Combining Data-Driven MT Systems for Improved Sign Language Translation

Sara Morrissey, Andy Way
School of Computing
Dublin City University, Ireland
{smorri,away}
@ computing.dcu.ie

Daniel Stein, Jan Bungeroth, Hermann Ney
Computer Science Department,
RWTH Aachen University, Germany
{bungeroth, stein, ney}
@i6.informatik.rwth-aachen.de

Abstract

In this paper, we investigate the feasibility of combining two data-driven machine translation (MT) systems for the translation of sign languages (SLs). We take the MT systems of two prominent data-driven research groups, the MaTrEx system developed at DCU and the Statistical Machine Translation (SMT) system developed at RWTH Aachen University, and apply their respective approaches to the task of translating Irish Sign Language and German Sign Language into English and German. In a set of experiments supported by automatic evaluation results, we show that there is a definite value to the prospective merging of MaTrEx's Example-Based MT chunks and distortion limit increase with RWTH's constraint reordering.

1 Introduction

Sign languages (SLs) worldwide are poorly resourced and lack political and social recognition. As a result, members of the Deaf community are often forced to communicate in a language that is neither their L1 nor natural to them. Given that on average the literacy competencies of a Deaf adult are poor to moderate (Traxler, 2000) and in most cases resemble those of a 10-year old (Holt, 1991), it is clear that using the *lingua franca* can cause some hindrance to them. This, coupled with the encouraging advancements in the field of machine translation (MT), data-driven MT in particular, leads to the intuition that MT could be employed to help alleviate the communication problems of the Deaf.

In this paper, we look at the data-driven approaches of two well-established MT research groups who have also broached the area of SL MT, namely the MaTrEx system developed at Dublin City University (DCU) and the Statistical Machine Translation (SMT) system developed at RWTH Aachen University. We investigate the methodologies of each for the translation of Irish Sign Language (ISL) and German Sign Language (DGS, Deutsche Gebärdensprache) into English and German.

The remainder of the paper is constructed as follows. In section 2 we introduce SLs and discuss their current status in Ireland and Germany. In section 3 we overview previous and current research in the area of SL MT. We present the data-driven systems of DCU and RWTH Aachen University in Section 4. The dataset we are using is outlined in section five. In section 6 we describe the experiments we carried out and include evaluation results. These results are discussed in section 7. Finally we conclude the paper in section 8 and outline the future direction of our work.

2 Sign Language

SLs are independent and self-contained means of communication used by the Deaf and many hard-of-hearing people. Since the languages have evolved naturally, it is no surprise that most countries have their own particular SL as well as local dialects. SLs are grammatically distinct from spoken languages and the grammar makes extensive use of the possibilities of a visual/gestural modality: locations, verb inflections, pronouns and many other linguistic devices are conveyed by spatial information in front of the signer. Apart from the obvious employment of the hands as information carriers, SLs also use affected facial expressions, tilts of the head and shoulder as well as the velocity of the sign to incorporate information such as comparative degree or subclauses.

In Ireland, ISL is the dominant and preferred language of the Deaf community consisting of approximately 5000 people. Despite being used in Ireland since the 1800s, ISL remains a poorly resourced minority language that lacks social and political recognition. A standardised form of the language is not taught to children in Deaf schools in the same way that English is in spoken language schools. Language development is slow as a result of "its users' lack of access to technical, scientific and political information" (Ó'Baoill & Matthews, 2000).

DGS is spoken and understood by approximately 200,000-300,000 people (including hearing). The language evolved naturally and was accepted as an official language in 2002. Since then, the Deaf are legally entitled to hire interpreters free of charge when dealing with federal authorities, and schools for Deaf children are beginning to adopt the language into their educational system. Although DGS has a completely different vocabulary, certain similar grammatical structures seem to be common in most Western European sign languages.

3 Sign Language MT

The lack of political recognition for SLs worldwide (Gordon, 2005) means that they are less resourced than spoken languages. This may be seen in the areas of SL linguistics and machine translation of SLs. Compared to their spoken language counterparts, both areas are relatively new with significant SL linguistic research beginning only 47 years ago with the work of (Stokoe, 1960). The earliest papers on research into sign language machine translation (SLMT) date back only 18 years.

Apart from the data sparseness, an additional issue for MT is the lack of a formally adopted, or even recognised, writing system. Depending on what the transcription is to be used for, existing systems also differ in accuracy and depth of detail. The earliest linguistic research on SL only dates back to the 1960s to the work of (Stokoe, 1960), which focuses mostly on the syntactic structure of the signs, namely the aspects of manual sign articulation: hand configuration, place of articulation and movement.

For our work, we use so-called 'glosses', a semantic representation of the sign language. As a convention, the meaning of the sign is written as the upper case stem form of the corresponding word in a spoken language, with additional spatial information and facial expression added. For our translation, it annotates all important sign language grammar features.

The example in (1) can be translated into English with 'The high pressure areas over the Atlantic Ocean are growing larger'.

(1) `ATLANTIC_a IX_a HIGH++ GROWING-(more)-hn`

The three signs are transcribed with the glosses 'HIGH', 'ATLANTIC' and 'GROWING' representing their meaning in English. The sign 'IX' is a pointing gesture to reference the same space '_*a*' used by the discourse entity 'ATLANTIC'. Signs repeated (for example to indicate plural forms) are annotated with a double-plus, mouth pictures are written in brackets, e.g. '(more)', '-hn' means that the signer is nodding during signing.

Since the inflection of verbs and nouns is taking place through spatial information, notations like '_*a*' in the above example pose quite a problem for the translation system. Including external morpho-syntactic parsing information usually greatly reduces errors, especially on sparse data sources, but no parsing algorithm exists for the morphologically rich sign languages. Therefore, these issues have to be addressed with proper pre- and post-processing steps.

To collect the data, one has to manually annotate SL video data in a highly time-consuming process (the ECHO project[1], an EU-funded scheme based in the Netherlands that has made fully annotated digitised corpora available on the Internet, reports one minute of video data takes approximately 100 minutes to annotate). This approach involves transcribing information taken from a video of signed data. The transcriber may decide the level of detail at which the SL video will be described. These categories include a gloss term of the sign being articulated by the hands and may also contain information about non-manual features or any other relevant linguistic information.

3.1 Previous Approaches

Given the relatively recent research into SLMT, most systems to date have used 'second generation' approaches. Transfer-based approaches have included the work of (Grieve-Smith, 1999) who translated English weather reports into American Sign Language (ASL) by mapping syntactic structures. (Van Zijl & Barker, 2003) also used a syntactic approach in their work on South African Sign Language with most of their focus on avatar production. (Marshall & Sáfár, 2002; Sáfár & Marshall, 2002) employ discourse representation structures and use HPSG semantic feature structures for the generation of ASL.

There have also been interlingual approaches adopted by (Veale et al., 1998) and (Zhao et al., 2000), the latter employing synchronised tree–adjoining grammars.

A second generation hybrid approach has been developed by (Huenerfauth, 2005) where interlingual, transfer and direct approaches are integrated.

[1]http://www.let.kun.nl/sign-lang/echo/data.html

3.2 Current Developments

More recently, SLMT has followed the more mainstream MT trend away from rule-based approaches toward data-driven methods. The following groups are active in their 'third generation' approaches:

- (Morrissey & Way, 2005; Morrissey & Way, 2006) investigate corpus-based methods for example-based sign language translation from English to the sign language of the Netherlands.
- (Chiu et al., 2007) present a system for the language pair Chinese and Taiwanese sign language. They show that their optimizing method surpasses IBM model 2.
- Basic work on Spanish and Spanish sign language was done by (San-Segundo et al., 2006). Here, a speech to gesture architecture is proposed.
- A complete system setup was discussed by Stein (Stein et al., 2006) for German and German sign language on the domain weather reports. Further, they describe how to improve the results with sign language specific pre- and post-processing methods.

4 Data-Driven MT in DCU and RWTH University

Over the last 10 years, the National Centre for Language Technology at DCU has developed a successful track record in research on Data-Driven MT. This is evident from the work of (Veale & Way, 1997) involving a template-driven approach to EBMT, to the Marker-Based segmentation research of (Gough & Way, 2004b) and more recently the work of (Stroppa & Way, 2006) on the development of the MaTrEx MT system (cf. section 4.1) which has performed well in international evaluations such as `IWSLT`.[2]

For over a decade, the RWTH University has been focussing research on SMT. The system has achieved very competitive results in all international evaluations in which it has participated (`TC-STAR`[3],`IWSLT`, `NIST`[4]).

In light of these developments, we have chosen to combine the approaches of these two prominent data-driven MT research centres and apply their approaches to the area of SL translation. These systems are described in more detail in sections 4.1 and 4.2.

[2]http://www.slt.atr.jp/IWSLT2006/

[3]http://www.tc-star.org/

[4]http://www.nist.gov/speech/tests/mt/

4.1 MaTrEx: The DCU MT System

MaTrEx (Machine Translation using Examples) is the Data-Driven MT system developed at DCU (Stroppa & Way, 2006). It is a hybrid system that combines Example-Based MT (EBMT) and SMT approaches. The system is modular in design consisting of a number of extendible and reimplementable modules. This modular design makes it particularly adaptable to new language pairs. An overview of the translation process is in Figure 1.

The decoder is fed by different example databases to translate new sentences. These chunk and lexical example databases are created using the the Word Alignment, Chunking and Chunk Alignment Modules that are themselves fed by aligned source-target sentences.

4.1.1 Word Alignment Module

Word alignment for the system is performed using Giza++ (Och, 2003), a statistical word alignment toolkit. A set of high-quality word alignments are extracted from the original uni-directional alignment sets using the "refined" method of (Koehn et al., 2003).

4.1.2 Chunking Module

The primary chunking strategy employed for our language pairs in this system is based on the Marker Hypothesis (Green, 1979). This method is based on the universal psycholinguistic constraint that languages are marked for syntactic structure at their surface level by closed sets of lexemes or morphemes. Lists of closed-class "marker" words (i.e. prepositions, conjunctions, determiners etc.) are used to segment the sentences and derive a new data source: a set of marker chunks. Each chunk consists of one or more marker words and at least one non-marker word to ensure contextual information is withheld in the chunk. Marker-based chunking has the advantage of being easily adaptable to new languages by simply providing the system with a relevant list of marker words.

4.1.3 Chunk Alignment Module

An 'edit-distance style' dynamic programming alignment algorithm is employed to align the chunks created in the chunking module. Rather than using the Expectation-Maximization algorithm for parameter estimation, instead these are directly computed according to the information within the chunks. This information is obtained from three sources: word-to-word translation probabilities, word-to-word cognates and chunk labels. The resulting aligned chunks are then combined with the SMT phrasal alignments. The two alignment styles

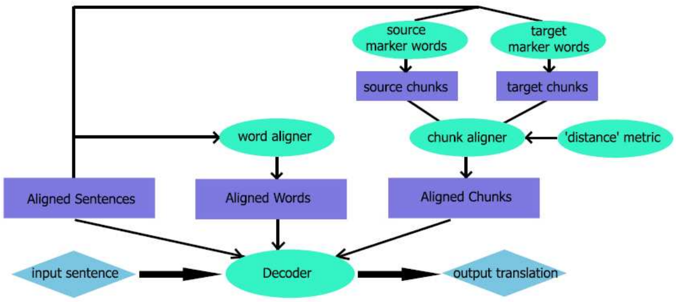

Figure 1: MaTrEx Architecture

are merged to help produce translations of a higher quality following the recent research of (Groves & Way, 2005; Groves & Way, 2006)

4.1.4 Decoder

The MaTrEx decoder is a wrapper around Moses (Koehn et al., 2007), a phrase-based SMT decoder. Minimum-Error-Rate training (Och, 2003) is implemented within a log-linear framework (Och & Ney, 2002) and the BLEU metric (Papineni et al., 2002) is optimized using the development set.

4.2 The RWTH MT System

We use a SMT system to automatically transfer the meaning of a source language sentence into a target language sentence.

Our baseline system maximizes the translation probability directly using a log-linear model (Och & Ney, 2002) shown in below:

$$p(e_1^I|f_1^J) = \frac{\exp\left(\sum_{m=1}^M \lambda_m h_m(e_1^I, f_1^J)\right)}{\sum_{\tilde{e}_1^I} \exp\left(\sum_{m=1}^M \lambda_m h_m(\tilde{e}_1^I, f_1^J)\right)}$$

with a set of different features h_m, scaling factors λ_m and the denominator a normalization factor that can be ignored in the maximization process. We choose the λ_m by optimizing an MT performance measure on a development corpus using the downhill simplex algorithm. For a complete overview of the translation system, see (Matusov et al., 2006).

4.2.1 Reordering constraints

When we look for the best translation we can reduce the search space by assuming monotone word dependency. This works well for closely related language pairs, such as Catalan-Spanish, that have a very similar grammatical structure and phrases containing similar sequences of words over large portions of the text. However, many other language pairs differ significantly in their word order. To keep computational costs at a reasonable level, we allow a larger search space but limit the permutation number by *reordering constraints*.

A reordering constraint is a directed, acyclic graph that allows limited word reordering of the source sentence. The edges of each possible path equal a permutation π of the numbers 1 to J.

In our work, we investigate the influence of three reordering graphs Figure 2 (Kanthak et al., 2005) on our translation results: the local constraint, the IBM constraint and the inverse IBM constraint. Each graph allows characteristic permutation types, constrained by a window size w: the local constraint allows each word in the sentence to be moved up to a maximum of $w-1$ steps toward the front or the end of the sentence. The IBM constraint allows up to $w-1$ words in the sentence to be moved to the end of the sentence, likewise, the inverse IBM constraint allows up to $w-1$ words to be moved to the sentence beginning.

The higher the window size w, the higher the amount of possible permutations has to be considered. A window size which is higher or equal to the sentence length J results in a search space that is equal to the maximum of permutations possible.

5 The Corpus

A prerequisite for data-driven approaches to MT involves a bilingual data set. The broader the domain and/or vocabulary, the higher the need for a sufficient amount of data to properly train the system. For spoken language, there are large amounts of data available for use in MT.

On the contrary, suitable SL corpora are not as easy to find. Most of the corpora available in SLs take the form of prose or conversational communication as they are primarily used for linguistic analysis and are unsuited to MT because of their open-domain content and use of flowery language. This is

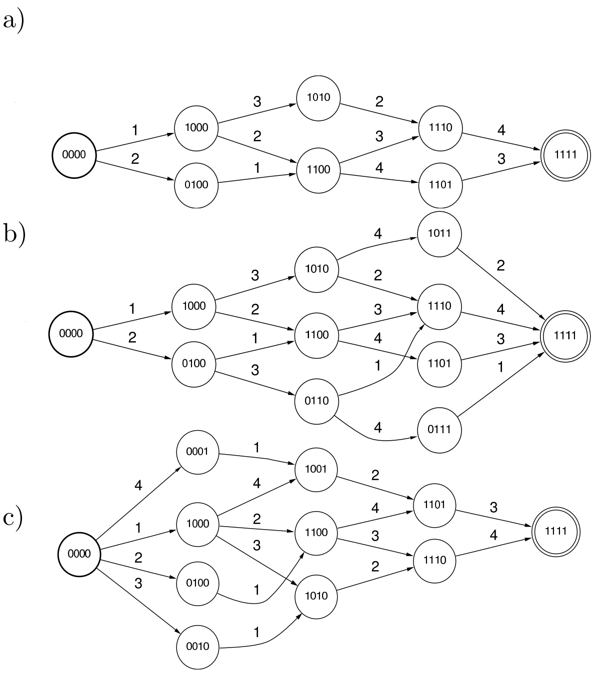

Figure 2: Permutation graph of a source sentence $f_1 f_2 f_3 f_4$ using a window size $w = 2$ for a) local constraints, b) IBM constraints and c) inverse IBM constraints

evident from the work of (Morrissey & Way, 2006), whose SL research made use of data from the ECHO project data. For the most part, that which is available is so small in terms of sentence quantity that it is unusable for data-driven MT. For these reason we chose to create our own corpora.

We found a suitable dataset in the ATIS corpus (Hemphill et al., 1990). The ATIS (Air Travel Information System) corpus is a dataset of transcriptions from speech containing information on flights, aircraft, cities and similar related information. This corpus is particularly suited to our MT needs as it is within a closed domain and has a small vocabulary. The domain itself has a potentially practical use for Deaf people.

The ATIS corpus consists of 595 English sentences. Although this is a significantly smaller dataset than that used in mainstream data-driven MT, it is sufficient to feed our systems, as demonstrated in section 6. We had this dataset translated into ISL by Deaf native ISL signers and recorded for video and then annotated with an English gloss. The English data (EN) was also translated into German (DE) and then DGS gloss annotation. This provided us with four parallel corpora, already sententially aligned, with the potential to work with four translation pair types containing twelve different language pairs, namely:

		EN	DE	ISL	DGS
Train	no. sentences	418			
	no. running words	3008	3544	3028	2980
	vocab. size	292	327	265	244
	no. singletons	97	118	71	84
Dev	no. sentences	59			
	no. running words	429	503	431	434
	vocab. size	134	142	131	119
Train	no. sentences	118			
	no. running words	999	856	874	877
	vocab. size	174	158	148	135
	trigram perplexity	15.7	12.4	28.3	11.39
	out of vocab.	22	22	30	15

Table 1: Corpus Overview

(i) from SL to spoken language (ISL–EN, ISL–DE, DGS–EN, DGS–DE),

(ii) spoken language to SL (EN–ISL, DE–ISL, EN–DGS, DE–DGS),

(iii) spoken language to spoken language (EN–DE, DE–EN)

(iv) and the novel translation pairings of SL to SL (DGS–ISL, ISL–DGS).

For this paper, we have chosen to work with type (i). These language pairings and directions were chosen to facilitate automatic evaluation of the translated output. With respect to type (ii) and the SL–SL pairings of type (iv), currently SL automatic evaluation is not possible. This is because traditional string-based evaluation metrics such as BLEU (Papineni et al., 2002) "are inappropriate for the evaluation of SLMT systems, where the primary goal is translation from an oral to a non-oral language, as there is no 'gold standard' underlying SL annotation available" (Morrissey & Way, 2006).

6 Experiments

The 595 sentences of the ATIS corpus were divided into training, development and testing sets of 418 sentences, 59 sentences and 118 sentences respectively. An overview of the corpus breakdown is given in Table 1.

6.1 MaTrEx Experiments

The baseline system for all MaTrEx experiments employs the modules as described in section 4.1 with the exception of the EBMT chunks.

To try to improve on these translations we introduced EBMT chunks into the system in two sets of experiments. In the first, 'type 1 chunks', we used the Marker Hypothesis described in section 4.1.2 to segment both the source and target sentences. The resulting chunks and corresponding alignments were added to the system.

		BLEU	WER	PER
ISL-EN	baseline	51.63	39.32	29.79
	DL = 10	*52.18*	*38.48*	*29.67*
	T1 chunks	50.69	37.75	30.76
	DL = 10	*51.31*	*37.39*	*30.63*
	T2 chunks	49.76	39.92	32.44
	DL = 10	*50.32*	*39.56*	*32.32*
ISL-DE	baseline	38.18	48.52	38.79
	DL = 10	*39.69*	*47.25*	*38.47*
	T1 chunks	40.67	46.72	38.58
	DL = 10	*42.13*	*45.45*	*38.16*
	T2 chunks	38.54	46.93	38.05
	DL = 10	*40.09*	*45.66*	*37.63*
DGS-EN	baseline	45.25	48.85	32.08
	DL = 10	*48.40*	*41.37*	*30.88*
	T1 chunks	44.74	50.66	31.72
	DL = 10	*47.22*	*44.14*	*31.12*
	T2 chunks	44.34	49.93	33.17
	DL = 10	*47.43*	*42.82*	*32.20*
DGS-DE	baseline	38.66	55.28	39.53
	DL = 10	*42.09*	*50.31*	*39.53*
	T1 chunks	34.86	56.65	39.53
	DL = 10	*39.38*	*51.37*	*38.79*
	T2 chunks	35.63	55.81	39.74
	DL = 10	*40.29*	*50.31*	*38.90*

Table 2: MaTrEx Evaluation Results

Given the natural lack of closed class lexical items in SLs, it was noted that frequently one word taken from the SLs would combine with a whole marker-based chunk from the corresponding target language. With this in mind we ran 'type 2 chunks' experiments where the marker-based chunks for the spoken target languages were aligned with SL chunks where each SL word formed its own chunk.

The default distortion limit for the decoder is set to allow for no jumps or 'block movements' to occur in translation. Given the differences in between SLs and spoken language grammar, particularly the sentence-initial positioning of time references and similar grammatical structures, we experimented using varying distortion limits.

The evaluation results for each set of experiments is given in Table 2.

The 'type 1 chunks' formula worked best for ISL–DE leading to an improvement in the BLEU score of 2.49% and a decrease in WER of 1.8%. This shows a relative increase of 6.5% and 3.7% respectively.

The addition of 'type 2 chunks' for DGS–DE improved the PER scores but the same improvement was not reflected in the BLEU score.

Allowing a distance range of 10 for block movements when decoding improves BLEU scores across all language pairs. This is particularly noted for the DGS–DE pairing where, for example, when combined with 'type 2 chunks', the BLEU score increases by 4.66% on the 'type 2 chunks' score alone, which shows a relative increase of 13%. This experiment also had the effect of lowering error rates for all language pairings.

		BLEU	WER	PER
ISL-EN	baseline	50.72	39.44	30.27
	inv-IBM reord.	52.62	37.63	28.34
ISL-DE	baseline	40.36	47.25	38.90
	inv-IBM reord.	40.40	46.40	38.58
DGS-EN	baseline	40.10	51.62	36.55
	inv-IBM reord.	43.16	46.32	31.36
DGS-DE	baseline	32.92	55.07	40.69
	inv-IBM reord.	35.69	49.15	38.68

Table 3: RWTH Evaluation Results

6.2 RWTH Experiments

The baseline includes minimum error rate training on the weighting parameters with the WER as the optimized evaluation measure. For the reordering constraints, the three permutations graphs, their window size, and the probability of the monotone (original) permutation have also been optimized on the development set. The local optimum was determined for inverse-IBM reordering with a window size of 3 for all four language pairs. The reordering probability was best for values around 0.6. The results can be found in Table 3.

7 Discussion

Based on the experiments we have carried out on the ATIS corpora, it is clear that taking into account the differences in SL and spoken language grammar and allowing more freedom of movement of units when decoding has paid off and produced successful improvements in evaluation scores. The initial experiments of adding various chunking style alignments has shown potential for helping to improve scores and that further chunking styles most suited to SLs are worth investigating.

Reordering constraints seem to work quite well for language pairs that include DGS. During translation, it was stressed by the deaf interpreters that information on time and location should always appear at the beginning of the sentence. Thus, many sentences benefited from the reordering since in the spoken language, time and location might not appear before the middle or the end of the sentence. An example is given in Table 4. Although the constraints also improved for translations starting from ISL, the difference in the error rate is not as dramatic.

Taking into account the improvements to the MaTrEx baseline by adding EBMT-style chunks to the SMT phrasal alignments and increasing the distortion limit for jumps in when decoding, it is clear that these methodologies would add to a general SMT system. Furthermore, the improvements to scores for the RWTH system baseline following the constraint

source sentence	CORK IX_a a_BIS_b SHANNON IX_b a_FLIEGEN_b WAS-qu
target sentence	what flights are there from cork to shannon
baseline translation	cork from shannon to what
+ inverse-ibm reordering constraint	what flights go from cork to shannon
source sentence	FREITAG BELFAST IX_a a_BIS_b DUBLIN IX_b a_FLIEGEN_b NAME AER LINGUS WAS-qu
target sentence	welche flüge von belfast nach dublin am freitag mit aer lingus?
baseline translation	freitag von belfast nach dublin, aer lingus welche
+ inverse-ibm reordering constraint	welche flüge von dublin nach belfast am freitag mit aer lingus?

Table 4: Influence of reordering constraints on the translation result

reordering shows that combining this approach with the previous MaTrEx methodologies could potentially further improve the translations for these data sets. With this in mind, we intend to combine these methodologies. Also, since the evaluation results of both systems are in the same range for all language pairs, we also plan on exploiting different system-specific error types with system combination of the translation outputs in the future.

Comparative work has been presented on Dutch Sign Language data of a similar size in (Morrissey & Way, 2006). Manual annotation was also used in this work but at a much more detailed level. As highlighted in section 5 of this paper, the open domain of fables and poetry was used in their work. They present a WER score of 119% and a PER of 78% for a test set of 55 sentences transating Dutch SL to English. No BLEU score is reported due to the lack of 4-gram matches found during evaluation. Based on these results, it can be assumed that the greatly improved scores attained by both the MaTrEx and RWTH systems described in this paper can partly be attributed to the more closed domain, and simpler annotation as well as the more sophisticated MT systems.

8 Conclusions

In this paper, we have investigated the methodologies of two data-driven MT systems, the MaTrEx system and RWTH's SMT system with a view to combining their methodologies for translating sign languages. Through sets of experiments carried out on ISL and DGS datasets, we have shown promising results for the addition of EBMT-style chunks, increasing distortion limits and reordering constraints. This shows some potential for producing improved translations if incorporated together in a data-driven system.

Our research has also highlighted the need for MT to be applied to SLs to aid communication with the Deaf and hearing communities and have outlined current developments in this area. With this in mind, we have begun to take our work further by adding an SL recognition tool to the front end of our current system to develop a fully automatic SL-to-spoken language MT system (Stein et al., 2007). For the ATIS corpus and its available video recordings in ISL, some preliminary but promising experiments have been carried out to connect the recognition and MT processes.

At a later stage, to facilitate a more practical use for the Deaf we hope to reverse the language direction and produce SL translations of spoken language through the medium of an avatar, thereby allowing Deaf people to translate and access information in their natural language. The development of both these language directions leads naturally to the merging of both systems to allow for translation from SL-to-SL, a novel area of research that could facilitate worldwide communication between Deaf communities.

9 Acknowledgments

We would like to thank the anonymous reviewers whose comments and suggestions helped improve the quality of this paper. We would like to express our gratitude to Ege Karar and Horst Sieprath for their linguistic knowledge and input for German Sign Language as well as Kevin G. Mulqueen and Mary Duggan for their assistance in the creation of the ISL corpus. This research is part-funded by the Irish Research Council for Science, Engineering and Technology (IRCSET)[5] and IBM.[6]

References

Chiu, Y.-H., 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. In *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(1):28–39.

[5] http://www.ircset.ie

[6] http://www-927.ibm.com/ibm/cas/sites/dublin/

Gordon, R. G., Jr. (ed.). 2005. *Ethnologue: Languages of the World, Fifteenth Edition*. Dallas, Texas.: SIL International.

Gough, N. and A. Way. 2004. Robust Large-Scale EBMT with Marker-Based Segmentation. In *Proceedings of the Tenth Conference on Theoretical and Methodological Issues in Machine Translation (TMI-04)*, Baltimore, MD., pp.95–104.

Green, T. 1979. The Necessity of Syntax markers: Two Experiments with Artificial Languages. In *Journal of Verbal Learning Behaviour*, vol. 18, pp.95–104.

Grieve-Smith, A. B. 1999. English to American Sign Language Machine Translation of Weather Reports. In D. Nordquist (ed.) *Proceedings of the Second High Desert Student Conference in Linguistics (HDSL2)*, Albuquerque, NM., pp.23–30.

Groves, D. and A. Way. 2005. Hybrid example-based SMT: the best of both worlds? In *Proceedings of the Workshop on Building and Using Parallel Texts: Data-Driven Machine Translation and Beyond, ACL 2005*, Ann Arbor, Michigan, pp. 183–190.

Groves, D. and A. Way. 2006. Hybrid data-driven models of MT. *Machine Translation, Special Issue on EBMT* (to appear).

Hemphill, C.T., J.J. Godfrey and G.R. Doddington. 1990. The ATIS Spoken Language Systems pilot corpus. In *Proceedings of DARPA Speech and Natural Language Workshop*, Hidden Valley, PA., pp. 96-101.

Holt, J. 1991. Demographic, Stanford Achievement Test – 8th Edition for Deaf and Hard of Hearing Students: Reading Comprehension Subgroup Results. *Journal of Deaf Studies and Deaf Education*

Huenerfauth, M. 2005. American Sign Language Generation: Multimodal NLG with Multiple Linguistic Channels. In *Proceedings of the ACL Student Research Workshop (ACL 2005)*, Ann Arbor, MI., pp.37–42

Kanthak, S., D. Vilar, E. Matusov, R. Zens and Ney, H. 2005. Novel reordering approaches in phrase-based statistical machine translation. In*ACL workshop on building and using parallel texts.* Ann Arbor, MI: Association for Computational Linguistics, pp. 167–174.

Koehn, P., F. Och and D. Marcu. 2003. Statistical Phrase-Based Translation. In *Proceedings of AMTA 2006*, Cambridge, MA., pp.48–54.

Koehn, P. 2005. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of the MT Summit X*, Phuket, Thailand, pp79–86.

Koehn, P., H. Hoang, A. Birch, C. Callison-Burch, M. Federico, N. Bertoldi, B. Cowan, W. Shen, C. Moran, R. Zens, C. Dyer, O. Bojar, A. Constantin, and E. Herbst. 2007. Moses:Open source toolkit for SMT. In *Proceedings of the ACL 2007 Demo and Poster Session*, Prague, Czech Republic, pp.177–180.

Marshall, I. and É. Sáfár. 2002. Sign Language Generation using HPSG. In *Proceedings of the 9th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-02)*, Keihanna, Japan, pp.105–114.

Mauser, A., R. Zens, E. Matusov and H. Ney. 2006. The RWTH Statistical Machine Translation System for the IWSLT 2006 Evaluation, In *Proceedings of the IWSLT*, Kyoto, Japan, pp.103–110.

E. Matusov, R. Zens, D. Vilar, A. Mauser, M. Popovic and H. Ney. 2006. The RWTH Machine Translation System. In *Proceedings of the TC-STAR Workshop on Speech-to-Speech Translation*, Barcelona, Spain, pp. 31–36.

Morrissey, S. and A. Way. 2005. An Example-Based Approach to Translating Sign Language. In *Proceedings Workshop Example-Based Machine Translation (MT X 05)*, Phuket, Thailand, pp.109–116.

Morrissey, S. and A. Way. 2006. Lost in Translation: the Problems of Using Mainstream MT Evaluation Metrics for Sign Language Translation. In *Proceedings of the 5th SALTMIL Workshop on Minority Languages at LREC'06.* Genoa, Italy, pp.91–98.

Ó'Baoill, D. and P. A. Matthews. 2000. *The Irish Deaf Community (Volume 2): The Structure of Irish Sign Language.* The Linguistics Institute of Ireland, Dublin, Ireland.

Och, F.J. 2003. Minimum Error Rate Training in Statistical Machine Translation. In *Proceedings of ACL 2003*, Saporo, Japan, pp.160–167.

Och, F.J., and H. Ney. 2002. Discriminative training and maximum entropy models for statistical machine translation. In *Proceedings of ACL 2002*, Philadelphia, PA, pp. 295–302.

Papineni, K., S. Roukos, T. Ward and W. Zhu. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association of Computational Linguistics (ACL-02)*, Philadelphia, PA., pp.311–318.

Sáfár, É. and I. Marshall. 2002. The Architecture of an English-Text-to-Sign-Languages Translation System. In *Proceedings of Recent Advances in Natural Language Processing (RANLP-01)*, Tzigov Chark, Bulgaria, pp.223–228.

San-Segundo, R., 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 *Proceedings of Interspeech 2006*, Pittsburgh, PA.

Stein, D., J. Bungeroth and H. Ney. 2006. Morpho-Syntax Based Statistical Methods for Sign Language Translation. In *Proceedings of the 11th Annual conference of the European Association for Machine Translation (EAMT'06)*, Oslo, Norway, pp.169–177.

Stein, D., J. Bungeroth, H. Ney, S. Morrissey and A. Way. 2007. Hand in Hand: Automatic Sign Language to Speech Translation. In *Proceedings of the 11th Conference on Theoretical and Methodological Issues in Machine Translation (TMI-07)*, Skövde, Sweden, (forthcoming).

Stokoe, W. C. 1960. An Outline of the Visual Communication Systems of the American Deaf. In *Studies in Linguistics: Occasional papers, No. 8*,Department of Anthropology and Linguistics, University of Buffalo, Buffalo, New York, pp.78.

Stroppa, N. and A. Way. 2006. MaTrEx: DCU Machine Translation System for IWSLT 2006. In *Proceedings of the International Workshop on Spoken Language Translation*, Kyoto, Japan, pp.31-36.

Traxler, C.B. 2000. The Stanford Achievement Test, 9th Edition: National Norming and Performance Standards for Deaf and Hard–of–Hearing Students. *Journal of Deaf Studies and Deaf Education*, **5**, 4, pp.337–348.

Van Zijl, L. and D. Barker. 2003. South African Sign Language Machine Translation System. In *Proceedings of the Second International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa (ACM SIGGRAPH)*, Cape Town, South Africa, pp.49–52.

Veale, T., A. Conway, and B. Collins. 2000. The Challenges of Cross-Modal Translation: English to Sign Language Translation in the Zardoz System. *Machine Translation* **13**(1):81–106.

Veale T. and A. Way. 1997. GAIJIN: A Bootstrapping Approach to Example-Based MachineTranslation. In *Proceedings of 2nd International Conference on Recent Advances in Natural Language Processing*, Tzigov Chark, Bulgaria, pp.239–244.

Zhao, L., K. Kipper, W.Schuler, C. Vogler, N. Badler and M. Palmer. 2000. A Machine Translation System from English to American Sign Language. In *Envisioning Machine Translation in the Information Future: Proceedings of the Fourth Conference of the Association for Machine Translation (AMTA-00)*, Cuernavaca, Mexico, pp.293–300.