An Example-Based Approach to Translating Sign Language

Sara Morrissey
School of Computing
Dublin City University
Dublin 9, Ireland
smorri@computing.dcu.ie

Andy Way
School of Computing
Dublin City University
Dublin 9, Ireland
away@computing.dcu.ie

Abstract

Users of sign languages are often forced to use a language in which they have reduced competence simply because documentation in their preferred format is not available. While some research exists on translating between natural and sign languages, we present here what we believe to be the first attempt to tackle this problem using an example-based (EBMT) approach.

Having obtained a set of English–Dutch Sign Language examples, we employ an approach to EBMT using the 'Marker Hypothesis' (Green, 1979), analogous to the successful system of (Way & Gough, 2003), (Gough & Way, 2004a) and (Gough & Way, 2004b). In a set of experiments, we show that encouragingly good translation quality may be obtained using such an approach.

Key-words: Example-based machine translation, sign languages, Marker Hypothesis, ECHO corpus.

1 Introduction

Just like speakers of a less widely spoken language are often not catered for properly with respect to the provision of documentation in their preferred language, users of sign languages (SLs) observe similar restrictions. Having to read documents in the *lingua franca* often causes them some hindrance. This is because a system of 'oralism' (the practice of teaching Deaf students through spoken language using amplification devices and lip-reading, to the exclusion of all sign language communication) is used in most Deaf schools. As the students lack the ability to hear the language, on average their literacy competencies remain at approximately that of a ten year old (Holt, 1991).

A small body of work has attempted to alleviate the situation for SL users by developing machine translation (MT) systems capable of translating texts written in natural languages into various SLs. This field of SLMT is still in its infancy with research into the area dating back approximately ten years. Many of the systems proposed to date are rule-based systems, based on transfer approaches (Grieve-Smith, 1999), interlingual systems (e.g. the Zardoz system, (Veale et al., 1998)), or hybrid models where these approaches are combined (Huenerfauth, 2004, 2005). On a rather smaller scale, corpus-based approaches have also been proposed (Bauer et al., 1999).

Example-Based MT (EBMT) has been around for over 20 years now, from the seminal paper of (Nagao, 1984) to the more recent collection of (Carl & Way, 2003) and beyond. However, as far as we are aware, no previous approaches to SL translation have used such a method. In the medium to long term, our main goal is to develop an EBMT system for the language pair English–Irish Sign Language (ISL), in both directions. However, at this early stage of the project no ISL corpus is available, though one is in the process of being constructed by the Centre for Deaf Studies[1] in Dublin.

In order to demonstrate proof-of-concept of our approach, therefore, we present a system which instead translates between English and Nederlandse Gebarentaal/Sign Language of the Netherlands (NGT). We obtained a corpus of NGT examples from the ECHO project website.[2] As consultants on the ISL corpus-building process, we are aware that the ISL corpus is being constructed using the same annotation process and toolkit as that of the ECHO corpus, so developing an English–NGT EBMT system is a reasonable approximation of the task with which we will eventually be confronted. In initial experiments, we devised a set of sentences for testing the system and used manual analysis to evaluate the results. At this preliminary stage, the results obtained are encouraging.

The remainder of the paper is constructed as follows. In section 2, we describe previous related research in this area. In section 3, we present some of the issues involved in projects of this type, in particular the ECHO project, by showing the internal representation of an NGT object and describing how an EBMT approach may avail of this data. In section 4, we briefly summarize the main ideas behind typical models of EBMT, as well as the particular system used here. Section 5 presents the results obtained by our prototype EBMT system, and discussion of the major findings. Finally, we conclude and present avenues for further research.

[1] http://www.tcd.ie/Deaf_Studies/
[2] http://www.let.kun.nl/sign-lang/echo/

2 Related Work

It is only in the last ten years or so that an interest has been taken in using MT techniques to automate the translation of sign languages. Most of the research that has been carried out has involved the development of a system for the language pair English–American Sign Language (ASL), although there have been a few other language pair models. The most common approaches to date have been rule-based with more SL corpora being created we can reasonably expect corpus-based approaches to become more prevalent in this field mirroring the situation in 'regular' MT. The majority of systems work at translating spoken languages in text format into sign language that is then reproduced as either an avatar of a signing mannequin or a literal orthography (written annotation of the sign language).

Transfer systems have been developed by:

- (Grieve-Smith, 1999), who modelled a system for English–ASL using the limited domain of Albuquerque weather reports;
- (Marshall & Sáfár, 2002), (Sáfár & Marshall, 2002), whose English–ASL system is semantically driven and uses HPSG semantic feature structures and Discourse Representation Structures to represent the internal structure of linguistic objects;
- (Van Zijl & Barker, 2003), who proposed a system for English–South African Sign Language.

In terms of Interlingual approaches:

- (Zhao et al., 2000) developed an English–ASL system that uses synchronised tree adjoining grammars;
- (Veale et al., 1998) developed the Zardoz system for translating English into ISL, ASL and Japanese Sign Language.

In addition, (Huenerfauth, 2004, 2005) has proposed a hybrid multi-path system where English is translated into ASL using a combination of an interlingua, transfer methods and direct methods. This work focuses in particular on models for classifier predicates.

Systems translating from sign language into written oral-language text have also been developed, one such system being that of (Bauer et al., 1999). This is a statistical MT (SMT) system that uses Hidden Markov Models in the recognition of signs before using a translation model and a language model for translation in the usual SMT manner.

3 Sign Language

Despite common misconceptions, sign languages are indigenous, fully accessible languages for Deaf people, with their own unique syntax and grammar. Each country has its own sign language and these languages can vary slightly from region to region just like the dialects of a spoken language. In recent years, more and more national sign languages have begun to be officially recognised in the countries where they are used, as they are the primary means of communication for Deaf people. Regrettably, in many others "provision is not made for deaf people to access public information, or receive vital services such as education and health in their first language" (Ó'Baoill & Matthews, 2000), namely sign language. This is also true for the accessibility of public or private information in the form of written documents. This is an area in which an automated translation of written text could prove invaluable to members of the Deaf community, particularly in areas of low interpreter availability.

3.1 SLMT Issues

The development of an SLMT system requires a number of issues to be taken into consideration. An SLMT system has to deal with some of the problems that models of translation for non-SLs have to handle, such as varying and free syntax, morphological issues (e.g. repetition and pluralisation), and lexical gaps. In addition, models of SLMT should also have the ability to deal with sign language-specific phenomena: non-manual features (NMFs), classifiers, the spatial nature of sign language and its discourse mapping onto the signing space, topic–comment structures, and co-articulation of signs. It should also have an adequate notation system/literal orthography to describe the sign language, as they have no officially recognised written forms.

3.2 Sign Language Corpora

Corpora of sign languages are not widely available and the few that are often contain little or no annotation. Annotation is necessary as the corpora usually take the form of sign language videos owing to the lack of a standardised written form for SLs. This is one way in which SLMT differs from spoken language text based MT. SignWriting[3] may fill this gap as there are SL corpora available in this form. In terms of its suitability as a candidate for use in an EBMT system, SignWriting lacks the explicit linguistic detail necessary for the generation of signs using an avatar. Annotated corpora on the other hand have the potential to carry varying degrees of granularity of linguistic detail, therefore bypassing the need to translate using SingWriting and then deriving such details from the resulting SignWriting symbol. Another issue with SignWriting is that the majority of signers are unfamiliar with it which lowers its appeal for use as final output translation.

By contrast to poorly or unannotated data, the ECHO project is a pilot venture to make fully annotated digitised corpora available on the Internet. The project is based in the Netherlands and contains annotated corpora in NGT, British Sign Language (BSL) and Swedish Sign Language (SSL). These corpora have been annotated using ELAN annotation

[3] http://www.signwriting.org

software.[4] ELAN provides a graphical user interface (Fig. 1) from which corpora can be viewed in video format with their corresponding aligned annotations. These can be seen in the lower half of Fig. 1 where the tiers are named on the left-hand column and the annotations appear horizontally in line with their corresponding tier.

The ECHO corpora have been annotated to include a time-aligned translation in the native spoken language and in English. Other annotation tiers include glosses of the signs articulated by the right and left hand in both spoken languages and various NMF descriptions. An example of some annotations used in one of the NGT corpora can be found in (1). The initial numbers indicate the time span of annotation, the text in brackets shows the name of the annotation tiers and the final text is the annotation itself:

(1) a. `1459490 1461360`
`(Gloss RH English)CONSCIOUS`

b. `1459490 1461360`
`(Gloss RH) BEWUST`

c. `1459490 1461310`
`(Mouth) 'bewussssss'`

3.3 Suitability of ECHO Corpora

Suitably annotated corpora, such as those provided by the ECHO project, are ideal for use in an EBMT approach to SL translation. The provision of an English translation in the form of an annotation tier for each signed sentence along with the other time-aligned annotations allows for easy alignment of corpora on a sentential level as annotations within the time limits of the English translation annotation can then be aligned with that annotation. The presence of time spans for each annotation also aids in the aligning of annotations from each annotation tier to form chunks that can then be aligned with chunks derived from the English tier. Time-aligned annotations are also useful for tackling the issue of co-articulation of signs. The phenomenon of co-articulation in sign languages is analogous to co-articulation in spoken languages where the articulation of a phoneme may be altered relative to its neighbouring phonemes (Jerde et al., 2003). In sign languages phonemes are articulated using the hands. Examples of sign language phonemes include handshape and palm orientation. Co-articulation can occur in fluent signing when the the shape of the hand for one sign is altered relative to the handshape for the subsequent sign. Even if signs are co-articulated in the videos, the annotations for the signs will be separate and either contiguous or overlapping in different tiers, either way they are easy to separate using the time span figures. As it is these annotations that are used in the translation output, the issue of separating co-articulated words is removed automatically.

[4] `http://www.mpi.nl/tools/elan.html`

These annotated corpora also provide a solution to the SL translation issue of NMFs. In sign language meaning is conveyed using several parts of the body in parallel (Huenerfauth, 2005), not solely the hands which is a common misconception. NMFs are sign language units that use parts of the body other than the hands to express semantic information. Some examples of NMFs are eyebrow, cheek or eye movements, mouth patterns, head tilting or upper body and shoulder movements. They are used to express emotion or intensity, but also can be used morphologically and syntactically as markers (Ó'Baoill & Matthews, 2000). The annotations of the ECHO corpora contain explicit NMF detail in varying tiers such as eye aperture and mouth that combine with other tiers to form complete signs and therefore more linguistically complete translations. The example in (2) shows the effect NMFs have on a sign. The *Gloss RH/LH English* is the manual hand sign articulated by the right and left hand, in this case showing that of a hare running. The annotation *n* on the *head* tier indicates a nod of the head. This combined with the furrowing marked in the *brows* tier (signified by *f*), the squinting marked in the *eye aperture* tier (signified by *s*) and the puffing of the cheeks marked in the *cheeks* tier (signified by *p*) shows the intensity of the running that the hare is doing. Without these NMFs the hare would be understood to be running at a normal running pace.

(2)
```
(Gloss RH English) running hare
(Gloss LH English) running hare
(Head) n
(Brows) f
(Eye Aperture) s
(Cheeks) p
```

In many cases NMFs are essential for providing the full sense of the sign. The more detail that is contained in the annotation tiers, the better the translation and the more suitable the translations will be for use with a signing avatar. Currently, research is focused on the translation modules of the system and it is for this reason that annotations are produced as final input as opposed to a signing avatar.

4 Example-Based Machine Translation

A prerequisite for EBMT is a set of sentences in one language aligned with their translations in another. Given a new input string, EBMT models use three separate processes in order to derive translations:

1. Searching the source side of the bitext for 'close' matches and their translations;
2. Determining the sub-sentential translation links in those retrieved examples;
3. Recombining relevant parts of the target translation links to derive the translation.

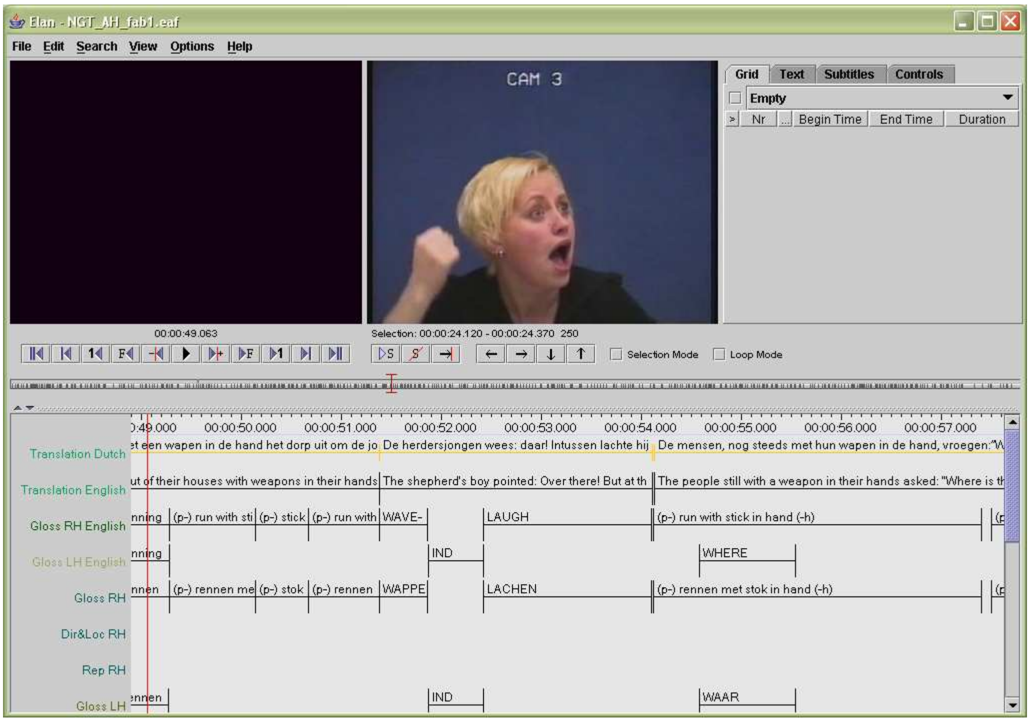

Figure 1: ELAN user interface

Searching for the best matches involves determining a similarity metric based on word occurrences and part-of-speech labels, generalised templates and bilingual dictionaries. The recombination process depends on the nature of the examples used in the first place: from aligning phrase-structure (sub-)trees (Hearne & Way, 2003) or dependency trees (Watanabe et al., 2003), to the use of placeables (Brown, 1999) as indicators of chunk boundaries.

Another method—and the one used in the EBMT system in our experiments—is to use a set of closed-class words to segment aligned source and target sentences and to derive an additional set of lexical and phrasal resources. (Way & Gough, 2003), (Gough & Way, 2004a) and (Gough & Way, 2004b) base their work on the 'Marker Hypothesis' (Green, 1979), a universal psycholinguistic constraint which posits that languages are 'marked' for syntactic structure at surface level by a closed set of specific lexemes and morphemes. In a pre-processing stage, (Gough & Way, 2004b) use 7 sets of marker words for English and French (e.g. determiners, quantifiers, conjunctions etc.), which together with cognate matches and mutual information scores are used to derive three new data sources: sets of marker chunks, generalised templates and a lexicon.

In order to describe this in more detail, we revisit an example from (Gough & Way, 2004a), namely:

(3) each layer has a layer number ⟹chaque couche a un nombre de la couche

From the sentence pair in (3), the strings in (4) are generated, where marker words are automatically tagged with their marker categories:

(4) <QUANT> each layer has <DET> a layer number ⟹<QUANT> chaque couche a <DET> un nombre <PREP> de la couche

Taking into account marker tag information (label, and relative sentence position), and lexical similarity, the marker chunks in (5) are automatically generated from the marker-tagged strings in (4):

(5) a. <QUANT> each layer has: <QUANT> chaque couche a

b. <DET> a layer number: <DET> un nombre de la couche

(5b) shows that *n:m* alignments are possible (the two French marker chunks *un nombre* and *de la couche* are absorbed into one following the lexical similarities between *layer* and *couche* and *number* and *nombre*, respectively) given the sub-sentential alignment algorithm of (Gough & Way, 2004b).

By generalising over the marker lexicon, a set of marker templates is produced by replacing the marker word by its relevant tag. From the examples in (5), the generalised templates in (6) are derived:

(6) a. <QUANT> layer has: <QUANT> couche a

b. <DET> layer number: <DET> nombre de la couche

These templates increase the robustness of the system and make the matching process more flexible. Now any marker word can be inserted after the relevant tag if it appears with its translation in the lexicon, so that (say) *the layer number* can now be handled by the generalised template in (6b) and inserting a (or all) translation(s) for *the* in the system's lexicon.

However, since SLs display a considerably reduced number of marker words, an alternative method is used for segmenting the SL texts. This is discussed in section 5.1.

5 Experiments and Results

Our corpus consists of 561 sentences with an average sentence length of 7.89 words, (min. 1 word, max. 53 words). The sign language side of the corpus consists of annotations that describe the signs used in the video. As the English translation annotation tier and the other annotation tiers are time-aligned, sentence alignments were easy to extract automatically.

5.1 Segmentation and Alignment

The Marker Hypothesis described in section 4 was used to segment the English sentences according to the same set of closed-class words used in (Way & Gough, 2003; Gough & Way, 2004a/b). This results in segments that start with a closed class word and usually encapsulate a concept or an attribute of a concept being described, for example the concept of darkness as shown in (7) where the angle-bracketed text refers to the marker tag representing the pronoun *it*.

(7) `<PRON> it was almost dark`

On the sign language side it was necessary to adopt a different approach as a result of the sparseness of the English closed class item markers in the SL text. This is normal in SLs, where often closed class items are not signed, as is the case with many determiners, or are subsumed into the sign for the neighbouring noun as is sometimes the case with prepositions. Initially experiments were performed on different divisions of the SL annotations. The NGT gloss tier was segmented based on the time spans of its annotations. The remaining annotations on other tiers were then grouped with the NGT gloss tier annotations within the appropriate matching time frame. It was found that these segmentations divided the SL corpus into concept chunks. Upon examination these concept chunks were found to be similar in form to the chunks that were formed using the the Marker Hypothesis on the English text and suitable for forming alignments, thereby providing a viable option for chunking the SL side of the corpus. The following example shows segments from both data sets and their usability for chunk alignment. (8) shows the results of the different chunking process on both sentences, (8a) being taken from the English chunking process and (8b) from the SL chunking process. (9) shows specific chunks that can be successfully aligned following the chunking process, (9a) being taken from the English chunked text and (9b) from the SL chunked text. Angled brackets contain the markers, round bracketed text names the tier, the remaining text is the annotation content of that tier and each tier is separated by a colon.

(8) a.
```
<DET> the hare takes off <PREP> in a flash.
```
b.
```
<CHUNK> (Gloss RH English) (p-) running hare :
(Mouth) closed-ao :
(Mouth SE) /AIRSTREAM/ :
(Cheeks) p :
(Gloss LH English) (p-) running hare :
(Gloss RH) (p-) rennen haas :
(Gloss LH) (p-) rennen haas :
<CHUNK> (Gloss RH English) FLASH-BY :
(Gloss RH) VOORBIJ-SCHIETEN :
(Mouth) closed, forward :
(Mouth SE) /PURSED/ :
(Eye gaze) rh
```

(9) a.
```
<DET> the hare takes off
```
b.
```
<CHUNK> (Gloss RH English) (p-) running hare :
(Mouth) closed-ao :
(Mouth SE) /AIRSTREAM/ :
(Cheeks) p :
(Gloss LH English) (p-) running hare :
(Gloss RH) (p-) rennen haas :
(Gloss LH) (p-) rennen haas
```

The main concept expressed in (9a) and (9b) is the running of the hare. The English chunk encapsulates this concept with the words *the hare takes off*. This same concept is expressed in the SL chunk in the combination of annotations. The 'Gloss RH English' and 'Gloss LH English' show the running of the hare and the additional semantic information of the effort involved in *takes off* as opposed to running at ease is expressed in the NMF tiers with the indication of puffing of the cheeks (*p* in the *cheeks* tier) and the closed mouth with breath being exhaled (closed-ao and */AIRSTREAM/* in the *mouth* and *mouth SE* tiers respectively). Despite the different methods used, they are successful in forming potentially alignable chunks.

5.2 Evaluation

As there is not formally recognised writing system for SLs and as annotation maybe be considered subjective to the author to a degree, it is uncertain that consistent gold standard sentences for evaluation purposes could be produced, (Huenerfauth, 2005).To better evaluate the performance of the system we decided to formulate our own test set. Test sets were manually constructed in four groups of ten sentences. The groups are as follows: (i) full sentences taken directly from the corpus, (ii) grammatical sentences formed by combining chunks taken from different parts of the corpus, (iii) sentences made of combined chunks from the corpus and chunks not in the corpus, (iv) sentences of words present in the corpus but not forming alignable chunks and of words not in the corpus. These test sets were constructed with a view to making the most of the limited data we had.

Each sentence was run through the translator and the resulting output manually evaluated based on the alignments of the corpus. The results are evaluated and divided into four categories depending on their quality: good, fair, poor and bad. Below is an explanation of the metric employed with examples using the sentence *it was almost dark*.

Good: contains the correct grammatical information (i.e. adverbs, prepositions that provide detail about the concept) and content (i.e. head noun or verb) information.

(10) `Gloss RH English`: DARK
`Gloss LH English`: DARK
`Mouth`: 'donker'
`Brows`: f
`Eye Aperture`: s.

Fair: contains the correct content information but is missing some of the grammatical detail.

(11) `Gloss RH English`: DARK
`Gloss LH English`: DARK
`Mouth`: 'donker'
(no brow or eye movement shown, alters meaning of phrase)

Poor: contains only some correct content information and either lacks grammatical detail or contains the incorrect grammatical detail.

(12) `Gloss RH English`: DARK
`Eye Aperture`: c.

Bad: contains an entirely incorrect translation.

(13) `Gloss RH English`: WHAT

5.3 Discussion

The manual evaluations performed on the test results show that the system is competent in translating sentences that occur fully intact in the corpus as would be expected from any EBMT system. These results also show that more than half the translations of sentences made up of chunks from the corpus provide reasonable-to-good translations. The system is able to segment the input and find adequate matches in the corpus to produce coherent translations for 60% of the sentences tested from (ii). This is also the case for almost a third of test sentences where data consists of combined corpus and external chunks (sentence type (iii)). The more data that is not present in the training set that is introduced in the test set the lower the rating, as can be seen from the results of type (iii) and (iv) where an increased amount of material not present in the corpus is tested. In these cases, translations are still produced but are of poor to bad quality. For sentence type (iii), only a third of the sentences were of fair quality. For sentence type (iv), more than two thirds of the translations were considered bad and the remainder poor. As with EBMT systems in general, were the corpus to be larger and to contain a richer word-level dictionary, the system would be able to produce closer, if not exact, matches for an increased number of chunks and words, thus improving the ratings. Currently the approach to aligning segments for the bilingual corpus is in its infancy. Further research and development in this area will also improve the quality of alignments and thus the translations.

6 Conclusions and Future Work

In light of the absence of documentation available to the Deaf in their first language, in this paper we aimed to test the applicability of EBMT techniques to SLMT with a view to developing a prototype MT system for SLs. Corpora of English-NGT data were obtained from the ECHO project website and their annotations were extracted. These annotations were then used as a written representation of NGT from which example alignments could be deduced following the segmentation phase. We found the Marker Hypothesis a sufficient approach for segmenting the English data but found it necessary to employ a time frame based technique to segment the SL annotations. We found that employing these segmentation approaches provided us with chunks of a similar format from which adequate alignments could be constructed for use in the translation process. Despite the small corpus and dictionary size, initial results are promising and indicate further development is plausible and worthwhile. Further research into the chunking and aligning processes, combined with an enhanced corpus and dictionary, will improve the quality of results and provide a clearer picture of the success of an EBMT system for sign languages. This prototype system has allowed us to identify some areas which require particular focus.

Subsequent to the work carried out to date, we intend to continue developing the system using the current language pair English–NGT. Initial plans include enhancing the annotation alignments by incorporating non-time-aligned annotations into the data set and using the information in the complete annotation set to determine closer matches with the English data and thus improve alignment at all levels. This should also allow for the automatic creation of generalised templates which would further aid the translation process. A large part of the work on this system will involve the improvement of the word-level dictionary. If possible, this task will be automated and the word alignments extracted from the corpora as opposed to an external source. We also intend to undertake increased manual evaluations of the translation results to determine specific problem areas that need work. Once a successful system has been produced for this language pair we intend to expand the system to translate from Dutch to NGT and to apply the system to other language pairs for which we have similar data, i.e. English-British Sign Language.

The ISL corpus under construction at the Centre for Deaf Studies (Dublin) will be much larger than the NGT corpus we are currently using and will contain richer annotations. The ISL corpus consists of roughly 20 hours of video data in comparison to the 40 minutes of the current NGT corpus we are using. This will allow for the creation of larger test–training sets, which should improve the results of the system on the basis that the more data a system has, the more possible matches can be found for input sentences. The richer annotations incorporated into the ISL corpus, including phonological information such as hand shape and palm orientation, will provide a more detailed translated output from which real sign language may be synthesised using an avatar. This is the ultimate goal for our work, to develop a fully automated text to sign language translation system where the signers can enter English written data and have it translated for viewing in their first language.

Acknowledgements

We would like to thank the anonymous reviewers whose valuable comments helped improve the quality of the paper. This research was partially funded by an IRCSET[5] PhD scholarship.

References

Britta Bauer, Sonja Nießen and Hermann Heinz. 1999. Towards an Automatic Sign Language Translation System. In *Proceedings of the International Workshop on Physicality and Tangibility in Interaction: Towards New Paradigms for Interaction Beyond the Desktop*, Siena, Italy.

Ralf Brown. 1999. Adding Linguistic Knowledge to a Lexical Example-based Translation System. In *Proceedings of the 8th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-99)*, Chester, England, pp.22–32.

Michael Carl and Andy Way (eds). 2003. *Recent Advances in Example-Based Machine Translation.* Kluwer, Dordrecht, The Netherlands.

Nano Gough and Andy Way. 2004a. Example-Based Controlled Translation. In *Proceedings of the Ninth EAMT Workshop*, Valetta, Malta, pp.73–81.

Nano Gough and Andy Way. 2004b. 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.

Thomas Green. 1979. The Necessity of Syntax Markers. Two experiments with artificial languages. *Journal of Verbal Learning and Behavior* **18**:481–496.

Angus B. Grieve-Smith. 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.

Mary Hearne and Andy Way. 2003. Seeing the Wood for the Trees: Data-Oriented Translation. In *MT Summit IX*, New Orleans, LA., pp.165–172.

Judith Holt. 1991. *Demographic, Stanford Achievement Test - 8th Edition for Deaf and Hard of Hearing Students: Reading Comprehension Subgroup Results.*

Matt Huenerfauth. 2004. Spatial and Planning Models of ASL Classifier Predicates for Machine Translation. In *The 10th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI 2004)* Baltimore, MD., pp.65–74

Matt Huenerfauth. 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

Thomas E. Jerde, John F. Soechting and Martha Flanders. 2003. Coarticulation in Fluent Fingerspelling. *Journal of Neuroscience* **23(6)**:2383–2393.

Ian Marshall and Éva 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.

Makoto Nagao. 1984. A Framework of a Mechanical Translation between Japanese and English by Analogy Principle. In A. Elithorn and R. Banerji (eds.) *Artificial and Human Intelligence*, North-Holland, Amsterdam, The Netherlands, pp.173–180.

Dónall Ó'Baoill and Patrick A. Matthews. 2000. *The Irish Deaf Community (Volume 2): The Structure*

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

of Irish Sign Language. The Linguistics Institute of Ireland, Dublin, Ireland.

Éva Sáfár and Ian 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.

Lynette Van Zijl and Dean 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.

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

Hideo Watanabe, Sadao Kurohashi and Eiji Aramaki. 2003. Finding Translation Patterns from Paired Source and Target Dependency Structures. In M. Carl & A. Way (eds.) *Recent Advances in Example-Based Machine Translation*, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp.397–420.

Andy Way and Nano Gough. 2003. Developing and Validating an EBMT System using the World Wide Web. *Computational Linguistics* **29**(3):421–457.

Liwei Zhao, Karin Kipper, William Schuler, Christian Vogler, Norman Badler, and Martha 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.