Statistical Sign Language Translation

Jan Bungeroth, Hermann Ney

Lehrstuhl für Informatik VI, Computer Science Department
RWTH Aachen University, D-52056 Aachen, Germany
{bungeroth, ney}@informatik.rwth-aachen.de

Abstract

In the field of machine translation, significant progress has been made by using statistical methods. In this paper we suggest a statistical machine translation system for Sign Language and written language, especially for the language pair German Sign Language (DGS) and German. After introducing the system's architecture, statistical machine translation in general and notation systems for Sign Language, the corpus processing is scetched. Finally, preliminary translation results are presented.

1. Introduction

The current progress in statistical machine translation suggests the usage of these methods on automatic Sign Language translation. This paper presents a first approach to such an application and discusses the advantages and disadvantages.

Deaf people, while fluent in their local Sign Language, often experience comprehension problems when they read written text or even lip-read spoken language. Thus for assisting the Deaf to communicate in a world of spoken languages, translation is needed. Currently human interpreters fill this gap, but their service is expensive and not always available. While a machine translation system can not fully replace an interpreter, it offers instant help in the everyday communication.

We therefore propose a system for translating a Sign Language into a spoken language and vice versa. Such a complete system translating from Sign Language to spoken language needs a gesture recognizer as input, the translation system and a speech synthesizer as output. The complete system translating from spoken language to Sign Language needs a speech recognizer as input, the translation system and a graphical avatar as output. In this paper the focus is held on the translation part. Figure 1 presents a schematic overview of such a system.

2. Related Work

In the recent years several groups showed interest in machine translation for Sign Languages.

- In our group, Bauer et al. (1999) proposed a framework for statistical-based Sign Language translation. The authors suggested to translate recognized video-based continuous Sign Language to spoken language.

- Other recent work was done by Sáfár and Marshall (2002) for translating English into British Sign Language using a rule-based approach. Here the grammar was modeled utilizing the HPSG formalism. The system is able to translate simple sentences.

- Huenerfauth (2004) introduces a rule-based concept for translating English text to American Sign Language (ASL).

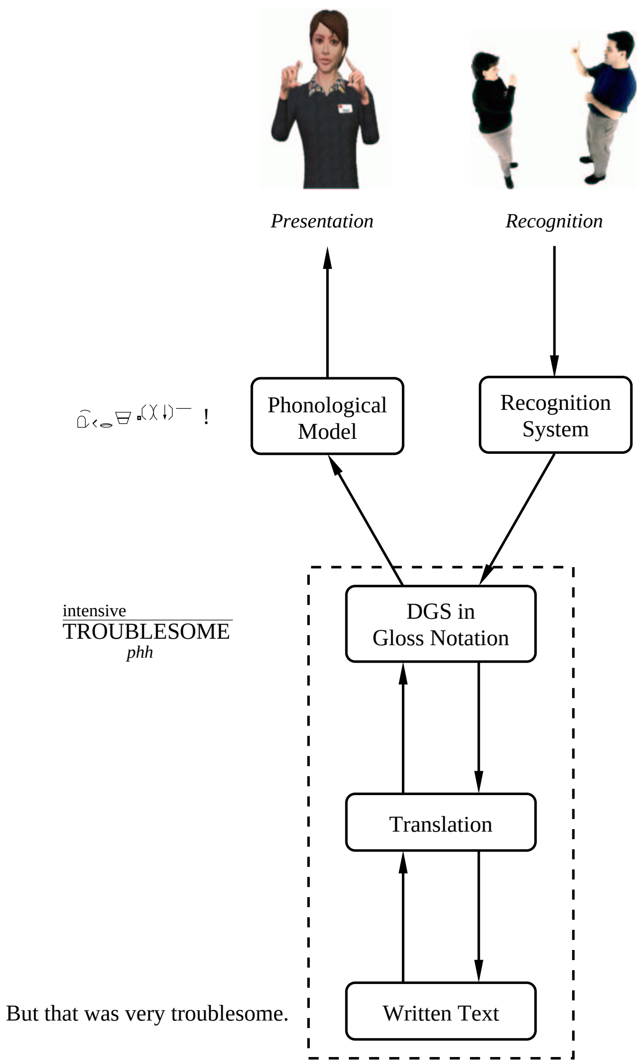

Figure 1: Automatic Sign Language translation system

- Also van Zijl and Barker (2003) propose another rule-based concept for translating English text to South African Sign Language (SASL).

Huenerfauth argues that a rule-based approach is better suited for Sign Language translation than statistical models because large corpora are difficult to obtain. He concludes that the use of a rule-based approach is more appropriate than the statistical. For our work, we do not think of this as an alternate option. Distinct corpora for Sign Languages are planned and already worked on. Additionally the optimization of the statistical translation process for scarce resources as suggested e.g. by Nießen and Ney (2000) allows for further improvement.

3. Statistical Machine Translation

Until recently, only rule-based systems were used for natural language translation. Such systems typically require hand written rules and dictionaries. However, over the last ten years a new approach has evolved, namely the statistical approach. This approach makes use of statistical decision theory and statistical learning. Such a system is trained using a set of sentence pairs. In recent evaluations like Chinese to English[1] and Arabian to English translations, it was found that these statistical approaches were comparable or superior to conventional systems.

In statistical machine translation a source sentence $f_1^J = f_1 \ldots f_J$ is transformed into a target sentence $e_1^I = e_1 \ldots e_I$ by choosing the sentence with the highest probability from all possible target sentences. This is given by Bayes' decision rule

$$\hat{e}_1^I = \underset{e_1^I}{\operatorname{argmax}} \{Pr(e_1^I) \cdot Pr(f_1^J | e_1^I)\}.$$

Several statistical models are used to estimate the free parameters with large training data (e.g. see Brown et al. (1993), Och and Ney (2000)). One target source word position is assigned to each source word position by alignments.

Figure 2 shows the general architecture of the statistical translation approach.

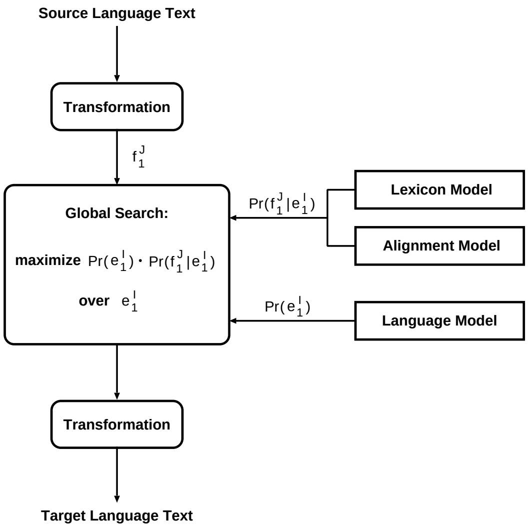

Figure 2: Architecture of the translation approach based on Bayes' decision rule

4. Notation Systems

Several different notations and phonological systems are common in Sign Language research. When dealing with Sign Language translation, an appropriate Sign Language representation is necessary to transfer data from and to the sign recognizer and the presentation avatar. Furthermore a word or phoneme based notation is needed for the internal alignment with the written words of the spoken language. A corpus based on such a notation system should qualify for learning and testing a statistical machine translation, but it might need pre- or postprocessing.

The following notation systems are introduced:

- Glosses are written words, where one gloss represents one sign. Additional markings provide further information, e.g. non-manual signs. Unfortunately no gloss standard exists, which results in inconsistent annotated corpora.

- The notation system introduced by Stokoe (1960) was the very first phonological symbol system of ASL. It divides signs into movement (sig), hand shape (dez) and location (tab) which occur simultaneously. As it focuses on ASL the application on other Sign Languages is not always possible. An ASCII encoding of the Stokoe system is available[2].

- The Hamburg Notation System HamNoSys (Prillwitz, 1989) is a more general form of the Stokoe system. Figure 3 shows an English sentence in gloss notation with markings and the corresponding HamNoSys glyphs.

- Liddell and Johnson (Liddell, 1984) suggest a sequential division of the sign stream into movement and hold segments. This avoids the simultaneous occurrence of phonemes.

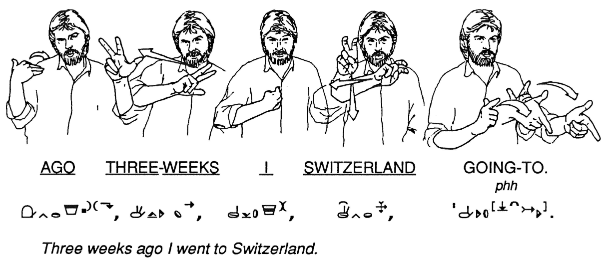

Figure 3: Example for HamNoSys and gloss notation taken from Prillwitz (1989)

5. Corpus Preparation

Statistical machine translation systems are trained using bilingual corpora containing full sentences. But two major problems arise when dealing with Sign Language. The first problem is the lack of large corpora. For example in written language, the Hansards corpus with French and English sentences from debates of the Canadian Parliament contains about 1,470,000 sentences. For Sign Language we have not found a corpus with more than 2000 sentences. The second problem is the lack of a notation standard. The

[1] http://nist.gov/speech/tests/mt/

[2] http://world.std.com/˜mam/ASCII-Stokoe.html

existent corpora use gloss notations which are too difficult to learn with limited corpora. Furthermore inconsistent use of the notation system complicates the problem.

For a starting basis the corpus collected by the DESIRE Team Aachen[3] consisting of 1399 sentences in DGS and German was investigated as it was one of the biggest available for us. Table 1 shows the details of this corpus, where singletons are words occurring only once. Note the very high number of singletons. This comes from the high diversity of the sentences. In addition, every word with a non-manual sign, e.g. (1), is counted as an extra word.

(1) $\overline{\text{HABEN}}^{\text{neg}}$
"not have"

	DGS	German
no. of sentence pairs	1399	
no. of running words	5480	8888
no. of distinct words	2531	2081
no. of singleton words	1887	1379

Table 1: DESIRE corpus statistics

This is not usable for statistical machine translation. Thus for first experiments a small corpus was built from the DESIRE corpus. Several considerations were made:

Brackets indicating a non-manual sign on a whole phrase or sentence are expanded. Consider the sentence (2).

(2) WAHL+ERGEBNIS $\overline{\text{WISSEN DU}}^{\text{qu}}$
"Do you know the election results?"

Table 2 shows the ASCII representation of this sentence before and after expanding the brackets.

```
WAHL+ERGEBNIS qu-{WISSEN DU}
WAHL+ERGEBNIS qu-WISSEN qu-DU
```

Table 2: Expanding brackets in the corpus file

Additional information to locus agreement was deleted as it can not be learned. E.g. in the phrase (3) the 'arbeit' refers to a place in signing space. This information is deleted. After the translation to DGS it can be partially reconstructed by rules.

(3) ARBEITEN X'arbeit'
"at work"

When suitable, the non-manual signs were treated as single words. As an example (4) is processed as seen in table 3, so it can be mapped to the German translation "nicht mögen". But (5) is kept so it can be mapped to the German "unmöglich".

(4) $\overline{\text{MÖGEN}}^{\text{neg}}$
"to like not"

(5) $\overline{\text{MÖGLICH}}^{\text{neg}}$
"impossible"

[3]http://www.germanistik.rwth-aachen.de/desire

```
neg-MÖGEN
neg MÖGEN
```

Table 3: Separating non-manual signs in the corpus file

These methods were used to form the new corpus of 200 sentences. In this corpus the number of singletons is kept low for better training. In addition most words or word forms have an entry in a bilingual manual lexicon. Table 4 gives an overview of the corpus. While this is not enough training data for a fully-fledged translation system, it allows the first experiments, we will discuss in section 6.

		DGS	German
Training:	no. of sentence pairs	167	
	no. of running words	845	828
	no. of distinct words	73	142
	no. of singleton words	15	48
Testing:	no. of sentence pairs	33	
	no. of running words	157	161
	no. of distinct words	43	74
	no. of singleton words	18	40

Table 4: The small DGS/German corpus statistics

6. Results

For translation experiments, training and testing data is needed, as-well as an objective error measurement. The corpus shown in table 4 is divided into training samples (83% of the sentences) and testing samples (17% of the sentences). The training is performed by using various statistical models like IBM Model 1-4 (Brown et al., 1993) and others like Hidden Markov Models HMM (Och and Ney, 2000). Figure 4 shows the alignment of a sentence pair which is obtained in training. For testing, the test sentences

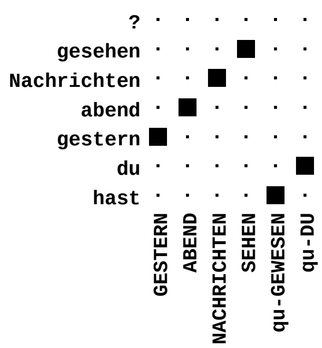

Figure 4: Trained alignment of a sentence pair

in the source language are translated and compared with the the known target sentences. These translation results are evaluated.

We use the following objective evaluation criteria for error measurement:

German	automatic DGS translation	manual DGS translation
du wartest darauf daß der Tee kommt	DU WARTEN BIS TEE KOMMEN	DU WARTEN BIS TEE KOMMEN
frische Bananen und Äpfel schmecken gut	FRISCH ÄPFEL UND BANANEN SCHMECKEN GUT	BANANEN FRISCH UND ÄPFEL SCHMECKEN GUT
ich mag nicht fliegen	ICH NICHT UNKNOWN_fliegen	FLIEGEN ICH neg MÖGEN

Table 5: Translated sentence pairs for German and DGS

- mWER:
  The word error rate (WER) is computed as the minimum number of substitution, insertion and deletion operations that have to be performed to convert the generated sentence into the target sentence. This performance criterion is widely used in speech recognition. This minimum is computed using a dynamic programming algorithm and is typically referred to as edit or Levenshtein distance. In addition for the multi-reference WER (mWER) not only one but a set of reference translation sentences is used. (Nießen et al., 2000)

- mPER:
  The position-independent word error rate (PER) compares the words of the two sentences without considering the word order. The PER is less than or equal to the WER. The multi-reference PER (mPER) again considers a set of reference translation sentences.

We performed the translation from German to DGS on the small corpus. Table 6 shows the mWER and mPER error rates for our experiments. As a reference the baseline is a single word-to-word translation. We then applied our models for the training of alignment models to improve the results.

	mWER [%]	mPER [%]
single word	85.4	43.9
alignment templates	59.9	23.6

Table 6: Testing results for German to DGS

The examples in table 5 show translations from our test corpus. The first sentence is a correct translation, while the second sentence is in partial disorder. The last sentence shows a wrong word order and missing words.

7. Summary

For the translation of spoken language into Sign Language, we propose statistical machine translation. Such a system is trained with bilingual corpora. While Sign Language corpora are still rare, we demonstrated how such a corpus can be prepared for the translation system. Furthermore we performed first experiments on a small German-DGS corpus and presented results. While this is meant only as a small-scale example and a proof-of-concept, we are confident of applying our methods to real-world conditions and corpora.

Future work includes the construction of a more suitable corpus and further improvement of the translation performance. Especially we expect performance gain from the use of better dictionaries and linguistic knowledge like morpho-syntactic information.

8. References

B. Bauer, S. Nießen, and H. Hienz. 1999. Towards an automatic Sign Language translation system. In *Proc. of the Int. Workshop on Physicality and Tangibility in Interaction*, Siena, Italy.

P. F. Brown, S. A. Della Pietra, M. J. Della Pietra, and R. L. Mercer. 1993. Mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 19(2):263–311.

M. Huenerfauth. 2004. A multi-path architecture for machine translation of English text into American Sign language animation. In *Proc. Student Workshop at Human Language Technolgies conference HLT-NAACL*, Boston, MA, USA.

S. Liddell. 1984. Think and believe: Sequentiality in American Sign Language. *Language*, 60(2):372–399.

S. Nießen and H. Ney. 2000. Improving SMT quality with morpho-syntactic analysis. In *Proc. on the 18th Int. Conf. on Computational Linguistics*, Saarbrücken, Germany.

S. Nießen, F.J. Och, G. Leusch, and H. Ney. 2000. An evaluation tool for machine translation: Fast evaluation for machine translation research. In *Proc. of the Second Int. Conf. on Language Resources and Evaluation (LREC)*, pages 39–45, Athens, Greece.

F. J. Och and H. Ney. 2000. A comparison of alignment models for statistical machine translation. In *Proc. on the 18th Int. Conf. on Computational Linguistics*, pages 1086–1090, Saarbrücken, Germany.

S. Prillwitz. 1989. *HamNoSys. Version 2.0; Hamburg Notation System for Sign Language. An Introductionary Guide*. Signum Verlag.

E. Sáfár and I. Marshall. 2002. Sign Language generation using HPSG. In *Proc. Int. Conf. on Theoretical and Methodological Issues in Machine Translation*, pages 105–114, TMI Japan.

W. Stokoe. 1960. Sign Language structure: An outline of the visual communication systems of the American deaf. *Studies in Linguistics, Occasional Papers*, 8.

L. van Zijl and D. Barker. 2003. South African Sign Language machine translation system. In *Proc. 2nd Int. Conf. on Computer graphics, virtual Reality, visualisation and interaction in Africa*, pages 49–52, Cape Town, South Africa.