

JOINING HANDS: DEVELOPING A SIGN LANGUAGE MACHINE TRANSLATION SYSTEM WITH AND FOR THE DEAF COMMUNITY

Sara Morrissey and Andy Way

National Centre for Language Technology
School of Computing, Dublin City University, Glasnevin, Dublin 9, Ireland
Phone: +353 1 700 6914, Email: {smorri, away}:computing.dcu.ie

Abstract: This paper discusses the development of an automatic machine translation (MT) system for translating spoken language text into signed languages (SLs). The motivation for our work is the improvement of accessibility to airport information announcements for D/deaf and hard of hearing people. This paper demonstrates the involvement of Deaf colleagues and members of the D/deaf community in Ireland in three areas of our research: the choice of a domain for automatic translation that has a practical use for the D/deaf community; the human translation of English text into Irish Sign Language (ISL) as well as advice on ISL grammar and linguistics; and the importance of native ISL signers as manual evaluators of our translated output.

Keywords: sign language, machine translation, D/deaf accessibility

1. Introduction

In this paper, we discuss a data-driven approach to Sign Language Machine Translation (SLMT) for translating English text into ISL. We use this work as a vehicle to acknowledge and demonstrate the role members of the D/deaf¹ community play in the research of accessibility aids.

The remainder of the paper is constructed as follows. In section 2 we give a brief overview of ISL, the primary SL used in our work. Section 3 outlines the general SLMT process and overviews previous and current research in this area. A description of the choice of domain and the data processing is given in section 4 and our own system is described in section 5. In section 6 we discuss the experiments we have carried out, their evaluation and results. We conclude the paper in section 7 and outline the future direction of our work.

¹ It is generally accepted (Callow, 2007) that 'Deaf' (with a capital 'D') is used to refer to people who are linguistically and culturally deaf, meaning they are active in the deaf community, have a strong sense of a Deaf identity and for whom SL is their preferred language. 'deaf' (with a small 'd') describes people who have less strong feelings of identity and ownership within the community, who may or may not prefer the local SL as their L1. Hard of hearing (HOH) is generally used to describe people who have lost their sense of hearing later in life and have little to no contact with the deaf community or SL usage for various social and cultural reasons. The boundaries of these categories are fuzzy and people may consider themselves on the border of one or another depending on their experiences and preferences.

2. Irish Sign Language

The work described in this paper is primarily concerned with the translation of ISL. Despite being the predominant and preferred language of the ~5000 members of the Irish Deaf community, it remains unrecognised as an official language of Ireland. As a result the language and its inherent culture remains poorly resourced, although this is improving through the efforts of organisations such as The Irish Society for the Deaf who work to “empower and enable Deaf people to participate in positive action to further their independence and full participation in the community”².

Despite common misconceptions, ISL is more related to French Sign Language/La Langue des Signes Française than British Sign Language due to its introduction into the schooling system in the mid 1800s. The language then developed into a language in its own right. It uses a one-handed alphabet, makes broad use of initialised signs and displays little regional or geographical variation, as most people were educated in the same schools. Due to its low political and social status and the use of ‘oralism’ as a teaching method rather than ISL, there is little access to a standard form of the language for the D/deaf communities so the language is still considered to be marginalised and oppressed (Ó’Baioill & Matthews, 2000).

3. Sign Language Machine Translation

SLs worldwide lack political recognition (Gordon, 2005) and are poorly resourced in comparison to their spoken language counterparts. This is evident in the area of SLMT research with the earliest papers in this area dating back only 18 years. Fewer than 10 groups within this time have attempted SLMT and for the most part these projects have been short-lived with varying degrees of success. In general, SLMT has followed the trend of mainstream MT towards data-driven approaches over rule-based or more linguistic approaches. Below is a list of current international activity in this area:

- (Morrissey and Way, 2005, 2006), (Morrissey et al., 2007): our work has centred on using example-based methodologies as part of a data-driven framework, where we have worked with Dutch Sign Language and more recently ISL as described in this paper and (Morrissey et al., 2007)
- (Stein et al., 2006) use Statistical methods for translating German Sign Language in the domain of weather reports. Their work involves a number of pre- and post-processing steps
- (Chiu et al., 2007) also present a Statistical approach for their work with Chinese and Taiwanese Sign Language
- (San-Segundo et al., 2006) propose a speech to gesture architecture for their work on Spanish to Spanish Sign Language
- (Huenerfauth, 2006) has primarily focused on American Sign Language generation and processing classifier predicates in MT

Essentially, an SLMT system for translating spoken language text into an SL will take a sentence as input, run it through the system looking for the most likely translation based on pre-described linguistic rules or the best statistical match, for example, and reproduce the sentence in a textual format of the SL. Some systems stop at this point of the process, focussing mainly on the translation process; others fit an avatar to the text output that will sign the translated sentence in real SL

Previous approaches have varied in their practical applications with many focussing on linguistic phenomena and are too broad in scope to be of use in real world situations. Some, such as the work of Stein et al. (2006), have constrained their translation to the area of weather reports, which itself has a limited practical use. The work of Morrissey & Way (2006) notes the practical limits of a SLMT system based on the domain of children’s fables and poetry using data from the ECHO project³. With a view to improving the practical applications of such SLMT systems, we have developed a data-driven MT system for translating English into ISL for the domain of airport announcement information. This choice of closed domain facilitates better data-driven MT as it has small vocabulary. The choice of this domain also facilitates its target user group as highlighted in section 4.1.

² <http://www.deaf.ie>

³ <http://www.let.kun.nl/sign-lang/echo/data.html>

4. Aiding Airport Announcement Accessibility for Deaf and Hearing-Impaired People

4.1 Choosing a Domain

One of the important goals of our research is to develop a translation system that is has a practical use within the D/deaf community. To help achieve this, we have sought the guidance and assistance of our colleagues in the Centre for Deaf Studies, Dublin. Given their personal experiences, we have identified the domain of airport information announcements as one where an SLMT system that translates such information directly into SLs could facilitate the D/deaf and HOH communities.

Typically, airport information announcements, such as gate changes and delays, are announced over a PA system and often such information does not appear on screens for some time if at all. This causes considerable hindrance to D/deaf and HOH people who may be left uninformed and possibly inconvenienced through no fault of their own. To help alleviate this, we chose to orient our MT system specifically toward this domain. In an airport scenario, should there be alterations to flight information, the relevant piece of information could be typed into the MT system, which would automatically translate the sentence into ISL. The information would then be signed in real ISL by a mannequin on video screens for people to view. An example of the generated mannequin that would sign the output is in Figure 1 taken from Poser 4 ProPack Software⁴.



Figure 1 Example of signing avatar

4.2 Data Selection and Processing

A data-driven approach such as ours necessitates a corpus of text in the source and target language. Having chosen our domain, we found a suitable base corpus in the ATIS (Hemphill et al., 1990) dataset, a corpus that is frequently used in NLP and was derived from a speech dialogue system of air traffic queries and responses (e.g. “What flights are there from Cork to Dublin?”). Our translation methodology requires a *bilingual* dataset. As the ATIS corpus is an English corpus it was necessary for us to translate the original datasets into ISL. To ensure the authenticity of our data, we liaised with the Irish Deaf Academy to employ two native Deaf ISL signers for translation and consultation work. During this process, the signers were encouraged to translate the sentence into an authentic ISL sentence irrespective of the choice of English words and grammar. In order to ensure fluency and consistency, each translation and signed sentence was discussed between the signers.

The lack of a formalised writing system for SLs leads to the issue of how to represent them during the translation process. Having considered methods such as Stokoe Notation (Stokoe, 1960), HamNoSys (Prillwitz, 1989) and SignWriting⁵, we have chosen to use manual gloss annotation to transcribe the ISL video data of the ATIS corpus for its adaptability. For this we used the ELAN video annotation toolkit⁶ to transcribe a semantic representation of the ISL in the videos. An example of an annotated sentence ‘*Early morning flights between Cork and Belfast*’ taken from the corpus is shown in (1).

⁴ <http://www.curiouslabs.com>

⁵ <http://www.signwriting.org>

⁶ <http://www.mpi.nl/tools/elan.html>

(1) EARLY MORNING BETWEEN be-CORK CORK FLY BELFAST BETWEEN ref-BELFAST ref-CORK

Given that annotating SL data is a time-consuming process, at this stage we have kept the level of detail to the basic semantic representation of the signs without non-manual or phonetic feature detail. It is intended that once a satisfactory level of translation has been achieved, more comprehensive features will be added.

5. System Description

Our data-driven approach to SLMT makes use of the MATREX MT system (Stroppa et al., 2006) developed at Dublin City University. It combines Statistical MT and Example-Based MT (EBMT) methodologies and has a modular design that makes it particularly adaptable, as modules can be extended or reimplemented at various stages in the translation process. This modularity also means that it is particularly adaptable to the translation of different language pairings or domains as one need only 'plug in' a bilingual corpus of the relevant languages on the desired topic. A diagram illustrating the architecture of our system is shown in Figure 2 below.

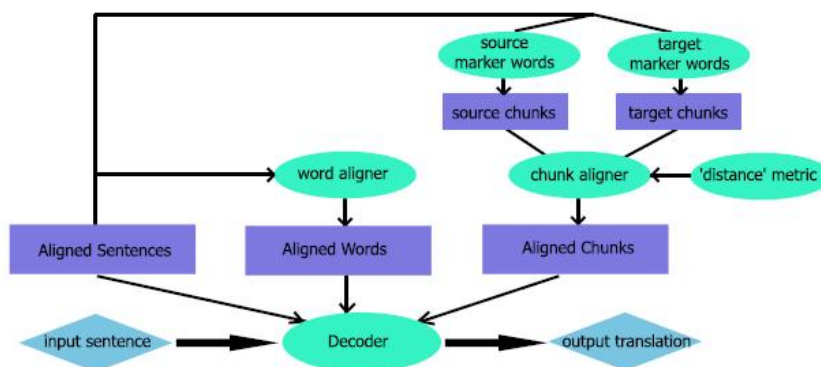


Figure 2 Architecture of the MaTrEx MT system

Within the system the 'decoder' is the main engine that takes an English sentence, for example, as input and produces the best ISL sentence it finds in annotated format. The decoder is fed by and makes its translation estimations based on three information pools of aligned data: groups of aligned sentences, aligned words and aligned chunks retrieved from the bilingual corpus. The statistical word alignment toolkit GIZA++ (Och, 2003) is used to derive singular word alignments between source and target. To derive sub-sentential chunk alignments, we primarily use the Marker Hypothesis (Green, 1979), where a stop list of closed class lexical 'marker' words are chosen as delimiters to segment each sentence from the source and target corpus. A minimum edit distance metric is then used to align potential matching chunks based on the least number of substitutions, insertions and deletions when the two candidates are compared.

During the translation process, the decoder takes in an input sentence, searches its three alignment databanks for candidate matches on a sentential, sub-sentential chunk and word level. MOSES (Koehn et al., 2007) is used to deduce the most likely phrase for translation, which is then produced as output.

6. Experiments

We have carried out three experiments on the data described in section 4.1, translating English into ISL. The first is the baseline system that employs the modules described in section 5 with the exception of the EBMT chunks.

The subsequent experiments exploit two EBMT techniques in an attempt to improve on the baseline SMT system. The first, 'chunking method 1', uses the Marker Hypothesis outlined in section 5 to segment both the source and target data and the resulting chunks and alignments were added to the system. The second, 'chunking method 2', takes into account the natural lack of closed class lexical

items in SLs and segments the ISL data so that each ISL word forms its own chunk which is then aligned with English chunks that have been derived using the Marker Hypothesis.

6.1 Results and Evaluation

At this stage of the translation process, automatic evaluation of the annotated output allows us to get an objective view of how the system is performing without annotation-to-avatar noise interfering. Evaluation scores for the test data of 118 sentences are shown in Table 1.

Given that translated output takes the form of annotated ISL, we have been able to automatically evaluate the output against a set of 'gold standard' annotations withheld for this purpose. The sentences produced are evaluated based on two types of error rate calculations. Word error rate (WER) computes the distance between the reference and candidate translation based on the number of insertions, deletions and substitutions in the words, divided by the number of correct reference words and takes the word order into account. Position-independent word error rate (PER) calculates the same distance as the WER but discounts word position. For error rates, a lower percentage score indicates better translations.

	WER %	PER %
Baseline	41.68	32.53
Chunking Method 1	40.96	29.75
Chunking Method 2	40.60	31.80

Table 1 Automatic Evaluation Scores for English to ISL MT using MaTrEx

6.2 Discussion

First testing of the system at baseline level indicates that the system does a reasonable job of translating English into ISL with scores comparable to mainstream speech-to-speech systems. At this level, more than two thirds of the words produced are correct and almost 60% of the time the word order is also correct.

Using the Marker Hypothesis to segment sentences improves both WER and PER scores, the latter by approximately 3% showing an increase in the number of correct words in the candidate translations. The effect of the second chunking method, while it does lower both error rate scores, is not as successful as the first methods in terms of the PER and only improves the WER by 0.36%. These results show that sub-sentential chunking of the training data improves the translation.

7. Conclusions and Future Work

Our collaboration with members of the Deaf community as both consultants and facilitators has allowed us to effectively channel our research in the area of SLMT towards a practical goal, namely aiding D/deaf and HOH people in the accessibility of airport information announcements.

To date, our research has primarily focused on the development and improvement of MT processes with the translated output being produced in annotated format. For the system to be of practical use to its intended users a signing avatar is required. With a view to this, it is intended to expand the annotation to include descriptive phonetic features of the signs which could feed into Poser 4 software shown in Figure 1 in section 4.1 to create on-the-fly signed sentences. The use of such a human-like mannequin allows for real ISL to be produced as similar to the natural language as possible. Ideally, in fully functioning software for airport use, announcements would be appearing on the screen in both text and avatar cover the preferences of the Deaf, deaf and HOH communities.

This final stage necessitates manual analysis by native ISL signers. For this, we propose the use of formalised accuracy and fluency scales for evaluating the translated output. This will allow us to assess the performance of the system in terms of complete translation but also to gauge the practical usability of our work as the evaluators are from the intended user group.

References

- Callow, L. (2007). Deaf Awareness Session at DCAL Summer School on Deaf Studies and Sign Language Research: Researching language and communication in another modality, London, United Kingdom.
- 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*, vol. 29(1), pp; 28–39.
- Gordon, R.G. Jr. (Ed.) (2005). *Enthnologue: Languages of the World*, Fifteenth Edition. Dallas, Texas.: SIL International.
- Green, T. (1979). The Necessity of syntax markers: two experiments with artificial languages, *Journal of Verbal Learning Behaviour*, vol. 18, pp.95–104.
- Hemphill, C., J. Godfrey and G. Doddington. (1990). The ATIS spoken language systems pilot corpus. In *Proceedings of the workshop on Speech and Natural Language*, pp. 96–101, Hidden Valley, Pennsylvania.
- Huenerfauth, M. (2006). *Generating American Sign Language Classifier Predicates for English-to-ASL Machine Translation*. Doctoral Dissertation, Computer and Information Science, University of Pennsylvania.
- Koehn, P., M. Federico, W. Shen, N. Bartoldi, O. Bojar, C. Callison-Burch, B. Cowan, C. Dyer, H. Hoang, R. Zens, A. Constantin, C. Moran and E. Herbst. (2007). Open source toolkit for statistical machine translation: factored translation models and confusion network decoding, In *Final Report of the Johns Hopkins 2006 Summer Workshop*.
- 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 2006*, pp. 91–98, Genoa, Italy.
- Morrissey, S. and A. Way. (2007). Towards a hybrid data-driven MT system for sign language translation. In *Proceedings of the MT Summit XI, (forthcoming)* Copenhagen, Denmark.
- Ó'Baoill, D. and P. Matthews. (2000) *The Irish Deaf Community (Volume 2):The Structure of Irish Sign Language*. The Linguistics Institute of Ireland, Dublin, Ireland.
- Och, F. (2003). Minimum error rate training in statistical machine translation. In *Proceedings of ACL 2003*, Sapporo, Japan, pp.160–167.
- Prillwitz, S. (1989). *HamNoSys Version 2.0: Hamburg Notation System for Sign Language. An Introductory Guide*. Signum Verlag.
- 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). The architecture of an English-text-to-sign-languages translation system, In *Proceedings of the 11th Annual conference of the European Association for Machine Translation (EAMT, '06)*, pp. 169–177, Oslo, Norway.
- 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, NY., [revised 1978 Lincoln Press]
- 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.

Acknowledgements: We would like to thank the anonymous reviewers whose valuable comments helped improve the quality of this paper. We would also like to thank staff and colleagues at the Centre for Deaf Studies, Dublin⁷ and the ISL Academy⁸ for their guidance and assistance. This research is funded by a joint IRCSET⁹ and IBM¹⁰ PhD scholarship.

⁷ <http://www.centrefordeafstudies.com>

⁸ <http://www.deaf.ie/ISLAcademy.htm>

⁹ <http://www.ircset.ie>

¹⁰ <https://www-927.ibm.com/ibm/cas/sites/dublin/>