpus RWTH-PHOENIX-Weather

Jens Forster, Christoph Schmidt, Oscar Koller, Martin Bellgardt, and Hermann Ney

Human Language Technology and Pattern Recognition
RWTH Aachen University, Germany
{forster,schmidt,koller,bellgardt,ney}@cs.rwth-aachen.de

Abstract

This paper introduces the RWTH-PHOENIX-Weather 2014, a video-based, large vocabulary, German sign language corpus which has been extended over the last two years, tripling the size of the original corpus. The corpus contains weather forecasts simultaneously interpreted into sign language which were recorded from German public TV and manually annotated using glosses on the sentence level and semi-automatically transcribed spoken German extracted from the videos using the open-source speech recognition system RASR. Spatial annotations of the signers' hands as well as shape and orientation annotations of the dominant hand have been added for more than 40k respectively 10k video frames creating one of the largest corpora allowing for quantitative evaluation of object tracking algorithms. Further, over 2k signs have been annotated using the SignWriting annotation system, focusing on the shape, orientation, movement as well as spatial contacts of both hands. Finally, extended recognition and translation setups are defined, and baseline results are presented.

Keywords: Sign Language, Corpus, Automatic Recognition and Translation

1. Introduction

Sign languages are the native languages of the deaf and partly of the hard-of-hearing communities worldwide. As deaf communities form a minority in their respective countries and the hearing majority typically has hardly any signing skills, there is an interest to build automatic systems to ease the communication between both groups. Two crucial building blocks to achieve this are automatic sign language recognition and automatic translation of sign languages. In sign language recognition, an automatic system extracts sign language from a video and represents the signs in a written intermediate notation, which then is translated into a written text of a spoken language. State of the art speech recognition and translation systems employ statistical models to facilitate recognition or translation itself. As such, statistical models require large amounts of labelled data to learn from in order to robustly generalize to new and unseen data. In terms of available data, sign languages are under-resourced languages with corpora that are typically recorded for linguistic research, not providing the type/token ratios needed for statistical natural language processing. Typically, this kind of data differs significantly from the real language encountered outside the research lab. One concept used particular in linguistic corpora is the concept of staged communicative events trying to elicit special aspects of sign language communication. Staged communication events focus on the interaction between one or more signers. While this makes the language encountered more natural, it raises automatic processing to a difficulty level not yet in focus of the machine learning and pattern recognition community.

To address the issue of under-resourced sign language corpora, (Forster et al., 2012) created the RWTH-PHOENIX-Weather 2012 corpus for German sign language (DGS) by recording and annotating "real-life" sign language footage aired in 2009 and 2010 by the German public TV station "PHOENIX" in the context of weather forecasts. Over the last two years, this corpus has been significantly enlarged by adding data from the years 2011 to 2013 with a bilingual annotation in DGS glosses and written German, creating the RWTH-PHOENIX-Weather 2014 corpus. Additionally, annotations for the spatial positions of the hands and face of a signer for over 40 000 video frames as well as annotations for hand shapes and orientations on the frame level and on the sign level using SignWriting (Sutton and Writing, 2000) have been added to the corpus.

In the following, Section 2. discusses related work. The extensions to the RWTH-PHOENIX-Weather 2012 w.r.t. recognition and translation setups are described in Sections 3. and 4.. Preliminary recognition and translation results using these setups are presented in Section 5.. The paper is concluded in Section 6..

2. Related Work

Currently publicly available, video-based sign language corpora fall in one of three categories depending on the scientific community they originated from.

First, there are corpora intended as video-based lexica for sign languages allowing to track and analyze changes in the vocabulary of sign languages from a linguistic point-of-view. "The American Sign Language Lexicon Video Dataset" (Neidle and Vogler, 2012) forms such a lexicon for American sign language (ASL), containing more than 3000 signs in multiple video views. The AUSLAN SignBank project[1] provides annotations on a variety of linguistic levels for 357 videos of Australian sign language.

Second, there are corpora intended for linguistic research on isolated signs and continuous sign language allowing to tackle questions like appearance of dialectic variances, differences in pronunciation and sentence structures. Typically, such corpora are created under lab-conditions focussing on certain aspect of sign languages. Corpus NGT (Crasborn and Zwitserlood, 2008) contains 12 hours of signing in upper-body and front view totalling 64 000 annotated glosses. Since 2008 the corpus has been extended

[1] www.auslan.org.au

by translations into various spoken languages. (Rutkowski et al., 2013) created a corpus for Polish sign language containing about 300h of video footage of 80 deaf signers performing predefined language tasks. The CopyCat corpus (Zafrulla et al., 2010) covers ASL spoken by children in 420 phrases formed from a vocabulary of 19 signs. For further reference, the University of Hamburg, Germany, created a summary on available linguistic sign language corpora[2].

Third, there are corpora either explicitly created or adapted for natural language processing and/or computer vision tasks. In contrast to the linguistic resources, corpora created for natural language processing tasks spot smaller vocabularies of a couple of hundred signs instead of thousands, higher type/token ratios and focus on a small number of closed language domains. The overall goal is to provide minimum statistics to allow for robust training of statistical models while refraining from focusing on special aspects of sign languages such as classifier signs. (Dreuw et al., 2010) give an overview on such corpora. Included in this survey are the RWTH-BOSTON corpora originally created for linguistic research at Boston University and adapted for pattern recognition purposes by RWTH Aachen University featuring multiple signers and up to 7 768 running glosses with a vocabulary size of 483 glosses. (Efthimiou et al., 2012) and (Braffort et al., 2010) present corpora for isolated and continuous sign language recognition for German, Greek, British and French sign language created in the course of the Dicta-Sign[3] project. The corpora include sign language videos shot in high-definition in frontal and side view under controlled lab-conditions. Similar to Corpus NGT, the Dicta-Sign corpora contain bird's eye views of the signers allowing for the study of hand movements in the signing space with regard to the distance from the upper-body of the respective signer. The SIGNUM corpus (von Agris et al., 2008) has been explicitly created for pattern recognition purposes foregoing linguistic considerations and consists of 25 signers and nearly 14 000 running glosses in DGS with a vocabulary of 450 glosses.

Finally, in terms of bilingual corpora, (San-Segundo et al., 2010) have built a Spanish-Spanish Sign Language corpus in the domain of identity document and driver's licence renewal using gloss notation. The corpus features 1360 sentences in Spanish Sign Language, and several Spanish translation variants have been produced for each sentence, leading to a total of 4080 sentence pairs. (Almohimeed et al., 2010) describe a 230 sentence corpus for Arabic sign language and written Arabic with a vocabulary of 710 signs. In a similar fashion, (Morrissey et al., 2010) created a bilingual English/Irish sign language corpus of 350 sentences.

3. RWTH-PHOENIX-Weather Corpus

The original version of the corpus, which we refer to as the RWTH-PHOENIX-Weather 2012 (Forster et al., 2012), has been created in the course of the SignSpeak[4] project. It consists of 190 weather forecasts aired from 2009 to 2010 by the German public TV station "PHOENIX". Due to the rather compact domain of weather forecasting, the overall vocabulary of the corpus is limited except for named entities such as rivers or places.

Over the last two years, we continued to record and annotate weather forecasts from the period 2011 to early 2013 following the original annotation scheme (Forster et al., 2012). Like the previous recordings, the new videos have not been recorded under lab conditions, but the lighting conditions and the positioning of the signer in front of the camera are rather controlled in the TV studio. All videos have a resolution of 210×260 pixel and 25 interlaced frames per second (FPS). In the day and age of high-definition consumer-priced cameras the temporal and spatial resolution of the videos seem to be at odds with current technical possibilities. Unfortunately, we have no influence on the broadcast method used by the TV-station which is 720×576 pixel interlaced and which is typically scaled to 1024×576 pixel by TV devices without adding information to the video stream. Figure 1 shows an example frame from the original video stream broadcast by PHOENIX. The broadcast of the weather forecast is overlayed with the sign language interpreter leading to the afore mentioned spatial resolution of 210×260 pixels.

Figure 2 shows example images of all nine signers present in the corpus as well as their distribution.

One of the challenges of the RWTH-PHOENIX-Weather corpus are motion blur effects due to high signing speed

Figure 1: RWTH-PHOENIX-Weather Example of original video frame. The sign language interpreter is shown in an overlay on the right of the original video frame.

Figure 2: Example images and percentage of data performed by signer in RWTH-PHOENIX-Weather corpus. Top, left to right signers 1 to 5, bottom signers 6 to 9

[2]www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/sl-corpora.html

[3]www.dictasign.eu

[4]www.signspeak.eu

Table 1: Final Statistics of the RWTH-PHOENIX-Weather corpus for DGS and announcements in spoken German

	2012 DGS	2012 German	2014 DGS	2014 German
# signers	7		9	
# editions	190		645	
duration[h]	3.25		10.73	
# frames	293,077		965,940	
# sentences	1,980	2,640	6,861	8,767
# run. glosses	21,822	32,858	75,107	123,532
vocab. size	911	1,489	1,558	2,589
# singletons	537	525	586	531

and the low temporal resolution of 25 FPS. Furthermore, the sign language content of the videos is closer to the grammatical structure of German than in other scenarios because it is created by hearing interpreters under real-time constraints. This has an impact on the spatio-temporal organization of the sign language utterances represented in the corpus. It differs from the spatio-temporal organization seen in sign language uttered by deaf signers and in a sense differs from "real-life" sign language in the linguistic sense. As such, the RWTH-PHOENIX-Weather corpus is not intended for linguistic research but for the development of automatic sign language recognition and translation systems and lends itself to linguistic research only in areas investigating effects due to fast signing, speech artefacts, facial expressions and mouthings. Nevertheless, the language in the corpus is "real-life" in the sense of it being produced without constraints imposed on it in the context of an research lab and w.r.t. signing speed, facial expressions and mouthings. The issues regarding the spatio-temporal organization of the utterance from a linguistic point-of-view do not affect findings for automatic sign language recognition or translation because the used statistical models learn directly from data.

Using the ELAN[5] tool, the newly recorded weather forecasts have been annotated with

1. gloss sentences including sentence boundaries,
2. the utterances of the announcer in written German, annotated with the help of a speech recognition system.

The RWTH-PHOENIX-Weather 2014 corpus consists of 6 861 sentence in DGS, not counting annotations labeled as "<PAUSE>", and 75 107 running glosses. The overall vocabulary comprises 1 558 different glosses. Table 1 summarizes the overall statistics of the RWTH-PHOENIX-Weather 2014 corpus in comparison to the original RWTH-PHOENIX-Weather 2012 version.

To obtain the text spoken by the announcer, the open-source speech recognition system RASR (Sundermeyer et al., 2011; Tüske et al., 2013) was applied to the audio stream of the videos. The recognition output was then manually corrected by native German speakers to obtain the references used for the machine translation experiments. To further speed up the manual correction, automatic true-casing and punctuation recovery have been applied.

The corpus statistics for the translation corpus can be found in Table 1 in the columns titled "German". Note that the number of sentences differs from those presented for DGS. First, the glosses were annotated by two deaf experts, who also defined the segmentation of the glosses into sentences. The number of these segments does however not necessarily correspond to the number of spoken sentences. For the translation corpus, we therefore re-segmented the glosses into sentences such that they correspond to the announcements in spoken German.

The RWTH-PHOENIX-Weather 2012 corpus included spatial annotations of the hand palms and the nose tip for a subset of 266 signs covering seven signers and 39 691 frames in addition to the annotation of the signs in gloss notation and spoken German. This annotation allows for the evaluation of hand and face tracking systems on real-life data in the context of sign language recognition but does not allow for evaluating the effect of tracking accuracy on sign language recognition performance.

To address this shortcoming, all frames of the *Signer Dependent Subset for Signer03* (single signer for short) presented in (Forster et al., 2012) have been annotated with the position of the nose tip and of both hands. Due to this large-scale spatial annotation, the single signer subset can be used to evaluate tracking algorithms in the context of sign language recognition, allowing for detailed analysis of tracking errors on the frame level. Further, it allows to develop and evaluate features and algorithms independent of the need to automatically track the hands or the face of the signer.

Based on the annotation of the position of the right hand in the training part of the single signer setup, the shape and orientation of the hand has been annotated on the frame level using cropped images of the right hand of size 50×70 pixels. This annotation allows the evaluation of clustering algorithms in the context of sign language recognition intending to automatically find shared hand-shapes between signs. Because of the low temporal resolution of the videos and the high signing speed, many frames show motion blur effects and were excluded for the hand-shape and orientation annotation. From the 46 282 video frames in the training set, 31 640 frames were automatically excluded because the hand moved more than five pixels in Euclidean distance between consecutive frames according to the position annotation, 4 056 were excluded after manual inspection, and 266 frames were excluded because the hand was not visible in the frame, leaving 10 320 frames with hand-shape and orientation annotations.

The hand-shape and orientation annotation has the form of `OR1-OR2-SHAPE` and is annotated from the annotator's point-of-view. `OR1` indicates the orientation the metacarpus of the hand is facing (red arrow in Figure 3) and `OR2` indicates the direction of a line orthogonal to the back of the hand (green arrow in Figure 3).

Both orientations are quantized into the six classes *UP*, *DOWN*, *LEFT*, *RIGHT*, *FRONT*, and *BACK* where *FRONT* indicates that the orientation is facing towards the annotator. Each orientation spans an angle of 45° in either direc-

[5] `www.lat-mpi.eu/tools/elan`

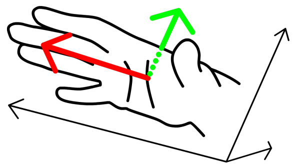

Figure 3: Illustration of the orientation annotation directions annotated: The direction the metacarpus is facing (red), as well as the direction of an vector, orthogonal to the back of the hand (green).

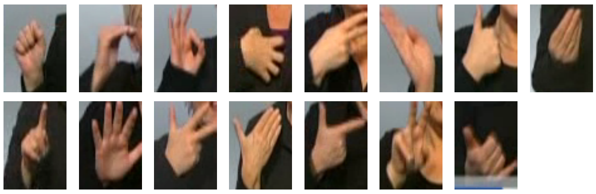

Figure 4: Example images of the 15 annotated handshapes: From left to right, top row: AHAND, CORNER, FHAND, GRAB, HHAND, NOSPREAD, ONE, PINCH ; bottom row: POINT, SPREAD, THREE, THUMBOUT, TWO, VHAND, YHAND

tion of the respective axis of an 3 dimensional Cartesian coordinate system as shown in Figure 3. Considering for example the orientation *UP*, the orientation spans an 45° angle on either side of the ordinate with regard to the abscissa for positive values on the ordinate.

The shape of the hand has been annotated according to 16 classes, one of them being a noise class. The 15 non-noise classes (see Figure 4) have been determined semi-automatically by first clustering the cropped images per orientation and then manually determining the classes across orientations. Clustering is performed by splitting the cropped images according to the manually annotated orientations `OR1` and `OR2` and then applying the expectation-maximization algorithm (Duda et al., 2001) using a single Gaussian mixture distribution until convergence for each orientation split resulting in a clustering result for each `OR1-OR2` combination. The resulting clusters contain at least 20 cropped images. Following the automatic clustering, clusters are manually cleaned leading to a number of clusters with less than 20 cropped images. Furthermore, visual similar clusters are manually joined. Finally, unique hand shapes are manually identified across orientations by annotators. Excluding orientation and shape combinations that occur less than ten times, the resulting annotation contains 98 classes (out of the possible 576 $(6 \cdot 6 \cdot 16)$) with the most frequent one occurring 800 times (see Table 2).

Furthermore, a subset of 2388 cut out signs (1391 from the train and 997 from the test set) has been annotated using the SignWriting annotation syten, for details on SignWriting refer to (Koller et al., 2013). The annotation comprises the shape and orientation of both hands (109 different classes), hand movement characterization in 3D and contact points between the hands and the body. Most signs manifest in a sequence of varying handshapes, as shown on the right of Figure 5. On average this yields 1.9 handshapes per sign.

Table 2: Ten most frequent orientation-shape classes in the *Signer Dependent Subset for Signer03* in the form `OR1-OR2-SHAPE`. For shape class information please see Figure 4 where `TOUT` is the abbreviation of `THUMBOUT`.

`UP-BACK-SPREAD`	800	`UP-LEFT-NOSPREAD`	471
`FRONT-UP-OTHER`	386	`FRONT-LEFT-AHAND`	337
`UP-LEFT-CORNER`	318	`UP-BACK-POINT`	315
`UP-BACK-OTHER`	311	`FRONT-LEFT-OTHER`	307
`UP-FRONT-TOUT`	290	`UP-LEFT-AHAND`	286

Figure 5: Example SignWriting annotations for glosses WEATHER (left) and 22 (right).

4. Evaluation Corpora

The RWTH-PHOENIX-Weather corpus allows for the creation and evaluation of automatic recognition systems for continuous and isolated sign language in a signer dependent, a multi-signer and a signer independent fashion as well as for the creation and evaluation of automatic translation systems for the language pair DGS and German. The spatial annotations allow for the evaluation of hand tracking and face tracking systems as well as for face detection systems. Furthermore, the new annotation of the hand and nose tip positions for every frame of the single signer setup originally defined in (Forster et al., 2012) allows for the evaluation of the impact of features and other components of a sign language recognition system independent of the algorithms employed to automatically track the movement of the hands. Additionally, the frame level annotation of the hand orientation and shape allows to evaluate clustering algorithms in the challenging context of sign language recognition on "real-life" data. Joint modelling of subunits within whole signs can be explored by help of the provided SignWriting annotations.

Extending the single signer setup for sign language recognition, a multi-signer subset has been defined, containing four signers in the training set and two signers in the test set. Table 3 shows the statistics of both setups.

Since the multi-signer setup is a true superset of the sin-

Table 3: Multi and Single Signer Recognition Corpora

	Single Signer		Multi Signer	
	Training	Test	Training	Test
# signers	1	1	4	2
signer overlap	-	all	-	2
# editions	41	31	92	47
duration [m]	31.09	4.5	67.6	8.2
# frames	46,638	6,751	101,496	12,296
# sentences	304	47	725	89
# running glosses	3,309	487	7,776	946
vocabulary size	266	118	349	156
# singletons	90	47	100	53
OOV [%]	-	1.6	-	0.6

Table 4: RWTH-PHOENIX-Weather Translation Corpus

		Glosses	German
Train:	sentences	8 495	
	running words	99 207	134 927
	vocabulary	1580	3 047
	singletons/voc	35.8%	37.9%
Dev:	sentences	250	
	running words	2 573	3 293
	OOVs (running)	1.3%	1.8%
Test : (single signer)	sentences	2× 73	
	running words	487	921
	OOVs (running)	1.6%	(5.4%)
Test : (multi signer)	sentences	2× 135	
	running words	946	1 753
	OOVs (running)	0.7%	(2.4%)

gle signer subset, it allows to evaluate how methods developed in the context of single signer recognition generalize to multiple signers. Additionally, the multi-signer setup allows to investigate inter-signer variability. As with the single signer setup, the multi-signer setup itself is a challenging task with a vocabulary of 349 glosses, out of which 100 occur only once in the training set, and an out-of-vocabulary (OOV) rate of 0.6%. OOV signs cannot be recognized by a recognition system with closed vocabulary, and the high fraction of singleton signs makes it difficult to train robust models. Furthermore, both setups are a challenge with regard to computer vision techniques because of motion blur. In building both setups from single-view video data, approaches to sign language recognition can be evaluated in real-life scenarios where additional information due to RGB-D cameras or stereo-depth information is not available.

For the use of machine translation, two setups have been defined, corresponding to the single signer and multi signer setup of the sign language recognition corpus. In this way, the output of the sign language recognition system can be passed to the translation system, leading to a video-to-text translation pipeline. The corpus statistics of the translation setup can be found in Table 4. As the identity of the signer hardly affects the translation system which is only trained on the glosses and the spoken text, the data for all signers has been included in its training data. Only the test data is different for both setups, as it is identical to the data from the sign language recognition system, featuring one signer in the single signer setup and several signers in the multi signer setup, respectively.

5. Preliminary Results

In this section, we present preliminary results on the RWTH-PHOENIX-Weather 2014 using the sign language recognition system described by (Forster et al., 2013) and our open-source translation system JANE, which was officially released in (Vilar et al., 2010).

Because ground-truth spatial annotation is not available for the majority of frames in the multi-signer setup, Histogram of Oriented 3D Gradients and movement trajectory features have been extracted using tracked spatial positions. Applying the sign language recognition system on the multi-signer setup using a trigram language model results in a Word Error Rate (WER) of 49.2%. The WER measures the minimum number of insertion, deletion and substitution operations needed to transform a hypothesized string into the ground truth string. Using the same feature extraction setup, 45.0% WER are achieved on the single-signer setup in contrast to 39.8% WER using ground-truth spatial annotations (Forster et al., 2013) underlining the need for robust and accurate upper-body tracking.

Table 5: Translation Results DGS to German

	Reference	BLEU	TER
old corpus	announcer	31.8	61.8
	multiple references	38.8	53.9
new corpus	announcer	33.4	60.1
	multiple references	40.9	50.5

Preliminary phrase-based translation results using the original test set from (Forster et al., 2012) are shown in Table 5. A detailed description of how to adapt a statistical machine translation system to the task of sign language translation can be found in (Stein et al., 2012). An advanced method to enhance gloss-based corpora such as the RWTH-PHOENIX-Weather corpus using a viseme recognition system was presented in (Schmidt et al., 2013).

The increase in data leads to an improvement in the translation of 2.1 BLEU (the higher the better) and 3.4 TER (the lower the better). Due to the time constraints of a real-time interpretation of the announcer's words which are spoken at a fast pace, the sign language interpreter sometimes leaves out some minor information. To compensate for this mismatch, we additionally provide another spoken reference which is more closer to the information contained in the glosses. The translation quality is then calculated with regard to the original spoken words and this additional reference. The results on multiple references show that the translation quality is better than the single reference indicates and that there is indeed some mismatch in information.

6. Conclusion

In this work, we introduced the extensions made to the video-based sign language corpus RWTH-PHOENIX-Weather 2012 over the last two years. The resulting RWTH-PHOENIX-Weather 2014 corpus triples the size of the gloss and translation annotation of the original corpus, forming one of the largest available corpora for video-based automatic sign language recognition and translation. The new frame level hand-shape and orientation annotations as well as hand shape and movement annotations in SignWriting enable research regarding automatic clustering of subunits in the context of sign language recognition. Further, the spatial annotation of the hand and nose tip in every frame in the single signer setup provides a standardized setup for the evaluation of tracking algorithms and allows for the evaluation of the impact of tracking errors on sign language recognition performance. Moreover, recognition and translation baselines are provided for further research.

7. Acknowledgments

Oscar Koller would like to thank the Janggen-Pöhn-Stiftung, Switzerland for partial funding of this work. Further, the authors would like to thank Zoltán Tüske from RWTH Aachen University for support in obtaining the transcriptions of the spoken German weather forecast using the RASR engine.

8. References

A. Almohimeed, M. Wald, and R. Damper. 2010. An arabic sign language corpus for instructional language in school. In *4th Workshop on the Representation and Processing of Sign Languages at LREC 2010*, pages 81 – 82, Malta, May.

A Braffort, L Bolot, E Chtelat-Pel, A Choisier, M Delorme, M Filhol, J Segouat, C Verrecchia, F Badin, and N Devos. 2010. Sign language corpora for analysis, processing and evaluation. In *LREC*, pages 453 – 456, Valetta, Malta, May.

O Crasborn and I Zwitserlood. 2008. The Corpus NGT: An Online Corpus for Professionals and Laymen. In *3rd Workshop on the Representation and Processing of Sign Languages at LREC 2008*, pages 44–49, Marrakech, Morocco.

P Dreuw, J Forster, and H Ney. 2010. Tracking benchmark databases for video-based sign language recognition. In *ECCV International Workshop on Sign, Gesture, and Activity*, Crete, Greece, September.

R. Duda, P. Hart, and D. Stork. 2001. *Pattern Classification*. New York: John Wiley & Sons, 2nd edition.

E. Efthimiou, S. Fotinea, T. Hanke, J. Glauert, R. Bowden, A. Braffort, C. Collet, P. Maragos, and F. Lefebvre-Albaret. 2012. Sign language technologies and resources of the dicta-sign project. In *5th Workshop on the Representation and Processing of Sign Languages at LREC 2012*, Istanbul, Turkey, May.

J. Forster, C. Schmidt, T. Hoyoux, O. Koller, U. Zelle, J. Piater, and H. Ney. 2012. RWTH-PHOENIX-Weather: A large vocabulary sign language recognition and translation corpus. In *LREC*, Istanbul, Turkey, May.

J. Forster, O. Koller, C. Oberdörfer, Y. Gweth, and H. Ney. 2013. Improving continuous sign language recognition: Speech recognition techniques and system design. In *Workshop on Speech and Language Processing for Assistive Technologies*, Grenoble, France, August.

O Koller, H. Ney, and R. Bowden. 2013. May the force be with you: Force-aligned SignWriting for automatic subunit annotation of corpora. In *IEEE International Conference on Automatic Face and Gesture Recognition*, Shanghai, PRC, April.

S. Morrissey, H. Somers, R. Smith, S. Gilchrist, and S. Dandapat. 2010. Building a sign language corpus for use in machine translation. In *4th Workshop on the Representation and Processing of Sign Languages at LREC 2010*, pages 172 – 177, Malta, May.

C. Neidle and C. Vogler. 2012. A new web interface to facilitate access to corpora: Development of the asllrp data access interface. In *5th Workshop on the Representation and Processing of Sign Languages at LREC 2012*, Istanbul, Turkey, May.

P. Rutkowski, J. Lacheta, P. Mostowski, S. Lozinska, and J. Filipczak. 2013. The design and compilation of the corpus of polish sign language (PJM). In *5th International Conference on Corpus Linguistics, CILC*, Alicante, Spain, March.

R San-Segundo, J M Pardo, J Ferreiros, V Sama, R Barra-Chicote, J M Lucas, D Snchez, and A Garca. 2010. Spoken spanish generation from sign language. *Interacting with Computers*, 22(2):123–139.

C. Schmidt, O. Koller, H. Ney, T. Hoyoux, and J. Piater. 2013. Using viseme recognition to improve a sign language translation system. In *International Workshop on Spoken Language Translation*, pages 197–203, Heidelberg, Germany, December.

D. Stein, C. Schmidt, and H. Ney. 2012. Analysis, preparation, and optimization of statistical sign language machine translation. *Machine Translation*, 26(4):325–357, December.

M. Sundermeyer, M. Nußbaum-Thom, S. Wiesler, C. Plahl, A. El-Desoky Mousa, S. Hahn, D. Nolden, R. Schlüter, and H. Ney. 2011. The RWTH 2010 quaero asr evaluation system for english, french, and german. In *ICASSP*, pages 2212–2215, Prague, Czech Republic, May.

V. Sutton and Deaf Action Committee for Sign Writing. 2000. *Sign writing*. Deaf Action Committee (DAC).

Z. Tüske, R. Schlüter, and H. Ney. 2013. Multilingual hierarchical mrasta features for asr. In *Interspeech*, pages 2222–2226, Lyon, France, August.

D. Vilar, D. Stein, M. Huck, and H. Ney. 2010. Jane: Open Source Hierarchical Translation, Extended with Reordering and Lexicon Models. In *ACL 2010 Joint Fifth Workshop on Statistical Machine Translation and Metrics MATR*, pages 262–270, Uppsala, Sweden, July.

U. von Agris, M. Knorr, and K.-F. Kraiss. 2008. The significance of facial features for automatic sign language recognition. In *FG*, pages 1–6, Amsterdam, September.

Z. Zafrulla, H. Brashear, H. Hamilton, and T. Starner. 2010. A novel approach to american sign language (asl) phrase verification using reversed signing. In *CVPR workshops*, pages 48–55, San Francisco, CA, USA, August.