Building a Sign Language corpus for use in Machine Translation

Sara Morrissey, Harold Somers, Robert Smith, Shane Gilchrist and Sandipan Dandapat
Centre for Next Generation Localisation
Dublin City University
Glasnevin, Dublin 9, Ireland
{smorri,hsomers,rsmith,sdandapat}@computing.dcu.ie,shane.gilchrist@gmail.com

Abstract

In recent years data-driven methods of machine translation (MT) have overtaken rule-based approaches as the predominant means of automatically translating between languages. A pre-requisite for such an approach is a parallel corpus of the source and target languages. Technological developments in sign language (SL) capturing, analysis and processing tools now mean that SL corpora are becoming increasingly available. With transcription and language analysis tools being mainly designed and used for linguistic purposes, we describe the process of creating a multimedia parallel corpus specifically for the purposes of English to Irish Sign Language (ISL) MT. As part of our larger project on localisation, our research is focussed on developing assistive technology for patients with limited English in the domain of healthcare. Focussing on the first point of contact a patient has with a GP's office, the medical secretary, we sought to develop a corpus from the dialogue between the two parties when scheduling an appointment. Throughout the development process we have created one parallel corpus in six different modalities from this initial dialogue. In this paper we discuss the multi-stage process of the development of this parallel corpus as individual and interdependent entities, both for our own MT purposes and their usefulness in the wider MT and SL research domains.

1 Introduction

This paper describes the planning and construction of a multimedia parallel corpus for the purpose of developing a machine translation (MT)-based approach to using technology to assist patients with limited English in a healthcare scenario. Focussing on the first point of contact a patient has with a GP's office, the medical secretary (receptionist), we are developing a corpus representing the dialogue between the two parties when scheduling an appointment. The corpus is a multimedia six-way parallel corpus consisting of (a) audio recordings of the original material, (b) written English transcription, translated into (c) Irish Sign Language (ISL) video recordings and (d) Bangla text. From the video recordings, transcriptions in (e) HamNoSys and (f) the corresponding SiGML notations have been made, the last of these being suitable to generate ISL with an animated computer figure (avatar). Each of these elements is discussed in this paper.

1.1. Assistive technology and appointment scheduling

There is no shortage of literature confirming that lack of knowledge of the host country's language and the ensuing communication difficulties constitute the single most important barrier to healthcare (e.g Jones & Gill, 1998; and many others), and an equally rich literature, which we will not review here, discusses traditional ways of addressing this problem, through use of interpreters and other services. While this observation usually applies to refugees and other immigrants, it applies equally to Deaf people (e.g. McEwen and Anton-Culver, 1988; and many others). On-going research has been investigating the use of various types of language technology to address this problem for oral languages, including (but not restricted to) MT (Somers and Lovel, 2004; Somers, 2006). In the field of spoken-language MT, cooperative goal-oriented dialogues such as appointment scheduling have always been the most widely targeted dialogue type, while the medical domain has become an important focus of research for speech translation, with its own specialist conferences (e.g. at HLT/NAACL06 in New York, and at Coling 2008 in Manchester).

1.2 SL translation

SL MT is in the early stages of development, in comparison with mainstream MT. Widespread documented research in SL MT did not emerge until the early 1990s. This is understandable given the comparatively late linguistic analysis of SLs (Stokoe, 1960). Despite this, and within the short time-frame of research, the development of systems has roughly followed that of spoken language MT from rule-based approaches toward data-driven approaches.

Rule-based systems, such as the Zardoz system (Veale et al., 1998) and the ViSiCAST project (Marshall and Sáfár, 2002, 2003) carry out a deep linguistic analysis on a syntactic and sometimes semantic level in order to define rules for translation. More recent systems developed at RWTH Aachen University (Dreuw et al., 2007) and Dublin City University (Morrissey, 2008) have employed data-driven approaches that eschew heavy linguistic analysis in favour of empirical and statistical data. Both methodologies are heavily dependent on the suitability of the transcription approach chosen.

In the remainder of this paper we discuss our methods and the issues and problems in each stage of the corpus building activity, ending with a preview of our intended uses of the corpus.

2 Elicitation method

Our first task was to collect an English-language corpus of patient–receptionist dialogues. A major difficulty in gathering genuine data in the medical field, or any domain where personal information is involved, is that the confidentiality and other ethical issues more or less preclude using genuine data collected *in situ*. This difficulty has long been recognised in medical training, where "standardized patients" (SPs) are used with medical students, that is, actors trained to simulate consistently the responses of a patient in a particular medical setting. Barrows (1993) describes some of the pros and cons of using SPs. As far as we could ascertain, no reported study has used SPs only for appointment scheduling, though this activity has been a (usually minor) part of many studies. Training SPs is of course a major undertaking in itself necessarily involving experienced experts, so for the purposes of this project we made a compromise in that we engaged an experienced GP's receptionist to participate in a number of role-play sessions with the native English speakers among the authors (HS, SM, RS). These were all recorded and later transcribed. Following the receptionist's guidance, we role-played a number of scenarios:

- general appointment scheduling with the GP or practice nurse, including scheduling on behalf of a third party (a child, an old person, or someone who doesn't speak English),
- emergency situations
- scheduling of specific activities, e.g. vaccinations, bringing in samples, collecting results, having stitches removed, etc.
- changing or cancelling appointments

Many of the dialogues involved negotiations of a general nature (e.g. exploring available days and times) or more specific to the individual person or purpose. In each case, the receptionist made suggestions based on her real-life experience of types of interactions that had not already been covered. In this way, we believe that our corpus contains samples that are realistic, and offer a broad coverage of our target domain, even if they are not genuine in the literal sense.

Our recordings comprise 350 dialogue turns. In transcription, this works out at just under 3,000 words (a very small corpus by any standards), each dialogue turn on average roughly 8 words.

3 Translation

The next stage in the process was to translate our English corpus into ISL (and Bangla). ISL is the main SL used in Ireland's Deaf community. Historically, Deaf children were taught separately according to their sex, leading to the rise of two main variants in ISL, i.e. male signs and female signs. Among the younger generation, there has been an acceleration in contact between varieties due to increasing social interactions between males and females, and thus contemporary ISL could be said to include both dialects. Older members of the community may not be familiar with variants from the other side.

Signed English (SE), promoted by a Deaf school in Dublin, is used by a number of Deaf people in the greater Dublin area, especially among the older generation. It is seen by some as prestigious, despite the more recent view that ISL is the way forward. There is a strong link between SE and the Church: for example the Lord's Prayer and Hail Mary are done in SE rather than ISL.

For the present project, a Deaf consultant was engaged to discuss the most suitable strategy. It was agreed that Deaf people who use SE are capable of following ISL no matter how fluent it is. On the other hand, native signers of ISL would have trouble following SE. It can be argued that SE is part of ISL (just as finger spelling is). In this context, when discussing ISL, we are talking about a register where there is very little influence from English and this in turn provides a challenge for translating since ISL is a minority language used in face-to-face communication while English is used when writing and reading. However , low levels of English literacy among Deaf people is a major motivation for this project, so it was agreed that our translations into ISL should show a minimal influence from English.

3.1 Challenges in translation

Translation between any languages, whether related or not, involves cases where closely following the source text (a "literal" translation, within the grammatical constraints of the target language) can result in a stilted, unnatural or, in the worst case, unacceptable translation. This is especially the case when translating between English and ISL which differ both typologically and (obviously) in the medium of expression.

A particular difference is the role of pragmatics in the two languages. ISL utterances tend to reflect the immediate context much more explicitly than English, so that it is difficult to provide an ISL translation of a given dialogue turn out of context. This also has serious implications for our approach to MT.

A good example is the dialogue in (1):

(1) A. Which doctor would you prefer?
B. I don't mind.

In ISL, A will depend on how many choices there are: if there are three people, they will first have to be identified, using the neutral space to show three different placements. Then <WHICH?> is signed,[1] spreading it across the neutral space. For the response B, the signer would just point at each placement then sign <EITHER>, then <DON'T MIND>. But just signing <DON'T MIND> without the context would be misleading or meaningless.

Interestingly, this exchange posed a similar problem for translation into Bangla where a literal translation (2a) is less preferable than a more explicit translation (2b).

(2) a. আমি কিছু মনে করব না।
āmi kichhu mane karaba nā
I don't mind.
b. যে কোনো একজনকে দেখালেই হবে।
ye kono ekajanake dekhālei habe
Can see either of them.

Open-ended questions in English are better translated into ISL with a range of possible answers. For example, we translated (3a) as (3b).

(3) a. How long will it take you to get here?
b. YOU-GET-HERE WHAT TIME? 10 MINUTES? 5 MINUTES?.

The strategy of "explicitation" is well known in translation studies (Klaudy, 1998). There are many examples of this in our corpus: for many conditions the sign includes location on the body, for example <PAIN> or <RASH>, the sign for which should indicate whether the condition is on the arm, on the back, on the face etc. One tactic, though against our principle of providing natural translations, is to fingerspell <R-A-S-H>.

4 Video recording

Although a number of SL video corpora have been collected, there are no agreed standard formats, often

[1] Our convention in this paper is to indicate signs with an approximate English gloss in small capitals.

because of differences in the underlying purpose behind the corpora.

The first batch of signing was recorded using an analogue TV camera at the DCU TV studio using miniDV tapes. Upon advice from technical staff at DCU School of Communication, for the remainder a Sony XCAM HDD digital camera was used. This resulted in a big jump in quality and ease of editing. The first batch was transferred to file using the DV deck which was highly time consuming and the quality was not good. The second batch showed a vast improvement in comparison.

Following the lead of the Signs of Ireland corpus project (Leeson and Nolan, 2008), the individual recordings were stored as .MOV files. They were edited using the Final Cut Pro video editing program on a Apple iMac G5 at the DCU School of Communication

Three days were spent translating the English sentences into ISL: often some trial and error was needed to arrive at a translation that was satisfactory.

After the initial recording session, our Deaf consultant reviewed the translations. Approximately 90 of the 350 sentences had to be redone for several reasons because they were felt to be too close to the English, because facial expressions were not appropriate, placement and neutral space not used correctly, and other performance frailties due to the signer's fatigue towards the end.

In retrospect, it probably would have saved effort if the reviewer had been present during the original recordings. Despite the budgetary implications, this would have saved time and energy, and would have improved the overall quality of the corpus.

This highlights one of the most interesting differences between translation into SLs and oral languages: because of the "performance" element of the SL, the step equivalent to revision in the (oral language) translation flow is considerably more demanding.

5 Transcription

The next stage was to transcribe the videos into a form suitable for textual manipulation. It is probably not necessary in the present forum to justify our use of a transcription that reflects the actual signs in a more explicit way than the widely used convention of glossing into quasi-English, even if that representation method is advantageous for ready reference, as in our discussion in the previous section.

Our choice here was guided by our main purpose, ultimately, to use the corpus of translations in a data-driven MT system to generate translations of (novel) English inputs as simulations of ISL using a computer graphic animated character (avatar).

After looking at several alternatives, it was decided to use the Hamburg Notation System (HamNoSys) and its related mark-up language SiGML.

5.1 HamNoSys

HamNoSys is a well-established transcription system developed by the Institute for German Sign Language and Deaf Communication at the University of Hamburg for all SLs (Prillwitz et al., 1989). HamNoSys is a phonetic notation system purpose-built for use by linguists in their detailed analytical representation of signs and sign phrases rather than as a writing system for SLs. According to Bentele (n.d.), it consists of about 200 symbols covering the parameters of hand shape, hand configuration, location and movement. The symbols are iconic so as to be more easily recognizable and learnable. The order of the symbols within a string is somewhat fixed, but it is still possible to transcribe a given sign in lots of different ways. The notation is essentially phonemic, so the transcriptions are very precise, but on the other hand also very long and cumbersome to decipher. Without doubt, the learning curve for a newcomer to HamNoSys is relatively steep.

Transcribing HamNoSys is all the more arduous because the most widely used annotation tool, ELAN,[2] does not handle HamNoSys. To our knowledge, the only transcription software available for HamNoSys that allows alignment with the video timestamp is iLex (Hanke, 2002), though we have not yet got access to this tool.

5.2 SiGML

Closely associated with HamNoSys is SiGML (Signing Gesture Mark-up Language) (Elliott et al., 2004), a form of XML which defines a set of XML tags for each phonetic symbol in HamNoSys. SiGML files are represented as plain text which means they can be easily handled by computer, e.g. for transmission, and by the MT system (see below). SiGML was developed by the Virtual Humans group at the University of East Anglia over a three year period to support the work of the EU-funded projects ViSiCAST (Elliott et al. 2000; Kennaway, 2001, 2003) and eSIGN (Kennaway et al., 2007), whose main focus was to provide communication tools in the form of computer-graphic animated figures (avatars) for members of the Deaf community.

The SiGML representation of the HamNoSys notation of the SL sequence is readable by the AnimGen 3D rendering software (Kennaway, 2003).

6 Avatars

Research into synthesising SLs is still in the early stages of development. Most existing systems use avatars to synthesise sign language in real-time (e.g. Grieve-Smith, 1999; Krňoul et al. 2007). Using a tool called eSIGNeditor (Kennaway et al., 2007) developed during the eSIGN project, we are able to compose HamNoSys scripts for the corpus and validate them in real-time by using the processing pipeline for synthetic SL generation also developed in the eSigns project. Using this system, it is not possible however to align the HamNoSys transcriptions to the time stamps on the video files as it would be with iLex.

State-of-the-art SL synthesis can be compared to the somewhat robotic and artificial nature of early speech synthesis output. Current problems with the avatar include the need for better collision detection, more naturalness and less jerkiness. Collision detection is a means to incorporate awareness of the physical space taken up by the human body. Getting the avatar to position its hands exactly where you want them, for example close to the face, requires quite subtle programming: by default the hands and arms will take the shortest route possible to their destination, sometimes passing through another part of the body. There is a trade-off between collision

[2] http://www.lat-mpi.eu/tools/elan/, accessed 20.3.10

detection and processing time, but this should be a matter for the underlying software rather than the SiGML transcription. Similarly, some improvements will be necessary to prevent the avatar from doing impossible things, such as turning or bending limbs and joints in an unnatural fashion. And in some cases, the avatar's movements are still sometimes jerky and robotic. As part of our project we hope to address key factors that would make the animations more natural and human, in collaboration with colleagues at UEA. In addition to the above issues, we wish to address three further factors:

- non-manual features (facial expressions, mouth movements)
- non-linguistic attributes of the avatar such as weight shift, involuntary movements
- natural variance in signs, such as lack of symmetry in two-handed signs.

These developments should deliver a more human-like avatar, thereby improving SL synthesis quality and increasing acceptability by the target audience.

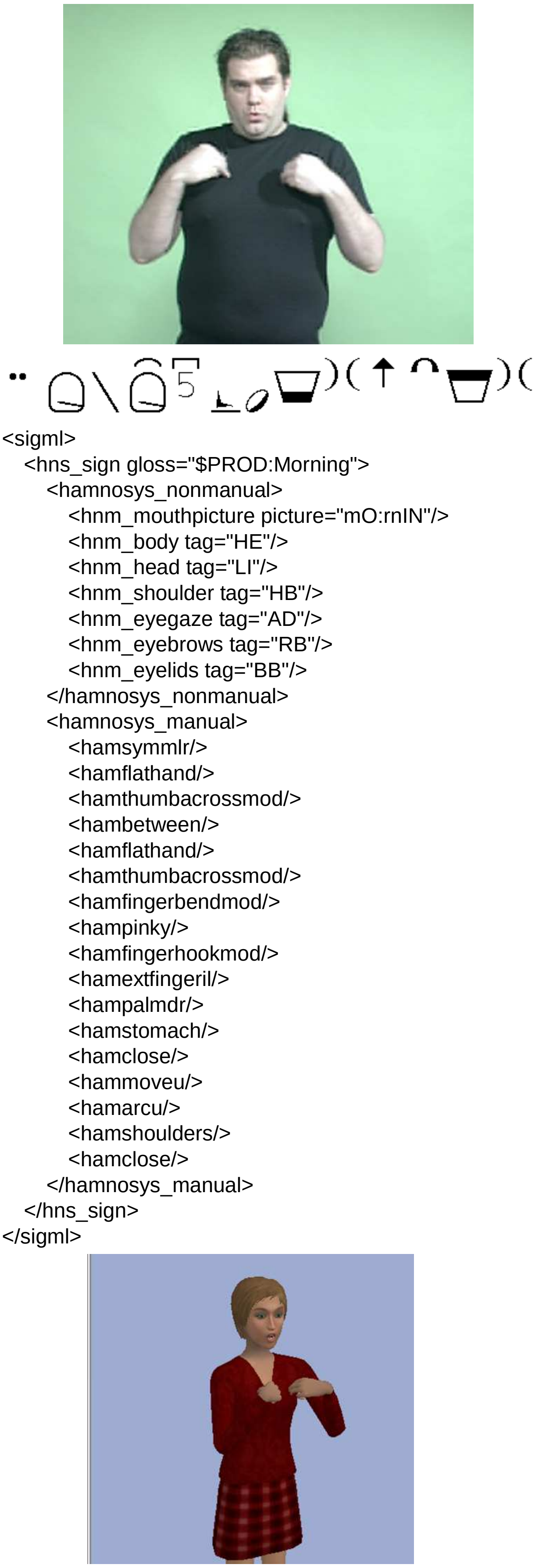
Figure 1 illustrates all the steps in the process for the word *morning* (found in several of our dialogue turns): a screen shot from the video corpus, transcribed into HamNoSys, the corresponding SiGML, and as synthesised by the avatar.

7 Proposed use for MT

Situated in a large and successful data-driven MT research group, we will adapt and use our MaTrEx MT system (Du et al., 2009; Ma et al., 2009) for the task of English to ISL translation. This system employs statistical- and example-based methods to perform translation. Statistical MT (SMT) is largely dependent on there being a large parallel corpus for training the system. Frequently, such systems train on several million sentence pairs (Du et al., 2009). Developmental constraints in our work have allowed us to create a toy corpus of only approximately 350 utterances. For this reason we will explore example-based methods which translate by analogy (Somers et al., 2009) and do not require the large amounts of data statistical models do.

Example-based machine translation (EBMT) is sometimes seen as an extension of the well-known translator's tool, the Translation Memory (although historically the two ideas were developed somewhat independently, and at about the same time – see Somers and Fernandez Diaz, 2004). In both, the input to be translated is compared with a database of previously done translations. If a direct match is found, the corresponding translation is used. If an imperfect match is found, it is then used as a model on which to base construction of the new translation. In the Translation Memory scenario, the translator takes the lead, while in EBMT this is done automatically, usually with the help of further examples that "cover" the differences. The reusable fragments in the source sentence and the found example(s) are extracted, aligned with the corresponding fragments in the translation, and then recombined to form the new sentence.

The English and SiGML modalities in our corpus will be used to drive this EBMT process. The marked-up text will be processed in the same way as MT data used in local-



```
<sigml>
   <hns_sign gloss="$PROD:Morning">
      <hamnosys_nonmanual>
         <hnm_mouthpicture picture="mO:rnIN"/>
         <hnm_body tag="HE"/>
         <hnm_head tag="LI"/>
         <hnm_shoulder tag="HB"/>
         <hnm_eyegaze tag="AD"/>
         <hnm_eyebrows tag="RB"/>
         <hnm_eyelids tag="BB"/>
      </hamnosys_nonmanual>
      <hamnosys_manual>
         <hamsymmlr/>
         <hamflathand/>
         <hamthumbacrossmod/>
         <hambetween/>
         <hamflathand/>
         <hamthumbacrossmod/>
         <hamfingerbendmod/>
         <hampinky/>
         <hamfingerhookmod/>
         <hamextfingeril/>
         <hampalmdr/>
         <hamstomach/>
         <hamclose/>
         <hammoveu/>
         <hamarcu/>
         <hamshoulders/>
         <hamclose/>
      </hamnosys_manual>
   </hns_sign>
</sigml>
```

Figure 1. Screen shot, HamNoSys, SiGML and avatar signing the word *morning*.

isation workflows (Du et al., 2010). Either the HamNoSys transcription or the SiGML code could form the text-based version of ISL required for MT processing. Both provide a level of granularity much finer than the usual approach to EBMT, which is usually based mainly on word-based matches, rarely on letter strings. It will be interesting, and a matter of research, to see the effect this has on the alignment and recombination phases of EBMT. For example subtle differences between signs that give different nuances of meaning and expression, for example in hand position, movement, or shape, will be captured by the system and used in the translation.

Using SiGML allows us to maintain the phonetic description of the signs required for animation by the avatar and avoids the use of glossing and other techniques that can misrepresent the language.

While current research efforts are focussed on English-to-ISL MT, we hope to expand the system in the future to include recognition components to allow for ISL-to-English MT, and thus a complete bidirectional translation system.

Acknowledgements

This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (http://www.cngl.ie) at Dublin City University. We would also like to acknowledge the contributions to this work of Eoin Campbell and Damien Hickey of the DCU School of Communication, Suzanne Lindfield, and Alvean Jones, and the assistance of our colleagues at UEA, Norwich.

References

Barrows, H.S. (1993) An overview of the uses of standardized patients for teaching and evaluating clinical skills. *Academic Medicine* 68, pp. 443–451.

Bentele, S. (n.d.) About the HamNoSys system. http://www.signwriting.org/forums/linguistics/ling007.html, accessed 20.3.10.

Dreuw, P., D. Stein and H. Ney. (2007) Enhancing a Sign Language translation system with vision-based features. In M. Sales Dias, S. Gibet, M.M. Wanderley and R. Bastos (eds) *Gesture-Based Human-Computer Interaction and Simulation, 7th International Gesture Workshop, GW 2007, Lisbon, Portugal, Revised Selected Papers* (LNAI 5085), Berlin (2009): Springer, pp. 108–113.

Du, J., He, Y., Penkale, S. and Way, A. (2009) MaTrEx: the DCU MT System for WMT 2009 . In *EACL 2009 Fourth Workshop on Statistical Machine Translation*, Athens, pp 95–99.

Du, J., Roturier, J. and Way, A. (2010) TMX markup: A challenge when adapting SMT to a localisation environment. Paper submitted to 14th Annual Conference of the European Association for Machine Translation, Saint-Raphaël, France.

Elliott, R., Glauert, J.R.W., Jennings, V. and Kennaway, J.R. (2004) An overview of the SiGML notation and SiGMLSigning software system. In *Fourth International Conference on Language Resources and Evaluation, LREC 2004*, Lisbon, pp. 98–104.

Elliott, R., Glauert, J.R.W., Kennaway, J.R. and Marshall, I. (2000) The development of language processing support for the ViSiCAST project. In *ACM SIGACCESS Conference on Computers and Accessibility, Proceedings of the fourth international ACM conference on assistive technologies*, Arlington, Virginia, pp. 101–108.

Hanke, T. (2002) iLex – A tool for sign language lexicography and corpus analysis. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, Las Palmas de Gran Canaria, Spain, pp. 923–926

Jones, D. and Gill, P. (1998). Breaking down language barriers. *British Medical Journal* 316 (7127), pp. 1476–1480.

Kennaway, R. (2001) Synthetic animation of deaf signing gestures. In I. Wachsmuth and T. Sowa (eds) *Gesture and Sign Language in Human-Computer Interaction, International Gesture Workshop, GW 2001, London, UK, 2001, Revised papers*, (LNCS 2298), Berlin (2002): Springer, pp.149–174.

Kennaway, R. (2003) Experience with and requirements for a gesture description language for synthetic animation. In A. Camurri and G. Volpe (eds) *Gesture-Based Communication in Human-Computer Interaction, 5th International Gesture Workshop, GW 2003, Genova, Italy, 2003, Selected revised papers*, (LNAI 2915), Berlin (2004): Springer, pp. 300–311.

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

Klaudy, K. (1998) Explicitation. In M. Baker (ed.) *Encyclopedia of Translation Studies*, London: Routledge, pp. 80–85.

Krňoul, Z., Kanis, J., Železný, M. and Müller, L. (2007) Czech text-to-sign speech synthesizer. In A. Popescu-Belis, S. Renals and H. Bourlard (eds) *Machine Learning for Multimodal Interaction, 4th International Workshop, MLMI 2007, Brno, Czech Republic, 2007, Revised selected papers* (LNCS 4892), Berlin (2008): Springer, pp. 180–191.

Leeson, L. and Nolan, B. (2008) Digital Deployment of the Signs of Ireland Corpus in Elearning. In *3rd Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, Marrakech, Morocco, pp. 112–121.

Ma, Y., Okita, T., Çetinoğlu, Ö, Du, J. and Way, A. (2009) Low-resource Machine Translation using MaTrEx: the DCU Machine Translation System for IWSLT 2009. In *Proceedings of the International Workshop on Spoken Language Translation (IWSLT 2009)*, Tokyo, pp.29–36.

Marshall, I. and Sáfár, E. (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.

Marshall, I. and Sáfár, E. (2003) A prototype text to British Sign Language (BSL) translation system. In *41st Annual Meeting of the Association of Computational Linguistics*, Sapporo, Japan, pp. 113–116.

McEwen, E. and Anton-Culver, H. (1988) The medical communication of deaf patients. *Journal of Family Practice* 26, pp. 289–291.

Morrissey, S. (2008). Data-driven machine translation for sign languages. PhD Thesis, Dublin City University, Dublin.

Prillwitz, S., Leven, R., Zienert, H., Hanke, T. and Henning, J. (1989) *HamNoSys. Version 2.0; Hamburg Notation System for Sign Languages. An introductory guide.* Hamburg: Signum.

Somers, H. (2006) Language engineering and the pathway to healthcare: A user-oriented view. In *HLT/NAACL-06 Medical Speech Translation, Proceedings of the Workshop*, New York, NY, pp. 32–39.

Somers, H., Dandapat, S. and Naskar, S. (2009) A review of EBMT using proportional analogies. In *Proceedings of 3rd Workshop on Example-Based Machine Translation*, Dublin, pp. 53–60.

Somers, H. and Fernandez Diaz, G. (2004) Translation Memory vs. Example-based MT: What is the difference? *International Journal of Translation* 16 (2), pp. 5–33.

Somers, H. and Lovel, H. (2003) Computer-based support for patients with limited English. In *Association for Computational Linguistics EACL 2003, 10th Conference of the European Chapter, Proceedings of the 7th International EAMT Workshop on MT and other language technology tools: Improving MT through other language tools, Resources and tools for building MT*, Budapest, pp. 41–49.

Stokoe, W. C. (1960). *Sign Language Structure: An outline of the visual communication system of the American deaf*, 2nd printing, Burtonsville, MD (1993): Linstok Press.

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, pp. 81–106.