Assessing Three Representation Methods for Sign Language Machine Translation and Evaluation

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

Abstract

Having no standard written format, sign languages must be transcribed in some way in order to be processable for machine translation (MT). Previous research into MT for sign languages (SLs) has shown little consistency or agreement on the appropriate transcription methodology for the SLs. In this paper, we take a corpus of 200 SL utterances and explore the effects of three different representations on the MT process using a randomly and a specially selected testset. We use the DCU MaTrEx MT system and show that using an XML-based markup achieves the best results over other formats in terms of BLEU scores. We discuss the meaning of these results in the context of evaluating a representation of a language as opposed to the final form.

1 Introduction

It is becoming increasingly accepted that sign languages (SLs) are natural, indigenous languages of the Deaf[1] communities worldwide. Along with this recognition, whether at a government or more local level, research in the area is beginning to expand both in terms of linguistic analysis and assistive technology. Although SL linguistics and SL technology are playing catch up when compared with those of spoken languages, there is a growth in research groups addressing the issues of SL analysis and language barriers through the development and adaptation of technology including machine translation (MT).

MT has the potential to improve communication for Deaf people through the automatic translation of text and speech into SL. While MT offers a whole host of possibilities to assist the communication with and of Deaf people worldwide a number of issues must be tackled in order to address this challenge. In this paper we address two of those issues, namely transcription and evaluation.

Given that there are no generally accepted written forms of SLs, this poses a problem for automatic processing of data. Of course, means of transcribing and representing SLs do exist (as discussed in Section 2) but none are widely accepted or considered adequate representations across the board and all have their limitations. This further complicates the task of MT. Large amounts of transcribed data are difficult to source, particularly if they are required for a particular domain. Typically researchers resort to the creation of databases themselves either through role play, collecting videos or translation of text into SL videos. Given that the collection and transcription process is so time-consuming this normally results in significantly small amounts of data for training and testing. When considered against the comparably gargantuan training sets used in spoken language MT systems with sentence pairs in the order of millions, the size of SL training data, with average training sets being composed of a few hundred to a few thousand sentence pairs, will have an impact on the ability of the MT engine to successfully perform translation on new data. Less data means fewer alignments from which to translate new data and much of the new data is often unseen and untranslated. We address this challenge

[1] In this paper we use the term 'Deaf' (with a capital D) as it is generally accepted (Leeson, 2003) and 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.

Mikel L. Forcada, Heidi Depraetere, Vincent Vandeghinste (eds.)
Proceedings of the 15th Conference of the European Association for Machine Translation, p. 137–144
Leuven, Belgium, May 2011

through our experiments and discussion in sections 5 and 6 where we show that many SL text-based formats are indeed viable options from a translation point of view and indeed produce promising evaluation scores.

The second main challenge facing SL MT, is that of evaluation. Given that SLs are visual-gestural languages, can it be said that automatic evaluation of representational output sufficiently qualifies to assess the quality of the output, especially considering that most forms of SL text-based representation are considered to be lacking in vital information for the production of the signed utterance? An alternative to automatic evaluation of text-based output is to have the output signed by an animated avatar. While this goes some way to addressing the gap between representation and actual signing, it is not without its own criticisms. In the discussion section of this paper we compare three transcription approaches and their automatic evaluation scores. We will show that automatic evaluation has its place but that human evaluation is integral to the proper assessment of output quality.

The paper is constructed as follows: section 2 introduces SL transcription for MT and covers a number of frequently used options. Section 3 discusses related research in the area of data-driven MT for SLs. We then give a description of the MT engine we use in section 4, and go on to describe our experiments using two different data sets as well as discussing our corpus creation and chosen transcription methods in section 5. The results are discussed in section 6 and in addition we comment on evaluation methods. Finally, we conclude in section 7.

2 Sign Language Transcription in Machine Translation

There is no agreed standard text-based representation for SLs. Although many transcription formats exist, some of which we discuss below, all are considered inappropriate or inadequate to some degree and are not universally adopted.

Although many approaches attempt to capture the linguistic components of a sign, the primary representation used for MT is the linguistic practice of 'glossing', meaning notating one language using (generally) another language. Glossing was initially used by SL linguists to notate SL videos for ease of reference and searching (i.e the Signs of Ireland corpus (Leeson et al., 2006)), but more

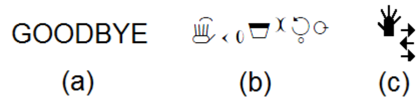

Figure 1: Images depicting gloss, HamNoSys and SignWriting representations of the sign 'goodbye'.

recently this notation form has been adopted for use in MT. Software such as the EUDICO Linguistic Annotator (ELAN)[2] allows simultaneous video viewing and manual notation with the option of multiple tiers of description as fine-grained as the user chooses.

The other two most widely recognised representations for SLs are the transcription language HamNoSys (the Hamburg Notation System) and SignWriting. HamNoSys (Hanke, 2004) uses a set of language-independent symbols to iconically represent the phonological[3] features of the SL including handshape, orientation, movement and location as well as some non-manual feature functionality. SignWriting (Sutton, 1995), on the other hand, was developed as a method of writing movement for dance and subsequently moved into the area of SL representation. It was created as a handwriting system and although it describes the signs phonologically, the system is more pictorial than the linear HamNoSys, using simple line drawings and arrows to represent parts of the body and movements or touching. This method is considered to be more intuitive and easier to learn and is expanding internationally through some research centres and schools for the Deaf. An example of each representation format for the sign 'goodbye' is shown in Figure 1 below. (a) shows the gloss representation, (b) shows the HamNoSys symbol equivalent and (c) shows the SignWriting version.

3 Related Research

Compared to mainstream MT for spoken languages[4] SL MT is still in the early stages with re-

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

[3]The term 'phoneme' in the area of SLs is analogous to its usage in spoken language in that it denotes the basic units of the articulated version of the language. Initially the term 'chereme' was coined for this purpose, but phoneme and its related terms are now in common usage for SLs and the term chereme is obsolete.

[4]Note that the term 'spoken languages' in the context of this paper refers to languages that are spoken as opposed to signed, and in this case does not specify only speech MT but also any text-based MT of those languages.

search in the area stretching back only about 20 years. Much research, with a few exceptions, has been the result of sporadic short-term projects as opposed to long–term research investment. While initial investigations (most probably following the predominant model of MT at the time, but also following the developing area of linguistic research into SLs - an area which only began in the 1960s with the seminal work of (Stokoe, 1960)) were focussed on rule-based methods, the area is shifting towards data-driven approaches following mainstream MT trends despite the fact that large amounts of SL data are difficult to source. For reasons of brevity, scope and relevance, only data-driven approaches that are comparable to the research described in this paper are discussed below. Less recent rule-based and interlingua approaches include the work of (Veale et al., 1998), (Speers, 2001), (Marshall and Sáfár, 2003), (Huenerfauth, 2004), (van Zijl and Fourie, 2007).[5]

Research at RWTH Aachen includes both SL recognition, Deutsche Gebärdensprache (DGS)/German Sign Language to German and German to DGS translation directions. Their research predominantly focuses on their weather domain corpus (RWTH-Phoenix-v3.0). Current work is focussing on adapting their techniques to working with small-scale data (Stein et al., 2010) and employing both phrase-based and hierarchical MT methods including a combined approach. Automatic evaluation scores (BLEU) range between 27% and 22% for glossed data with the combined approach showing the best score. No manual evaluations on recent work is documented.

Although previous work was rooted in a rule-based approach, (San-Segundo et al., 2010) have more recently redirected their focus to include opensource SMT tools. Using a corpus of just over 4,000 dialogue utterances in the domain of Drivers' Licence and Identity Document renewal they employ a rule-based model as well as a phrase-based SMT and finite state transducer SMT model. Using gloss annotation for testing and training, they report the rule-based approach achieved the best result with a BLEU score of 68%. A mock-up scenario was created to manually assess the animated output of their system. An exceptionally high translation accuracy of >90% is reported.

A combination of SMT and Example-Based MT (EBMT) approaches was explored by (Morrissey, 2008). The MATREX MT system from DCU (described in the next section) was employed. The EBMT approach was based on the Marker Hypothesis (Green, 1979). Multiple experiments were carried out comparing SMT and EBMT approaches, comparing 4 language pairs, translating in both directions for each (English–Irish SL (ISL), German–DGS, English–DGS, German–ISL). They report a BLEU score of 38.85% for EN–ISL experiments on an Air Traffic Information System corpus consisting of 595 sentences. They also perform EBMT experiments on the same data, but subsequent research showed that, despite some improvement in BLEU scores, the results were not statistically significant. Manual evaluations for intelligibility and fidelity to the source sentence were also carried out resulting in over 40% of the animated translations being considered intelligible and an excellent representation of the source sentence.

4 The MaTrEx System

Having described previous approaches, we now detail the MT system used in the experiments in the next section. MATREX (Machine Translation using Examples) is the data-driven MT system developed at the National Centre for Language Technology, DCU (Stroppa and Way, 2006). It is a hybrid system that can avail of both EBMT and SMT approaches by combining the resulting chunk- and phrase-alignments to increase the translation resources. We use only the SMT modules of the MATREX system, which is a basic wrapper around MOSES (Koehn et al., 2007),[6] in the experiments described in the next section, no EBMT experiments have yet been carried out.

The MATREX system is modular in design consisting of a number of extensible and reimplementable modules that can be changed independently of the others. This modular design makes it particularly adaptable to new language pairs and experiments can be run immediately with new data without the need to create linguistic rules tailored to the particular language pair at hand. It also facilitates the employment of different chunking methods to facilitate an EBMT approach, as de-

[5]Unfortunately, most early approaches pre-date the mainstream use of MT evaluation metrics, simply did not use them, or evaluations were manual inspections of the signing avatar and are therefore not comparable to the work here.

[6](http://www.statmat.org/moses)

scribed in our previous work (Morrissey, 2008). The databases of stored entities feed the decoder, including aligned sentences, phrases and words as well as chunks for EBMT, although not used in this implementation.

4.1 Word and Phrasal Alignment

The word alignment module takes the aligned bilingual corpus and segments it into individual words. Source words are then matched to the most appropriate target word to form word-level translation links. These are then stored in a database and later feed the decoder.

Word alignment is performed using standard SMT methods, namely GIZA++ (Och and Ney, 2003), a statistical word alignment toolkit employing the "refined" method of (Koehn et al., 2003). The intersection of the unidirectional alignment sets (source-to-target and target-to-source) provides us with a set of confident, high-quality word alignments. These can be further extended by adding in the union of the alignments. *n*:*m* alignments are produced here. Probabilities for the most likely translation alignments are estimated using relative frequencies.

The phrase alignment module is based on that of MOSES where phrases are extracted using the word alignments and scored using phrase translation probabilities and lexical weighting.

4.1.1 Decoder

Source language sentences are translated into target language sentences via a decoder module. We have used a wrapper around MOSES, a state-of-the-art phrase-based SMT decoder. The decoder chooses the best possible translation by comparing the input against the source side of the bilingual databases that feed it.

5 Experimental Setup

In this section we describe the experiments we have carried out using MATREX. We begin by outlining our dataset, its construction and the individual transcription methods chosen, before describing the experiments themselves.

5.1 Corpus Description

Our work centres around developing assistive technology for patients with limited English, given that the primary barrier to healthcare for non-native speakers is that of language. Avoiding the sensitive area of direct doctor-patient dialogue, we instead favour the domain of appointment scheduling and use a 5-way parallel multimodal corpus of English–Irish Sign Language (ISL) medical receptionist dialogue (Morrissey et al., 2010) on which to base our general work on minority language assistive technology. The 5 versions of the corpus are: audio recordings, English transcriptions, ISL videos, HamNoSys transcriptions, Signing Gesture Markup Language (SiGML) notation.

The data was first collected during a role-play session with a medical receptionist. Collecting real data in such a sensitive context was unpracticable as was the training and hiring of 'standardised patients', actors that are trained to perform as a real patient would. Audio recordings of the original material was made and subsequently transcribed into English. A native ISL signer manually translated and signed the dialogue in ISL and the results were recorded on video with the aid of a second ISL monitor. These videos were then transcribed using HamNoSys via the eSignEditor tool developed at the University of East Anglia (Kennaway et al., 2007). This software tool provides a platform for writing HamNoSys script, offering tables of HamNoSys symbols to select as appropriate to the sign being transcribed. Transcriptions can be stored in a lexicon with an attached gloss term for ease of referencing. Each individual sign transcription is also assigned an ID code. The eSignEditor also functions as an animation tool, transforming HamNoSys symbols into an XML code called SiGML. This in turn can drive an animation engine that produces a moving avatar that signs according to the HamNoSys code it is fed. The glosses, ID codes and SiGML description provide the various transcription methods we use in the experiments to follow.

Below we describe the three experiments and the reasoning behind them. Two data sets were selected for testing. The first was a set of 22 randomly chosen sentences removed from the training set using a 90:10 training:testset split. Given the especially small training data set, and the propensity of phrases in the domain to be of a similar nature, a special test set of 32 sentences was created using vocabulary and phrase structures from the training data, so that the test and training data are mutually exclusive at sentence level, all vocabulary in the test data is present in the training data. The same training sentences were used for both experiments, the training set consists of 199 parallel aligned English-ISL sentence pairs. No development sets were used for tuning in these experiments, for two reasons. Firstly, removing a portion of the training data for development would reduce an already meagre dataset further. Secondly, the use of such a small devset (say 30 sentences, roughly comparable to the size of the testset) is likely to lead to overfitting of the MT engine to those specific sentences and potentially diminish evaluation scores. We plan in future to investigate the use of comparable data for this purpose.

An example of the eSignEditor HamNoSys SiGML code from which the sign IDs, glosses and HamNoSys tags are taken for the three experiments is shown below.[7] For reasons of space our example is a one-word phrase (not to be confused with the contents of the individual words in the lexicon):

(1) ⟨ sign gloss "GOODBYE" signid "16"⟩
⟨ mouth⟩ Ba:⟨ /mouth⟩
⟨ src editable="false"/⟩
⟨ gol editable="false"/⟩
⟨ loc editable="false"/⟩
⟨ hand/⟩ ⟨ limbs/⟩ ⟨ facialexpression/⟩
⟨ hamnosys⟩ hamflathand, hambetween, hamfinger2345, hamextfingerul, hampalmr, hamlrat, hamchest, hamclose, hamparbegin, hammover, hamarcu, hamreplace, hamextfingerur, hampalml, hamchest, hamlrbeside, hamclose, hamparend
⟨ /hamnosys⟩ ⟨ /sign⟩

5.1.1 Sign IDs

We first chose to explore translation via sign ID numbers. As mentioned above, each sign in the corpus (variations included) has an individual and unique ID code attributed to it within the eSignEditor system. Based on the argument of spoken language glosses potentially misrepresenting signs (Pizzuto et al., 2006), we decided to use this non-language-based alternative to represent the signs in the translation process. The sign ID for each sign in an annotated sentence is extracted and forms the new text-based representation of that sentence in both the training and reference texts. An example of the sign ID format for the above phrase 'goodbye' is shown in (2a). The second example in (2b) shows the sign ID sequence matching the phrase 'When do you want to come in?'.

(2) a. `16`
b. `15 27 17 18 15`

Using sign IDs allows for the detailed description of the sign (provided by the associated HamNoSys and SiGML code, as indicated in (1) previously) to remain intact and outside of the translation process. The MT output produced would also take the form of sign IDs and these can then be searched in a stored lexicon of sign IDs and corresponding SiGML code, which would ultimately be joined with the other sign IDs from that particular output and be produced as one single animated video sequence.

5.1.2 English Glosses

Given that by fact of using eSignEditor we had access to a gloss-based representation of each sign, for comparative purposes, we extracted the upper case glosses from the SiGML output to use as the next text-based representation. This allowed us to draw a more concrete comparison between the use of ID tags and glossing than comparing results with previous experiments on different data. It should be noted that sign IDs and glosses are parallel, for each sign there is one unique gloss and one unique ID tag that correspond for all instances of that particular version of the sign. Corresponding glosses for the above examples are shown in (3a) and (3b) respectively.

(3) a. GOODBYE
b. WHEN WANT COME_AS_1 IN WHEN

For the purposes of MT, all glosses are converted to lower case. The suffix '_AS_1' refers to an alternative sign for 'come'. If there is more than one distinctly different sign in the database with one meaning, a suffix is added to distinguish them. In a similar way to using ID tags, we propose that upon production of gloss output the gloss terms may be searched in a lexicon database and the corresponding SiGML code joined and reproduced via the signing avatar.

5.1.3 SiGML code

SiGML code provides us with the HamNoSys tags that directly correspond to the HamNoSys symbols used to describe the phonetic features of

[7]It must be noted that each of the representations described below are parallel and describe the same unique objects but using different formats. It is these different representations of the same signed concepts that we are interested in comparing and assessing which is best handled by the MT process.

the signs in the ISL videos. Our next approach involves extracting those HamNoSys tags from the SiGML code and using these constructions to represent each sign. An example for the sign for 'goodbye' is exactly what is shown in (1).[8] Given the verbose nature of the HamNoSys tags, for reasons of space we will not show an example of a full sentence. In order distinguish between each individual sign, we substitute all spaces in the code with underscores '_' and insert a space ' ' in between the code for each sign to delineate each one clearly and tokenise the data.

6 Evaluation and Discussion

For testing, training and test sets were created based on ID numbers, glosses and SiGML code respectively. Translating from English to ISL, the output of each was evaluated against one reference sentence using BLEU (Papineni et al., 2002), word error rate (WER) and position independent WER (PER) were also calculated. Table 1 below shows the evaluation results for the 6 experiments carried out above. Random and special testsets are indicated respectively by Test^R, Test^S.

	System	BLEU %	WER %	PER %
ID nums	Test^R	3.82	112.09	112.09
	Test^S	16.05	104.86	104.86
Glosses	Test^R	31.84	80.37	73.20
	Test^S	43.03	61.33	48.80
SiGML	Test^R	55.43	54.46	46.10
	Test^S	45.64	54.79	46.10

Table 1: Evaluation Scores for Medical dialogue corpus EN-to-ISL experiments

6.1 Discussion

From the above results we see that using ID numbers gives by far the lowest evaluation scores across the board.[9] Examination of the output texts shows that large amounts of source text are present indicating no translation has taken place for at least some part of most sentences. Gloss and SiGML datasets achieve more comparable scores with SiGML achieving an improvement of almost 24 BLEU points over glossing for the random test set. As anticipated the specially selected test set achieves better results than the randomly selected testset, with the exception of the BLEU score in the SiGML experiments, however the error rate scores are almost the same for both SiGML testsets. Data sparseness in the randomly selected testset would account for this.

Although it appears from first look that the SiGML experiments illustrate their superiority as a choice of transcription method, an examination of the output and SiGML code in general shows that there is a significant overlap in the code (tags) for each individual sign in a sentence to the extent that the difference between the code for one sign and the next is a matter of different tags, ID numbers and gloss words.[10] The similarity of tags and code grossly outweighs the differences resulting in the inflated BLEU scores. The frequent repetition of these tags (regardless of the actual sign) means the presence of an incorrect tag in the output will have an almost insignificant effect on the evaluation result. For this reason we believe that a direct comparison via traditional MT automatic evaluation metrics of the output SiGML code against a reference set of the same format is not sufficient to ascertain if the translation quality is good or not.

Although glosses are not considered an adequate representation of a sign (or indeed of SL) by many, despite the small dataset, the results are comparable with those in the related research literature and may indeed, based on the previous assumption that gloss terms would have an equivalent animatable form in a predefined lexicon, provide clear and correct SL output. Given the above consideration of SiGML evaluation, it appears that until a more appropriate evaluation method is developed, glosses are the most effective means of translating SLs.[11] However, we must bear in mind the caveat that all these means of representation may not appropriately encapsulate inflectional and other linguistic information (if at all) that is integral to the understanding of a signed utterance. Until the output is produced via a signer or signing avatar if is difficult to ascertain what information is present, and indeed this is one of the arguments for the inadequacies of transcription methods for SLs, that the complex gestural-spatial performance nature of

[8]The order of the tags is significant, not to HamNoSys, but to the eSignEditor. Incorrect ordering of these tags can result in the sequence not being accepted for animation.

[9]It is possible to obtain a WER and PER of more than 100% if there are fewer words in the reference translations than in the candidate translations.

[10]Note that the order of the tags is important for later input into the animation system.

[11]This is strictly from an MT point of view and what is of merit to MT may not be of merit to the SL itself.

SLs and the linguistic artifacts that are concomitant with the performance (including non-manual features such as eyebrow movement and head-tilts, for example) may be omitted. However, given the limited range of representation methods available, it is up to MT researchers to make what is available as effective as possible.

6.2 Comment on Evaluation Methods

Evaluation is integral to the completion of any MT experiment to assess its performance, if there is only one reference text this is somewhat limiting. But with SL MT we are faced with the interesting fact that the MT output is not in fact the final product, but an intermediate representation of it. In fact there are two opportunities for evaluation in SL MT, evaluating text-based and animated output. In the first instance, automatic evaluation metrics (as used in our experiments above) can be employed to assess the text-based representation output of the MT engine with varying success. This method cannot be said to evaluate the actual translation of the input sentence into the SL as the output is a mere representation, not the actual language itself. However, it must be recognised that automatic evaluation of this output is useful for researchers to assess the progress of their own systems on a somewhat superficial basis. Our second option is to make use of an animated avatar tool that will take in the output of the MT engine and reproduce it via a signing animation. While this method allows for an evaluation of actual signed output, it is not without its problems. Ultimately what is being evaluated here is not just the translation but the whole process including the animation figure, how realistic it is and how fluid the signing is, as well as how the animation converts the text to real signing. This is coupled with the fact that manual evaluation by humans is a subjective process, whereby examiners are likely to be swayed by external factors affecting the evaluation such as their experience and opinion of animated signing and avatars, their level of SL and what they are being asked to evaluate against, and if they are in favour of such technology being developed for their use.

In sum, we believe both automatic and manual evaluations have their place in the SL MT process, and regardless of their flaws, both must be performed in order assess the quality of the output as best we can until something more objective is developed.

7 Conclusion

We have presented a set of SMT experiments using two different testsets to compare and assess three different forms of SL representation and how they perform through the MT process. We used traditional MT evaluation metrics to examine the quality of the output of each experiment. We have shown that all methods are MT–compatible, but although SiGML code provides the highest scores, the methods of automatic evaluation do not clearly indicate that is indeed the best system. We also discussed how appropriate current MT evaluation metrics are for assessing these intermediate representations and showed that while they are useful for ascertaining internal progress within experiments, in some instances they are not appropriate, but ultimately should be accompanied by human evaluation of actual signed output for completeness.

In order to better assess and evaluate the transcription methods, in future work we plan to implement a signing version for each of the output sets of data and compare them using manual evaluation.

Acknowledgements

This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (www.cngl.ie) at Dublin City University.

References

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

Hanke, T. (2004). HamNoSys – Representing Sign Language Data in Language Resources and Language Processing Contexts. In *Workshop on the Representation and Processing of Sign Languages at the Languages and Resources Evaluation Conference (LREC 04)*, pages 1–6, Lisbon, Portugal.

Huenerfauth, M. (2004). Spatial and Planning Models of ASL Classifier Predicates for Machine Translation. In *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-04)*, pages 65–74, Baltimore, MD.

Kennaway, J., Glauert, J., and Zwitserlood, I. (2007). Providing signed content on the Internet by synthesized animation. *ACM Transactions on Computer-Human Interaction*, (14):1–29.

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (2007). Moses: open source toolkit for statistical machine translation. In *Proceedings of Demonstration and Poster Sessions at the 45th Annual Meeting of the Association of Computational Linguistics (ACL-07)*, pages 177–180, Prague, Czech Republic.

Koehn, P., Och, F., and Marcu, D. (2003). Statistical Phrase-Based Translation. In *Proceedings of the Combined Human Language Technology Conference Series and the North American Chapter of the Association for Computational Linguistics Conference Series (HLT-NAACL)*, pages 48–54, Edmonton, Canada.

Leeson, L. (2003). Sign Language Interpreters: agents of social shcange in Ireland? In *M. Cronin and C. Ó Cuilleanáin(Eds.), Languages of Ireland*, pages 148–164. Four Courts Press: Dublin, Ireland.

Leeson, L., Saeed, J., Macduff, A., Byrne-Dunne, D., and Leonard, C. (2006). Moving Heads and Moving Hands: Developing a Digital Corpus of Irish Sign Language. In *Proceedings of Information Technology and Telecommunications Conference 2006*, [no page numbers], Carlow, Ireland.

Marshall, I. and Sáfár, E. (2003). A Prototype Text to British Sign Language (BSL) Translation System. In *the 41st Annual Meeting of the Association of Computational Linguistics (ACL-03) Conference*, pages 113–116, Sapporo, Japan.

Morrissey, S. (2008). *Data-driven Machine Translation for Sign Languages*. PhD thesis, Dublin City University, Dublin, Ireland.

Morrissey, S., Somers, H., Smith, R., Gilchrist, S., and Dandapat, S. (2010). Building a Sign Language corpus for use in Machine Translation. In *Proceedings of the 4th Workshop on Representation and Processing of Sign Languages: Corpora for Sign Language Technologies at LREC 2010*, pages 172–177, Valetta, Malta.

Och, F. and Ney, H. (2003). A Systematic Comparision of Various Statistical Alignment Models. *Computational Linguistics*, **29**(1):19–51.

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

Pizzuto, E., Rossini, P., and Russo, T. (2006). Representing signed languages in written form: questions that need to be posed. In *Proceedings of the 2nd Workshop on the Representation and Processing of Sign Languages at the Languages and Resources Evaluation Conference (LREC 2006)*, pages 0–13, Genoa, Italy.

San-Segundo, R., López, V., Martín, R., Sánchez, D., and García, A. (2010). Language resources for spanish - spanish sign language (lse) translation. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies at LREC 2010*, pages 208–211, Valetta, Malta.

Speers (2001). *Representation of American Sign Language for Machine Translation*. PhD thesis, Georgetown University, Washington, DC.

Stein, D., Schmidt, C., and Ney, H. (2010). Sign Language Machine Translation Overkill. In Federico, M., Lane, I., Paul, M., and Yvon, F., editors, *International Workshop on Spoken Language Translation*, pages 337–344, Paris, France.

Stokoe, W. C. (1960). *Sign language structure: an outline of the visual communication system of the American deaf*. Studies in Linguistics, Occasional Paper, 2nd printing 1993: Linstok Press, Burtonsville, MD.

Stroppa, N. and Way, A. (2006). MATREX: DCU Machine Translation System for IWSLT 2006. In *Proceedings of the International Workshop on Spoken Language Translation (IWSLT)*, pages 31–36, Kyoto, Japan.

Sutton, V. (1995). *Lessons in Sign Writing, Textbook and Workbook (Second Edition)*. The Center for Sutton Movement Writing, Inc., La Jolla, CA.

van Zijl, L. and Fourie, J. (2007). Design and Development of a Generic Signing Avatar. In *Proceedings of Graphics and Visualization in Engineering*, pages 95–100, Clearwater, FL.

Veale, T., Conway, A., and Collins, B. (1998). The Challenges of Cross-Modal Translation: English to Sign Language Translation in the Zardoz System. *Machine Translation*, **13**(1):81–106.