Mach Translat (2013) 27:25–64
DOI 10.1007/s10590-012-9133-1

Manual labour: tackling machine translation for sign languages

Sara Morrissey · Andy Way

Received: 10 December 2008 / Accepted: 16 October 2012 / Published online: 14 March 2013

Abstract This article explores the application of data-driven machine translation (MT) to sign languages (SLs). The provision of an SL MT system can facilitate communication between Deaf and hearing people by translating information into the native and preferred language of the individual. In this paper we address data-driven SL MT predominantly for Irish SL (ISL) but also for German SL (DGS/Deutsche Gebärdensprache). We take two different purpose-built corpora to feed our MaTrEx MT system and in a set of experiments translating both to and from the SLs, we investigate the effects of SL data on statistical MT (SMT). Exploiting the bidirectionality of the MaTrEx system, we demonstrate how additional modules, such as recognition and SL animation, can potentially build a full SL MT model for spoken and SL communication in addition to promising evaluation scores. A secondary focus of the article is on the two main issues affecting SL MT, those of transcription and evaluation. We offer a discussion on both these common problems before concluding.

Keywords Machine translation · Sign languages · Manual communication · Annotation · Data-driven machine translation · Statistical machine translation · Transcription · Evaluation

The work described in this article was, for the most part, carried out while at Dublin City University.

S. Morrissey
Centre for Next Generation Localisation, School of Computing, Dublin City University, Dublin, Ireland
e-mail: smorri@computing.dcu.ie

A. Way (✉)
Lingo24, Greenfield, Greater Manchester, UK
e-mail: andy.way@lingo24.com

1 Introduction

Communication is problematic when languages differ, but this is exacerbated still further between languages of different modalities, namely spoken and sign languages (SLs).

While simple spoken language communication problems can often be partly resolved by a combination of basic foreign language knowledge, similar sounding words, and gesticulations, the cross-modal nature of SL-spoken language communication poses additional challenges. Not only are the words and structure different, but so is the mode of communication. Instead of oral/aural–oral/aural, we are faced with oral/aural–gestural/visual. Furthermore SLs are not wholly iconic in that the signs do not always visually represent what is being signed and many signs represent abstract concepts.

To assume that this barrier can be overcome by lip-reading and speech on the part of the Deaf[1] person, and speaking and watching a signer mime on the part of a hearing person, is to be ignorant of the rich, complex and fully expressive language that is an SL. Lip-reading assumes a good grasp of the spoken language by the Deaf person and a slow, clear articulation by the speaker. In addition, SLs are full and articulate languages, far more complex and independent from spoken languages than mime. Furthermore, person-to-person communication within the confines of language barriers usually requires one or other party to break from using their native language, something which may not be possible for either party in the context of Deaf–hearing communication.

While direct cross-modal person-to-person communication can cause problems, the Deaf community can also have issues with indirect communication, namely written language. Studies have shown that the average Deaf adult has the literacy competencies of a 10-year-old (Traxler 2000). In addition, a study of the educational background and employment status of Deaf adults in Ireland showed that 38 % of the study participants did not feel fully confident reading a newspaper, and more than half were not fully confident writing a letter or filling out a form (Conroy 2006, p. 11). Regardless of the literacy competencies of a Deaf person, it should be their right to have information available to them in SL, their first and preferred language, and a language that is most natural to them. There are approximately 7 million D/deaf people worldwide, and in Ireland alone, there is a total population of approximately 50,000 people who know and use ISL (Leeson 2001), with over 5,000 members of the Deaf Community (Ó'Baoill and Matthews 2000, p. 7), yet "there are no public services available in this language and provision of services in an accessible language—in all domains of life—is relatively ad hoc" (Leeson 2003, p.150). This may be due in part to the lack

[1] It is generally accepted (Leeson 2003) 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. The boundaries of these categories are fuzzy and people may consider themselves on the border of one or the other depending on their experiences and preferences. In this paper, we use 'Deaf' to describe people whose preference is for communicating through SLs, and 'hearing' (note lowercase 'h') to describe all members of the non-D/deaf communities.

of official recognition of ISL as one of the indigenous languages of Ireland, a privation that is shared by most countries, although some have made the move to official recognition, such as Sweden, and the USA under the Equality Disability Act (Ó'Baoill and Matthews 2000).

One way of breaking through the communication barrier is through SL interpreters (SLIs). SLIs provide an invaluable intermediary communication service for Deaf people, but low interpreter availability (1:250 for Deaf people in Ireland) and high cost can make the service prohibitive. Matters of confidentiality can also affect the use of SLIs and one can imagine scenarios where interpretation or translation is required for only a short time, e.g. reading websites and documents or short daily interactions in public such as shopping or booking appointments.

Clearly, some means of interpretation or translation is required where the privacy of the individual is protected, which does not require the pre-booking of an SLI for short interactions and which can allow Deaf people access to information and services immediately and in their own language. Advancements in technology have provided some answers to this problem. Teletype systems are available for telephone conversations, and closed-captioning and subtitles are available on all DVDs nowadays. While these tools do alleviate a Deaf person's problems with hearing, they assume good literacy skills and speed of reading and understanding (Huenerfauth 2006). Often, for reasons of space and timing, subtitles are simplified, which means that some of the important information can be missing. In addition, this simplification can be seen as insulting to the Deaf communities and may be considered a dumbing down of the information. More recently, the release of Smart phones with online video communication technology such as the iPhone 4 allow Deaf people to communicate face to face with each other, albeit with the added cost of data transfer for such services. Aside from the often prohibitive cost factor[2] this service allows Deaf-Deaf communication but does not offer a solution to the communication barrier between Deaf and hearing.

There is, however, one particular type of technology that can tackle the communication and comprehension problems of the Deaf community, namely machine translation (MT). Over the last 60 years or so, since the first proposal on using computational methods to automate translation (Weaver 1949), advances in MT research have shown to significantly bridge communication and understanding between different spoken languages. This is demonstrable through the proliferation of online MT systems such as Google Translate, of large-scale research centres being funded for MT research and development (such as the Centre for Next Generation Localisation at Dublin City University, the University of Edinburgh, and RWTH Aachen) and of international MT competitions (such as the International Workshop on Spoken Language Translation[3] and the Workshop on Machine Translation[4]).

[2] According to a study by Conroy (2006), Deaf people in Ireland suffer from an above expected rate of unemployment and 70 % of those who are in employment are earning low wages.

[3] http://iwslt2011.org/.

[4] http://www.statmt.org/wmt11/.

The prosperity of this technology, coupled with the need for practical, objective tools for translation in the Deaf community, has motivated the research described in this article, the remainder of which is organised as follows.

In Sect. 2, we outline SLs, including an overview of SL linguistics that can affect SL MT as well as various text-based representations for SLs. We provide an overview of the literature in the field of SL MT in Sect. 3, highlighting the different approaches that have been taken over the years to address different aspects of SLs and their translation. In Sect. 4 we describe the MT system we have used in our own experiments and go on to detail a wide variety of experiments using two different corpora in Sect. 5. The main issues facing SL MT, namely transcription and evaluation, are outlined in Sect. 6, before we conclude in Sect. 7.

2 Sign languages

SLs are the first languages of members of the Deaf communities worldwide (Ó'Baoill and Matthews 2000). As in naturally occurring and indigenous languages, they can be as eloquent and as powerful as any spoken language through their visual–spatial modality. It is this alternative communicative channel that poses interesting challenges for the area of MT that lie outside of the general spoken language MT field.

SL research is still a relatively new area when compared to research into spoken languages. Significant study into the linguistic structure of SLs only began in the 1960s with the seminal work of Stokoe (1960), whose research on American Sign Language (ASL) paved the way for social recognition of SLs as real languages. More recent acknowledgment of this is included in the works of linguists such as Pinker (1994, p. 36) and Chomsky (2000, pp. 100–101). Increased recognition of SLs as fully formed, independent languages following political acknowledgment, such as the Resolution of the European Parliament in 1988,[5] has led to some level of research being carried out on SLs in most countries.

There are many common misconceptions surrounding SLs. One is that there is only one SL. On the contrary, SLs are not universal (Leeson 2003). The Ethnologue of world languages[6] catalogues 130 SLs in their own right that have evolved naturally within Deaf communities or have been borrowed from other SLs. International SL does exist[7] but is rather a vocabulary of signs, iconic in nature used to facilitate communication in international contexts and without any native users.

The second common misconception is that SLs are mimed or gesture-based versions of spoken languages. While Signed Exact English (SEE) does exist where there is a sign for each English word (and often each English morpheme), this is not true native signing. Real SLs are independent of spoken languages and make the best use of the modality of communication. SEE does not allow the same ease of communication through the spatial channel as real signing does.

[5] http://www.policy.hu/flora/ressign2.htm.

[6] http://www.ethnologue.com.

[7] http://listserv.linguistlist.org/cgi-bin/wa?A2=ind0202&L=slling-l&P=8; http://www.deaflibrary.org/asl.html.

2.1 Sign language linguistics

While SLs are natural languages in their own right, there are some distinguishing characteristics that differentiate SLs from spoken languages and which pose interesting challenges for MT. These include:

- Articulation: compared to spoken languages where the primary articulators are the throat, nose and mouth, the main articulators in SLs are the fingers, arms and hands, in addition to body and facial movement. Signs are analogous to morphemes and the articulations of the hands and body can be categorised as phonemes (Stokoe 1972).
- Non-Manual Features: (NMFs) are predominantly concomitant with manual signs, and consist of movements or expressions of parts of the body other than the hands that can express emotion, intensity or act as morphological and syntactic markers.
- Classifiers: these are grouping terms to denote the shape or arrangement or consistency of objects. They are generally preceded by a citation form of the lexical item, which is followed by the relevant classifier which can demonstrate the action or location of the cited object.
- Signing Space: SLs are articulated within the space in front of the signer, which extends from just above the head down to the waist and outwards about as far as the arms extend. SLs are articulated in such a way as to make the best use of the space in which articulation can take place. This includes using locational points in the space as anaphoric reference points.

2.2 Text-based representation

One of the striking differences between sign and spoken languages is the distinct lack of a formally adopted, or even recognised, writing system for SLs, meaning that SLs remain as visual–spatial languages that cannot be 'read' as spoken languages can. There have been many attempts at creating writing systems for SLs, but most are not usable by the general public as they consist of numeric codes or symbols to encapsulate the phonetics or phonology of signs and are not easily learned or written, nor is there a standardised accepted form.

However, given that data-driven MT, the approach described in this article, requires a text-based format, we look here at some of the most common representation formats for SLs.[8]

2.2.1 Symbolic

Stokoe Notation (Stokoe 1960) and the Hamburg Notation System (HamNoSys) (Prillwitz et al. 1989; Hanke 2004) use sets of predefined arbitrary symbols to represent

[8] This intermediate representation of SLs may be likened to the interlingua of a rule-based MT approach. While MT requires some sort of representation for SLs, it is not specifically intended as an interlingua, but rather as a means to describe the SL in question. Regardless of the transcription method used for SLs, the representation usually strives to be dependent on the source language, which somewhat separates the annotation from the language-dependent classification of a standard interlingual structure.

Fig. 1 Stokoe Notation: *don't know*

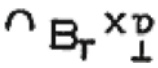
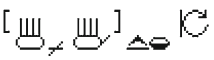


Fig. 2 HamNoSys Notation: *nineteen*



SLs. Stokoe Notation was developed in the 1960s for ASL and describes four factors to be taken into account, each with their own set of symbols: *tabulation* referring to the location of a sign; *designator*, referring to the hand shape; *signation*, referring to the type of movement articulated; and *orientation* describing the orientation of the hand shape. This system is primarily useful for notating single, decontextualised signs.

Figure 1 shows the Stokoe notation for the sign *don't know* on the left. This example shows the use of tabulation, signation, designator and orientation symbols for a single sign. The first symbol, like an upside-down *u*, is a tabulation symbol referring to the forehead, brow or upper face. The letter *B* that follows it is a designator symbol indicating a *B* hand shape where all fingers are extended and side-by-side but with the thumb folded in. The subscripted capital T symbol beside it refers to the orientation of the designator, in this case facing the signer. The superscripted x symbol indicates a signation where the designator hand shape *B* moves to touch the tabulation location of the upper head. The next superscript symbol, similar to an inverted, reversed lower case *a*, is also a signation symbol indicating the palm is turned down. The final symbol, a subscripted, upside-down *T*, indicates that the hand now faces away from the signer.

HamNoSys uses a set of language-independent symbols to iconically represent the phonological features of SLs (Ó'Baoill and Matthews 2000) that was not developed for computational usage. This system, rooted in the Stokoe system, allows even more detail to be described, but in most cases it is only a description of the hand shape; further NMF details have been incorporated in certain instances (Hanke 2002).

HamNoSys code for the sign *nineteen* is shown in Fig. 2. It illustrates most of the description categories used in HamNoSys notation. The first colon-like symbol indicates that the hand arrangement is a two-handed sign where the hands are parallel but not mirroring each other. The square brackets denote a grouping of simultaneously occurring movements. Within these braces, the oval symbol with four vertical lines indicates that one hand has four fingers extended and aligned together, while the same symbol repeated with a fifth diagonal line denotes the same hand shape but with the thumb extended this time. The horizontal line with the dash through it in between these two hand shapes shows the movement is repeated from the starting point of the original movement. Outside the brackets, the next symbol, a horizontal line with an upward facing arrow above it, indicates that the hand orientation is away from the body, while the next symbol, an oval with the lower half shaded, signifies that the palm of the hand is facing down. The final symbol, a vertical line with a clockwise circular arrow to its right, shows that the movement of the hands is circular, moving vertically and away from the body.

Symbolic systems can be a laborious means of transcribing SL data, and while descriptive, their meaning is not immediately salient without some form of decoding. While both systems can be used to manually transcribe SL videos, there are software

Fig. 3 Example of SignWriting: *deaf*

tools available that facilitate this process such as eSignEditor (Kennaway et al. 2007) and iLex (Hanke and Storz 2008). eSignEditor is used in the work described later in this article and provides a set of HamNoSys symbols from which the transcriber can choose those that best describe the signs they are annotating. The tool also provides an animation facility where the HamNoSys symbols are converted to SL via an animated signing avatar.

2.2.2 Pictorial

An alternative, yet similar approach, SignWriting, was developed in Sutton (1995). This approach also describes SLs phonologically, but unlike the others, it was developed as a handwriting system. Symbols that visually depict the articulators and their movements are used in this system, where NMFs articulated by the face (pursed lips, for example) are shown using a linear drawing of a face. These simple line drawings, such as that shown in Fig. 3, make the system easier to learn as they are more intuitively and visually connected to the signs themselves.

This example of SignWriting shows a circle that denotes the head of the signer with two sets of two parallel opposing diagonally slanting lines indicating eyes and a three further lines joined to form a smile. The square to the right of the head with a line extending from it indicates a hand with the index finger pointing out, and the semi-shaded square denotes that the back of the hand is facing outward. The asterisk symbols above and below the hand symbol and against the top right and lower right of the face indicate that the index finger touches these points.

The linguistic detail of SLs is encoded in the line drawings of SignWriting. Essentially, a secondary 'language' such as ASCII code representation of the SignWriting pictures or some form of recogniser would be needed to transfer the necessary detail from SignWriting text into a representation suitable for use in SL MT.

2.2.3 Annotation

This approach involves transcribing information taken from a video of signed data. Spoken language stems are normally used to gloss what is being signed.[9] It is a subjective process where a transcriber decides the level of detail at which the SL in the video will be described. These categories can include, for example, a gloss term of the sign being articulated by the right and left hands (e.g. FLIGHT if the current sign being articulated is the sign for the word 'flight'), as well as information on the corresponding NMFs, whether there is repetition of the sign and its location or any other relevant linguistic information (cf. SL linguistics in the Sect. 2.1). The annotations

[9] See the ECHO SL corpus project: http://www.let.kun.nl/sign-lang/echo/.

may be time-aligned according to their articulation in the corresponding SL video. The EUDICO Linguistic Annotator (ELAN),[10] is a gesture researching tool used for annotation. This tool is the standard for creating and developing SL corpora, and was specifically designed for language and gesture analysis (Leeson et al. 2006).

An example of an annotation extracted from the data used in the experiments described in Sect. 5 is shown in (1). Each gloss term occurs in capital letters, as is the standard (Ó'Baoill and Matthews 2000).

(1) English: Which are the morning flights?
Gloss: `WHICH MORNING FLIGHT WHICH?`

Translating ISL annotated with English words into English raises the question of the annotation being a pidgin form of English that is being translated into English. Such a process would require little 'translation', being forms of the same language. Indeed a simplified form of English is being used in our approach and can be likened to the vocabulary of an interlingual system. But the purpose of this reduced-English description is to represent the signed sentences in the absence of a standard lexical orthography, rather than trying to make the translation task easier. Unlike a pidgin language, this annotation version is not intended to be understood or used for communication, but rather is used simply for transcription purposes. However, it should be noted that annotation such as this is considered by many to be an under-representation of the SL, forcing some constraints of the spoken language onto the SL. In Sect. 5.1.5 we support this through a set of experiments comparing the evaluation scores of just the annotation as a translated 'output' with evaluations performed on SMT output.

From this overview of SLs, we can see that there are a number of challenges facing SL MT that do not occur in the field of spoken language MT. In the next section, we outline previous and current SL MT approaches.

3 Overview of sign language machine translation

SL MT is still a relatively novel area when compared with the proliferation of MT products, on-line tools and large research projects being funded for spoken language MT and localisation. Little more than a dozen systems have ever tackled this area of translation. Despite mainstream movement in MT communities toward data-driven methodologies, SL MT research has predominantly been rule-based, with only a more recent development using empirical approaches. In this section we give an overview of other SL MT approaches.

3.1 Rule-based approaches

3.1.1 The Zardoz system

This interlingual approach describes a multilingual translation system using a blackboard control structure (Veale et al. 1998) for translating English text into Japanese

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

SL, ASL and ISL. This is a comprehensive Artificial Intelligence approach employing multiple task-based components including a PATR-based unification grammar to produce the interlingua. The system offers a complete generation phase including a detailed avatar animation phase (Conway and Veale 1994). However, no translation evaluation experiments are noted.

3.1.2 Albuquerque weather

Grieve-Smith (1999) describes a transfer-based system for translating English into ASL within the domain of weather forecasting. A coded glossing technique called Literal Orthography is employed where letters are used as codes for different phonetic features. Weather data collected from the web is broken into concept chunks using lexical tags and parsed into a semantic representation. ASL parse trees are produced from the transfer rules. There is no additional avatar described, nor are any experiments and evaluation documented.

3.1.3 The ViSiCAST and eSIGN projects

Another system employing a transfer approach has been developed within the ViSiCAST project (Marshall and Sáfár 2002, 2003; Sáfár and Marshall 2002). Working with the language pair English–British SL, they collected sign narratives and information presentations as their choice of domain. SL data for the output takes the form of HamNoSys notation. Although Marshall and Sáfár employ a transfer approach that is language pair-specific, in the initial analysis phase they do parse to a semantic as well as syntactic level drawing similarities with an interlingua-style approach. Discourse representation structures and a head-driven phrase structure grammar are also employed. The system does cater for a wide array of SL linguistic phenomena but it does lack functionality for NMFs, and no specific experiments and results are detailed in this work. The projects do, however, focus on animation generation and we discuss this in more detail in Sect. 5.1.6.

3.1.4 The TEAM project and South African SL MT

The TEAM project (Zhao et al. 2000) was developed using an interlingua-based approach for translating English into ASL. Gloss notation is used along with some embedded parameters such as facial features. Synchronous tree-adjoining grammars (STAGs) are employed and source and target language parse trees are created simultaneously. In the second phase of the translation process the information from the IR is fed into the sign synthesiser where default parameters for the avatar model are appropriated to each sign. As with many of the other systems described in this section, this project does not include information on any evaluations carried out on its performance. However, it must be noted that these systems pre-date automatic evaluation metrics, and that most non-SL MT systems of this time did not include formal evaluations either.

The TEAM project described above is the basis of part of the translation work carried out on South African SL (SASL) (van Zijl and Combrink 2006; van Zijl and Olivrin

2008). This system variant differs from the TEAM project's interlingual approach by using a more language-dependent transfer methodology and augments it with the use of STAG. SASL grammar trees and transfer rules were manually constructed from a prototype set of sentences. The database consists of word and phrase lists of hand-annotated SASL videos. Deictic references and NMFs are taken into account as well as sentence type analysis. The authors have not yet integrated the translation and animation phases, but anticipate that generation will be performed by mapping the notation of the SASL trees to a graphical description which is then plugged into a generic signing avatar (Fourie 2006).

3.1.5 The ASL workbench

A transfer approach using Lexical Functional Grammar (LFG) is described by Speers (2001). The focus of this work is on the representation of ASL rather than the development of a complete system. Speers adopts the Movement-Hold model of sign notation (Liddell and Johnson 1989). This method divides and documents signs as a sequence of segments according to their phonological representation where each phoneme comprises a set of features specifying its articulation. LFG correspondence architectures are used in the transfer module to convert the English f-structure into a proposed ASL f-structure. Then an ASL c-structure is derived in the generation phase with a corresponding p(honetic)-structure representing spatial and NMF detail using the Movement-Hold notation. Speers notes that the output 'document' would ideally be produced in various forms including a gloss-based output, a linguistic notation and animated signing, but these have not yet been designed. No specific MT experiments are documented in the dissertation nor are any evaluation methods described.

3.1.6 Spanish SL MT

San-Segundo et al. (2007) use a transfer approach and incorporate a speech recogniser component for the domain of railway stations, flights and weather information, translating Spanish speech into Spanish SL via avatar. The Phoenix v3.0 parser (Ward 1991) is employed and uses a context-free grammar to parse the input into semantic frames in the analysis phase. The result of the transfer phase is a sequence of parse slots representing the SSL where each slot is a semantic concept that is assigned appropriate SSL gestures. While no specific MT experiments are noted, this system is one of the few described in the literature that does perform some amount of evaluation. Preliminary human evaluations were carried out where Deaf people were employed to assess the system's output. In preliminary experiments, the avatar produced only the letters of the alphabet. The evaluators found that less than 30 % of letters were difficult to understand. This shows a poor level of accuracy, something that may be attributed to the use of a 2D animation figure, further described in Sect. 5.1.6. Regardless of its level of accuracy, their testing method does not test the functionality of the MT system itself, but rather that of the animations.

Later work on Spanish SL focuses on the domain of Driver's Licence and Identity Document renewal and uses a corpus of 4,080 sentences (San-Segundo et al. 2010). Three MT approaches are detailed, a phrase-based SMT model and a finite

state transducer SMT model, as well as a rule-based approach using hand-crafted translation rules. The corpus is annotated using glosses, with 715 sentences also being annotated with HamNoSys, SEA[11] and SiGML.[12] The systems were evaluated using both automatic metrics for text-based MT output and human evaluation for animation-based output. Based on BLEU scores, the rule-based approach achieved the best score at 0.68. Based on 48 dialogues, a mock-up of a real-life situation was carried out with 2 government officials and 10 Deaf users. The users reported an exceptionally high translation accuracy of >90 % despite comments by users about problems with naturalness and normalisation in the SL output.

3.1.7 A multi-path approach

Huenerfauth (2006) describes an approach that combines interlingual, transfer and direct methodologies in what he terms a 'multi-path' approach. Although his work primarily focuses on the generation of classifier predicates (CPs), he describes the architecture of the English to ASL MT system despite the non-CP elements not being implemented. A partition/constituent methodology (Huenerfauth 2004) is used to create 3D trees encoding the ASL information. Focusing on classifier predicates, Huenerfauth proposes an interlingual approach where the interlingua is a 3D visualisation of the arrangement of the objects in the sentence of English input. Generating 3D visualisation models is computationally expensive, so for sentences that would not involve CPs in ASL, a transfer approach is proposed. For input sentences that cannot be handled by the rules in the transfer approach, a direct system is proposed that would produce SEE. Native ASL signers evaluated the CP animations rating the animations on a scale of one to ten for understandability, naturalness of movement, and ASL grammatical correctness. Signers were also asked to compare the output against animations created by a native ASL signer articulating CPs wearing a motion-capture suit, animations created by digitizing the coordinates of hand, arm, and head movements during the performance, and using the information to create an animation of a virtual human character. The CP animations produced by the prototype system scored higher than both other animations in terms of grammatical correctness, naturalness of movement and understandability. The CP animations attained an average score of just over 8/10 for grammaticality and understandability, and almost 7/10 for naturalness. In all cases these are at least 2 points higher than the other animation types.

3.2 Data-driven approaches

3.2.1 RWTH aachen group

The first statistical approach to SL MT emerged with the novel translation direction of SL to spoken language text using DGS and German (Bauer et al. 1999), and the

[11] Sistema de Escritura Alfabetica (SEA) is a is an alphabet-based phonetic transcription system developed for SLs.

[12] Signing Gesture Markup Language (SiGML) is an XML-based coding used to represent the HamNoSys code for signs.

RWTH Aachen group have maintained a diligent focus on developing their SL MT research ever since. The language direction here is the reverse of previous approaches due to this system's focus on recognition technology as a component of the broader MT architecture. Within the domain of 'shopping at the supermarket', they employ gesture recognition technology to a database of signed video for training which is then fed into an SMT system. Each sign is modelled with one Hidden Markov Model (Rabiner and Juang 1989). Based on the trained dataset, an input sentence of a signed video is entered and the best match found resulting in a stream of recognised signs being produced and converted into a meaningful sentence in German using Bayes' decision rule to find the best match. Bauer et al. report that for 52 signs they achieve a recognition accuracy of 94 % and a score of 91.6 % for 100 signs. From this the authors surmise that the translation tool would achieve a PER of 10 %.

More recent, yet unrelated, work also taking place in RWTH Aachen focuses primarily on the German–DGS language pair with translation both to and from SLs. Initial experiments on DGS data were performed within the domain of weather reports (Stein et al. 2006) and subsequent work has addressed the more practical domain of airport information announcements[13] (Stein et al. 2007). The baseline MT engine used for their SL translation is the phrase-based SMT system developed at RWTH (Matusov et al. 2006). The language model employs trigrams that are smoothed using the Kneser–Ney discounting method (Kneser and Ney 1995). Monotone search is used to find the best path as well as various reordering constraints (Kanthak et al. 2005). Such constraints address the variation in word order of different languages and involve acyclic graphs that allow limited word reordering of source sentences. Experiments using three variations of reordering constraints showed that the local constraint that allows each word in the sentence to be moved a maximum of w − 1 (where 'w' indicates the window size) steps to the beginning or end of the sentence is the most successful approach with a window size of 2 yielding an almost 10 % reduction in error rates. Their avatar was developed as part of the ViSiCAST system, as described in Sect. 3.1.3, and a manual evaluation was performed where 30 test sentences were evaluated with an average score of 3.3 on a scale of 1–5, with 5 being the highest.

Currently this group are focussing on adapting their techniques to working with small-scale data (a frequent problem of SMT is scarcity of training data) (Stein et al. 2010, 2012). Using the RWTH-PHOENIX-v3.0 corpus of just over 3,000 sentence pairs, they employ a phrase-based model (PBT Zens and Ney 2008), a hierarchical phrase-based system (the JANE system Vilar et al. 2010) as well as a system combination approach described in Matusov et al. (2008). No manual evaluation is carried out as there is no mention of the output being animated, but automatic evaluation is carried out on the gloss output. BLEU scores range between 0.17 and 0.22, with the combined approach showing the best score.

In addition, this research group are involved in the European FP7-funded project SignSpeak[14] which focuses on continuous SL recognition and translation (Dreuw et al. 2010).

[13] This is parallel data to the set described in Sect. 5.1.

[14] www.signspeak.eu.

3.2.2 Chinese SL MT

Wu et al. (2007) describe a hybrid transfer-based statistical model for translating Chinese into Taiwanese SL (TSL). The authors use a bilingual corpus of 2,036 sentences of Chinese with parallel annotated sign sequences of corresponding TSL words from which CFG rules are created and transfer probabilities derived. A Chinese treebank containing 36,925 manually annotated sentences is also noted and both corpora are used to derive a probabilistic CFG. The Viterbi algorithm is used to deduce the optimal TSL word sequence and best translation. This system is shown to outperform IBM Model 3 (Och and Ney 2000) across the board for their Chinese–TSL translation, with comparative BLEU scores of 0.86 for the proposed model and 0.8 for the IBM Model 3. Manual evaluation was also performed with approximately half of the output translations (in the form of sequences of TSL signs/words) considered to be good.

4 The MaTrEx machine translation system

Having described related research in the previous section, we now describe the approach we have adopted. In this section we detail the MaTrEx MT engine we use to carry out experiments, and in the following section we discuss our data sets, the experimental set up and evaluation results achieved.

MaTrEx is the data-driven MT system developed at the National Centre for Language Technology, DCU (Hassan et al. 2007; Ma et al. 2008; Tinsley et al. 2008; Penkale et al. 2010). It is a hybrid system that can avail of both Example-Based MT (EBMT) and SMT approaches (Stroppa and Way 2006) by combining the resulting chunk- and phrase-alignments to increase the translation resources. We use the MaTrEx system in the experiments described in the next section.

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. The blue boxes in Fig. 4 show the range of options available for integration within the broader system. The sections below describe the components employed in the experiments in Sect. 5.

4.1 Word alignment module

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 into the decoder.

Word alignment for the experiments described in this paper 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

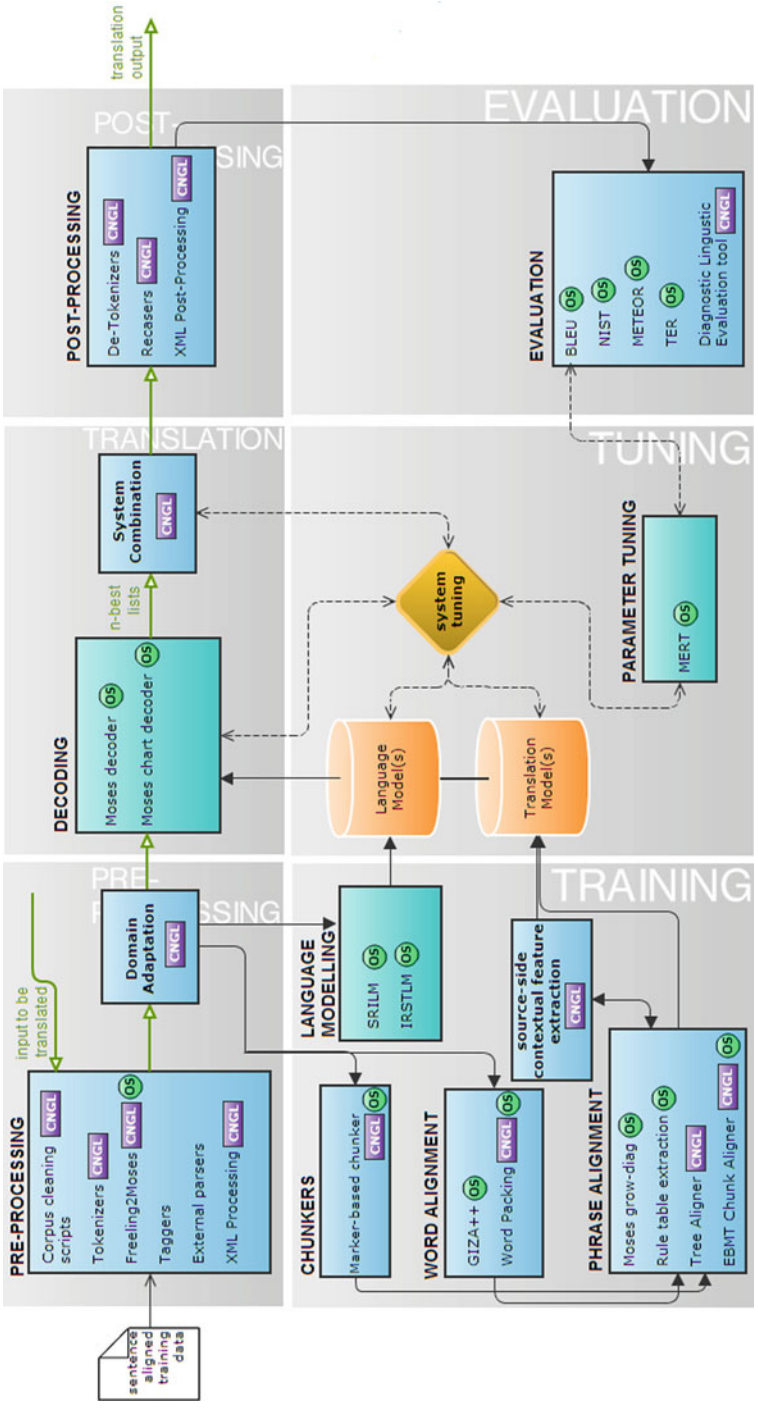


Fig. 4 The MATREX Architecture

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.

4.2 Phrase alignment module

The phrase alignment module is based on that of MOSES (Koehn et al. 2007), where phrases are extracted using the word alignments and scored using phrase translation probabilities and lexical weighting. Only phrasal alignments are used in the experiments described in Sects. 5.1 and 5.2.[15]

4.3 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 Experiments in SL MT using the MATREX system

We now discuss our own data-driven SL MT experiments (Morrissey 2008a, 2011) carried out on purpose-built ISL data sets using the MATREX system. First, we will describe experiments carried out on the Air Traffic Information System (ATIS) corpus, followed by a description of more recent experiments carried out on our medical-receptionist dialogue corpus and ending with a comparison of results. In each section the ISL corpus is introduced and we provide details of data collection and transcription processes. We demonstrate that despite the small amounts of data and initial problems concerning those resources, corpus-based SL MT is not only possible, but can achieve automatic evaluation scores comparable to mainstream spoken language MT.[16]

5.1 Translation using the ATIS corpus

A prerequisite for a data-driven MT approach is a bilingual corpus. The lack of a standardised representation format for SLs, as noted in Sect. 2, means there are few such corpora available, and fewer still with data in a restricted domain and of a large enough size for use in MT as shown in experiments in Morrissey and Way (2005, 2006).[17] As

[15] We performed comparative EBMT experiments on this data, but the results did not prove statistically significant so we have not included them here. More information on these experiments can be found in Morrissey (2008b).

[16] We address the issues facing SL MT evaluation, both automatic and manual inspection, in Sect. 6.

[17] This is beginning to change with the recent development of the RWTH-Phoenix corpus as part of the SignSpeak FP7-funded project. This corpus, as described in Stein et al. (2012) is comprised of just over 3,000 gloss-annotated sentences and although still exceedingly small compared with the hundreds of thousands of sentences typically used in mainstream MT, is the largest SL corpus currently available.

a result, we chose to create our own with a practical and restricted domain considered a necessity. On consulting with Deaf colleagues, we chose the area of air travel and found a suitable English corpus in the ATIS corpus (Hemphill et al. 1990) consisting of 595 utterances. The ATIS corpus is a dataset of transcriptions from speech containing information on flights, aircraft, cities and similar related information. This corpus is particularly suited to our MT needs as it is within a closed domain and has a small vocabulary. Having originated from speech, we believe that the corpus is particularly suitable for translation into SL, given that signing may be considered the equivalent of speech, both being a direct and person-to-person means of communication. The ATIS corpus has subsequently been translated into German, DGS and SASL and used for SL MT in Germany and South Africa (Bungeroth et al. 2008).

Furthermore, the domain itself has a potentially practical use for Deaf people. In airports and train stations, announcements of changes in travel information are usually announced over a PA system; often such information does not appear on the information screens until later if at all. It is also not displayed in the first and preferred language of the Deaf. For this reason, many Deaf people find themselves uninformed about changes to schedules and gates through no fault of their own.

In many airports and train stations worldwide, travel information is entered into a system that announces the changes in an electronic voice. This system could be extended to accommodate SLs without too much difficulty. The limited range of statements and information used in these circumstances could be compiled into a corpus and the information that is announced could be translated into SL and displayed on the video screens for the Deaf to view.

It is important to be guided by the potential users of the technology we are developing as we outlined in Morrissey and Way (2007). For this reason, we engaged the assistance of the Irish Deaf Society[18] to find two native ISL signers to translate the English data to form our bilingual corpus, and to advise us on ISL grammar and linguistics in the area of corpus development. During this signing phase, it was suggested by the signers to change the cities to Irish locations in order to localise the content, but also to facilitate signing. Many Irish towns and cities have either a specific sign or fingersign[19] to represent them, whereas cities and towns outside of Ireland that are not in common usage and are fingerspelled can disturb the flow of signing, particularly if they are orthographically long. For these reasons, we chose to alter the ATIS corpus to use Irish locations and create an amended version for our use. A sample set of sentences is shown below:

- I'd like a flight
- What flights leave from Dublin to Kerry?
- Which are the morning flights?

[18] http://www.deaf.ie.

[19] A fingersign, or lexicalized fingerspelled sign, is a recognised lexical item that began as a sign spelled in English orthography: *bus*, for example, is signed by signing each letter sequentially, which has subsequently shortened in spelling through common use and is "perceived as consisting of one overall movement of a particular handshape and …considered to belong to the sign vocabulary of ISL" (Ó'Baoill and Matthews 2000, p. 118).

We chose annotation as our representation method as it does not require symbolic notation formats to be learned, and it allows for flexibility in the granularity of annotation, so these can be as simple as glossing signs or as complex as including a phonetic description. Moreover, SL linguistic trends seem to favour annotating data [cf. SL linguistic projects described in Neidle et al. (2001), Leeson et al. (2006), and Crasborn and Zwitserlood (2008) as well as the corpus development work as part of the SignSpeak project (Stein et al. 2012)], which increases the likelihood of larger, already annotated corpora being available for MT use in the future.

We manually annotated the ISL videos created with the assistance of the Deaf signers using the ELAN software tool (see Sect. 2.2.3). Given that annotation is a time-consuming process, we limited our annotation to a glossing of the hands. We kept the glosses themselves quite simple using basic root forms of English words (see Sect. 2.2.3), as is standard in linguistic annotation of SL videos. Our annotation included some of the linguistic phenomena outlined in Sect. 2 of this article, such as deictic references and classifiers, but we chose to omit the inclusion of NMFs at this point.

Omission of NMFs from the annotation means important grammatical and semantic information is absent from the annotations which will ultimately affect the translation quality. Further annotation of the data to include NMFs would not have allowed for the comprehensive set of experiments, including the addition of an animation module, to be performed. Furthermore, while we feel that the addition of NMFs is crucial for complete SL representation and MT, we had sufficient data to begin experimentation with the intention of the post hoc addition of NMFs including an evaluation comparison. NMFs are not entirely neglected, however, as we have included them in the animation module, as noted in Sect. 5.1.6.

The 595 sentences of the English (EN) ATIS corpus were also translated into German (DE) and then DGS gloss annotation by Daniel Stein at RWTH Aachen.[20] This provided us with four parallel corpora, already sententially aligned, with the potential to work with four translation pair types for a total of 12 language pairs:

(i) from SL to spoken language (ISL–EN, ISL–DE, DGS–EN, DGS–DE),
(ii) spoken language to SL (EN–ISL, DE–ISL, EN–DGS, DE–DGS),
(iii) spoken language to spoken language (EN–DE, DE–EN),
(iv) and the novel translation pairings of SL to SL (DGS–ISL, ISL–DGS).

Each data set underwent a preprocessing step to extract the aligned sentences and annotations in preparation for the next phase: translation.

In the following sections, we describe the experiments we ran using the amended ATIS corpus in the MaTrEx system. Experiments were performed in both directions with supplementary modules added as described in the following sections. For all experiments the 595 sentences of the ATIS corpus were divided into 418 training sentences, 59 development sentences and 118 test sentences. Corpus statistics are provided for the English, German, ISL and DGS data in Table 1.

[20] Jan Bungeroth at RWTH Aachen arranged annotation of the data in SASL. This work was carried out separately and is not included in this paper.

Table 1 ATIS corpus overview

	EN	DE	ISL	DGS
All				
No. sentences	595			
Avg. sent. length (words)	7.34	7.67	7.4	6.74
Standard deviation	3.49	3.68	3.48	3.31
Train				
No. sentences	418			
No. running words	3008	3544	3028	2980
Vocab. size	292	327	265	244
No. singletons	97	118	71	84
Dev				
No. sentences	59			
No. running words	429	503	431	434
Vocab. size	134	142	131	119
Test				
No. sentences	118			
No. running words	999	856	874	877
Vocab. size	174	158	148	135

We used standard automatic evaluation metrics for assessing improvement across experiments, namely the string-based method of BLEU (Papineni et al. 2002) as well as two error rate methods, word error rate (WER) and position-independent word error rate (PER). BLEU is a precision-based metric that compares a system's translation output against reference translations by summing (roughly) over the 4-grams, trigrams, bigrams and unigram matches found, divided by the sum of those found in the reference translation set. It produces a score for the output translation of between 0 and 1. A higher score indicates a more accurate translation. WER computes the distance between the reference and candidate translations based on the number of insertions, substitutions and deletions in the words of the candidate translations divided by the number of correct reference words. PER computes the same distance as the WER without taking word order into account.

5.1.1 Translating sign language gloss to spoken language text

We performed the following experiments translating ISL glosses into English text:

1. Baseline of annotation, without translation,
2. MaTrEx SMT components only,
3. Distortion limit increase.

The baseline experiment was performed using only the ISL annotations, as shown in Sect. 2.2.3, without performing any MT but evaluating on the annotations as a 'translation'. As described in Stein et al. (2010), this provides a 'sanity check' to assess if

simply lowercasing the gloss version of the SL data provides an adequate translation of the text (from SL).

The first MT experiment uses solely the SMT components of the MaTrEx system, without the addition of sub-sentential chunking.

A further experiment was carried out across the MT sub-experiments which assessed the benefits of allowing jumps (Leusch et al. 2006) by changing the distortion limit (Morrissey et al. 2007b). The distortion limit function allows for jumps or block movements to occur in translation to account for the differences in word order of the languages. The limit is set to 0 jumps as the default and we found that allowing a distance range of 10 jumps when decoding produced the most significant increase in scores. Given that the average sentence length in words for all data is approx 6–7 words with an average standard deviation of around 3, the high distortion limit allows for jumps across the whole sentence in these experiments and we believe accommodates the freer word order of SLs compared with spoken languages.

Comparative scores for all the above experiments are shown in Table 2. While it can be considered that some degree of human 'translation' is performed when annotating the SL videos, we can see from Table 2 that the SMT systems outperform the annotation baseline with the BLEU score alone doubling. Similar comparison of annotation baseline against MT output is performed by Stein et al. (2010) which corroborates our results and finds annotation is significantly inferior to MT output.

Comparative sample candidate translations produced by each experiment are shown in (2).

(2) a. *Baseline annotation format:*
WHICH MORNING FLIGHTS WHICH ?
b. *Reference translation:*
Which are the morning flights?
c. *SMT translation:*
which is the morning flights

In the above candidate translations, although the sentences are all similar and capture the gist of the translation (i.e. a query about morning flights), we can see the differences that are reflected in the evaluation scores. For example, the baseline annotation format lacks any verbal information, while the SMT translation has produced a grammatically incorrect form of the verb *to be*. All MT outputs omit punctuation information. Although this is only a sample set of translations, it illustrates how even slight deviations from the reference translation can affect evaluation scores.

Table 2 MaTrEx Evaluation results for ISL–EN gloss-to-text experiments

	BLEU	WER	PER
ISL–EN			
ISL annotation baseline	25.20	60.31	50.42
SMT	51.63	39.32	29.79
Dist. Limit = 10	52.18	38.48	29.67

Table 3 MaTrEx comparative evaluation results for all gloss-to-text experiments

	BLEU	WER	PER
ISL–DE	39.69	47.25	38.47
DGS–EN	48.40	41.37	30.88
DGS–DE	42.09	50.31	39.53
ISL–EN	52.18	38.48	29.67

From the scores in Table 2, we can also see that the distortion limit improves results. We found that allowing a distance range of 10 jumps for block movements when decoding produced the most significant increase in scores. Compared to the SMT scores, BLEU score improved by 0.55 % (1.07 % relative), WER by 0.84 % (2.14 % relative) and PER by 0.12 % (0.40 % relative). The change in the distortion limit to 10 may increase the number of correct words found, thus lowering the error rates, but this may decrease the number of correct *n*-grams in the candidate compared to the gold standard. One way of improving this would be to have multiple gold standard reference texts, but this was not possible in our experiments due to lack of parallel data.

5.1.2 Comparative experiments using DGS and German parallel data

Having had the ISL data translated into both German and DGS, we ran comparative experiments. The results are shown in Table 3.

Using only the basic SMT components of MaTrEx, all systems score within a 13 % BLEU score range of each other, which indicates clearly that translation performance is dependent on the language pair.

Table 3 also addresses, albeit inconclusively, the question as to whether the annotations may be considered as a pidgin form of English/German that is being translated. While it is likely that the ISL–EN scores would be boosted by the use of English annotations and the evaluation scores, when comparing ISL–EN with ISL–DE, certainly appear to indicate this, the large difference between ISL–EN and DGS–DE evaluation scores (10.09 % absolute BLEU score difference, an almost 20 % relative decrease from ISL–EN to DGS–DE) does not indicate that this is also the case for DGS–DE. The evaluation scores of SL-to-spoken language experiments reflect that the use of German glosses does not indicate a bias toward improved results for the DGS–DE language pair over DGS–EN. In fact, the DGS–EN pair achieves better scores (44.34–48.40 % BLEU score) across the board than DGS–DE (34.86–42.09 % BLEU score).

5.1.3 Translating sign language video to spoken language text

We next sought to expand the system to make it more practical with the addition of SL recognition technology. An SL system that only produces annotation and an MT system that only accepts annotation as input are not usable systems by themselves, but when combined, they have the potential to contribute to the development of a fully automated SL-to-spoken language text system.

As the primary focus of our work is the translation component, we cooperated with the Sign Language Recognition group in RWTH Aachen to extract data in a glossed format from the same ISL videos that we manually annotated in Sect. 5.1. Their automatic sign language recognition system is based on an automatic speech recognition system (Dreuw et al. 2007) with a vision-based framework.

Our ATIS video data was taken for recognition experiments using the RWTH system (Stein et al. 2007). The experiments focussed on the dominant hand glosses and consisted of a basic feature setup to begin with. Despite the system's previous success, a good result was not obtained using our ATIS videos due to only a small set of features being used. The preliminary recognition of the videos had an error rate of 85 % (consisting of 327 deletions, 5 insertions and 175 substitutions out of 593 words). Given the initial poor score for the ISL data and the fact that combining any two systems will introduce additional errors, it is apparent that to use such inaccurate data to seed an MT system would be inappropriate at this time. Had a more in-depth feature recognition process been performed, we estimate that, based on previous success using ASL video data (Dreuw et al. 2007) the recognition output would be sufficient for use with our MT engine.

5.1.4 Translating sign language to synthetic speech

Having already explored many of the translation possibilities for SL and spoken language, we decided to further exploit the possible uses of our MT engine by adding on a speech synthesis module (Morrissey et al. 2007a). In the context of a fully functioning, bidirectional SL MT system, speech (as opposed to text) is a more natural and appropriate output for the spoken language. This is because speech is more akin to signing than text as they are both direct means of communication, produced face-to-face.

In order to explore this avenue, we collaborated with the Multilingual Ubiquitous Speech Technology: Enhanced and Rethought (MUSTER) research group in University College, Dublin.[21] The MUSTER group has developed the Jess system for synthesising speech in various languages (Cahill and Carson-Berndsen 2006). It is a modular system that allows for different synthesiser algorithms to be plugged in and tested using the same source and target data. Voice data is stored in four formats: utterance, word, syllable and phoneme. Text can be input in orthographic form and the system estimates a phonetic transcription of this from which the best pronunciation output is calculated.

For our experiments, we provided the MUSTER group with some of the English text output from the experiments described in Sect. 5.1.1. This was fed into the Jess system to produce English speech. Taking only minutes to complete the entire process, this was an easy task and seeds further collaboration between our groups, which will take place as part of the EUR 30m Centre for Next Generation Localisation project.[22]

[21] http://muster.ucd.ie.

[22] http://www.cngl.ie.

Table 4 Evaluation scores for spoken language-to-SL experiments: comparison with best SL-to-spoken language scores shown

System	BLEU	WER	PER
EN–ISL	38.85	46.02	34.33
ISL–EN	52.18	38.48	29.67
DE–ISL	25.65	57.95	46.62
ISL–DE	42.13	45.45	38.16
EN–DGS	49.77	45.09	29.59
DGS–EN	48.40	41.37	30.88
DE-DGS	47.29	45.90	28.67
DGS–DE	42.09	50.31	39.53

5.1.5 Translating spoken language text to sign language gloss

Given the satisfactory results of translating SLs into English and German text in Sects. 5.1.1 and 5.1.2, we decided to exploit the bidirectional functionality of our data-driven system and reverse the translation process. Translating spoken language into SLs has the potential to be directly useful to the Deaf community, enabling them to access information without the need for an interpreter for brief scenarios where short, immediate translation is required (as opposed to interpretation of a whole conversation or whole document).

Our initial experiments translating from English and German to SLs took text as input and produced ISL and DGS glosses. The results are shown in Table 4.

In addition, we note the different range of scores for each language pair. The EN–DGS pair achieves the highest scores across all experiments and DE–ISL achieves the lowest, with an approximate difference of 25 BLEU points between them. This drastic difference in scoring may be attributed to similarities and differences between the languages in question, meaning EN–DGS may have achieved better results if English and DGS annotation are more similar in format than the German and ISL pair. However, manual examination of the texts did not show an obvious similarity between English and DGS or German and ISL to substantiate this.

We also note from these results that the overall scores are not as good as for the reverse language direction, with BLEU scores alone showing a difference of between 13.07 and 13.33 % for the EN–ISL pair (25.05 and 25.55 % relative difference respectively).

Given that German is a morphologically rich language which can cause problems for MT (as noted in Stein et al. 2012) we would also have expected the DE–DGS scores to be more comparable with DE–ISL, yet it achieves a much better score (5.02 BLEU point improvement) over the reverse language direction. We anticipate that this is due to the language direction and that translating out of German poses fewer problems than reproducing this morphologically complex language.

The MT framework and translation methodologies are unchanged for these experiments, and it is likely that the evaluation metrics do not adequately capture the intelligibility of the output translations but more assess their fidelity with respect to the single gold standard. Additional gold standard data would provide a broader range

of reference translations which may provide better matches for candidate sentences. Given the somewhat freer word order of SLs than, say, English, it is likely that such set of variations is not out of reach but would require a team of ISL speakers and linguists for compilation.

It is likely that a trained human judge could better evaluate the output for intelligibility. We investigated this by performing human evaluations of animated output as discussed in the next section. Furthermore, producing ISL glosses as a 'translation' does not facilitate the wider Deaf community as it is still not in their native language. For these reasons, in the next section we further develop our research and produce animated SL from the translated glosses.

5.1.6 Translating spoken language text to animated sign language

Translating into SLs has a significant practical use for the Deaf community, but a system that produces gloss output is not of much use to a Deaf person and is more likely to be confusing by providing spoken-language stem words in an SL syntax. For this reason, the next natural step was to produce real SL output in the form of an animated computer figure or avatar.

For this SL generation module, we chose 50 randomly selected sentences from the output of our EN–ISL experiments described in Sect. 5.1.5. Each annotated gloss, or 'token', from the MT output was made into a separate video and these videos were joined to form SL sentences signed by our 3D avatar. Before describing our our work in this area, first we will briefly outline what others have done in this area.

Many of the previous approaches described in Sect. 3 have employed avatars that the authors themselves have developed:

- Zardoz: The authors describe a sign synthesis methodology in earlier work (Conway and Veale 1994). Their focus is on the importance of describing the internal phonological structure of SLs as an essential factor for the synthesis of native SLs in order to generate fluid signing and allow for inflectional variation.
- ViSiCAST: Marshall and Sáfár (2002), Sáfár and Marshall (2002), and Marshall and Sáfár (2003) HamNoSys symbols of their translation output are converted to an animated avatar using an interface developed at the University of East Anglia. The symbolic representation must satisfy grammatical constraints and a library of 250 BSL signs contains both fixed description animations as well as parameters that can be instantiated with variables to allow for directional verbs, for example.
- RWTH Aachen: this group also avail of the ViSiCAST avatar, whereby their annotated output is converted into HamNoSys before being fed into the animation interface (Stein et al. 2006). A similar parameterised approach to the ViSiCAST system is employed by the TEAM project (Zhao et al. 2000), where they use parameterised motion templates in their ASL dictionary that are combined with parameters from the intermediate representation to animate the avatar.
- SASL: the team have developed an avatar (Fourie 2006), but it has not yet been integrated with the MT system to accept translated output.
- Spanish SL: this group have produced an avatar where the SSL semantic concepts produced by the MT system are aligned in *n*:*m* alignments with the SSL ges-

tures (San-Segundo et al. 2007). A basic set of body positions and facial features are described, while continuous signing is achieved via interpolation strategies between these basic positions.

– Multi-path: Huenerfauth (2006) continued his focus on classifier predicates to the avatar stage. Previously calculated discourse models, predicate-argument structures and the visualisation scene are stored in a look-up table for each English sentence along with a library of ASL hand shapes, orientations and locations. These representations are then fed into an animation system developed by the Centre for Human Modelling and Simulation at the University of Pennsylvania.[23]

The avatar technology used in these approaches varies as to how realistic the animated models look. While some approaches, such as the SSL project, favour basic 2D cartoon-like models formed from geometric shapes, others such as the ViSiCAST project have developed 3D models that are as lifelike as a computer-generated character from a computer game. Given the preference by Deaf people for human signing over computer-generated signing (Naqvi 2007), it is likely that the more realistic avatars will prove more successful with the Deaf community.

As our research is primarily focussed on MT rather than animation production, in order to achieve the most lifelike avatar, we chose a commercial animation tool, as opposed to developing our own. We found suitable characters for our own avatar in Poser Animation Software, Version 6.0.[24] The Poser software tool enables the animation of 3D human figures similar to those generated for in computer game software.

Taking the 50 candidate annotation translations from our EN–ISL experiments, we segmented the data into individual words, giving 246 tokens with 66 individual types. Each of these 66 ISL types was then created into an individual video using the manual animation process described at the beginning of this section. However, this method compensates for the loss of NMF detail in the transcription by providing it in the animation process thereby providing some disambiguation[25] and defining inflection for example. Ideally this process would be automated through the MT process via a more detailed annotation.

Rather than developing our own mannequin, we chose the businessman figure from the selection of pre-created characters rather than some of the more video-game-like characters available. The mannequin is shown in Fig. 5.

Ideally, each animated sign would blend seamlessly with the next in fluid, natural articulation. As each sign produced by our system is animated manually and individually, this currently presents some fluidity problems when joining the signs to form sentences. It is not possible to join the manually created signs together in Poser in real time to avail of the interpolation technique and thus avoid 'jumping' between animations. In order to minimise this problem, we formatted each animation so that

[23] http://cg.cis.upenn.edu/.

[24] http://www.curiouslabs.de/poser6.html?\&L=1.

[25] Given the nature and type of the discourse chosen, there is little ambiguity to be found in the text and typically each discourse has only the one interpretable sense. However, the human nature of annotation would mean the disambiguation of terms is carried out before MT should ambiguity arise in the data, and that this affects the MT process by simplifying the task somewhat.

Fig. 5 Robert: The POSER 6 Avatar

the mannequin began and finished each individual sign animation in a neutral position with his hands resting in front of and close to his body and his face relatively expressionless, as in Fig. 5. While it is not natural to pause in this neutral position between each sign in normal discourse, this smoothing methodology helps to avoid sudden jumps between different hand configurations and positions, but does not remedy the issue of fluidity. Further interpolation methodologies such as parameterised templates (cf. the ViSiCAST system and TEAM project animation approaches described above) as input to the Poser avatar would be more appropriate here and will be adopted for future development for optimum transition fluidity, naturalness and comprehensibility.

The original ISL videos were examined to confirm the correct articulation of the 66 signs. These videos were also employed to provide examples of natural body movement and NMF detail (NMFs have been included at this level to help compensate for their omission in the annotation phase), particularly mouth and eyebrow patterns. This enabled us to improve the already human-like mannequin by adding in natural body movement to make him less stiff and robotic in nature.

Given that gesture recognition technology is not yet at a level where it can evaluate animated SL, we chose to have our animations manually evaluated. Following manual evaluation procedures for spoken language MT, such as those outlined by Pierce et al. (1966) and van Slype (1979), we chose to recruit 4 ISL evaluators, or 'monitors', for the task. These monitors were native Deaf signers sourced via the Irish Deaf Society[26] who use ISL as their main means of communication and are active members of the Deaf community. Each of the 50 animated translations was evaluated in terms of *intelligibility* (assessing how understandable an animation was) and *fidelity* (assessing how good a translation of the English original the animation was). A scale of 1–4 with qualifying descriptions were used for each metric:

Intelligibility:

1. Incorrect or too confusing to grasp the meaning

[26] http://www.deaf.ie.

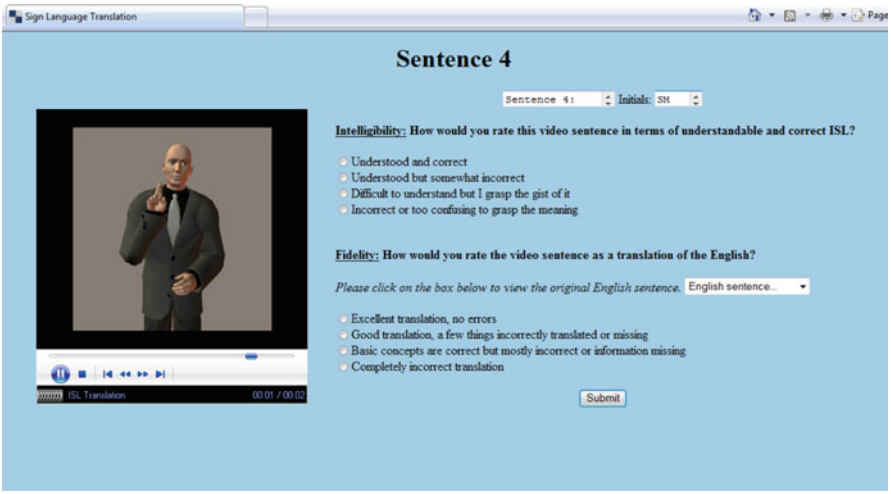


Fig. 6 ISL Evaluation Sample Sentence Evaluation Page

2. Difficult to understand but I grasp the gist of it
3. Understood but somewhat incorrect
4. Understood and correct

Fidelity:

1. Completely incorrect translation
2. Basic concepts are correct but mostly incorrect or information missing
3. Good translation, a few things incorrectly translated or missing
4. Excellent translation, no errors

Roughly speaking, this allows the evaluator to attribute a completely negative, mostly negative, mostly positive or completely positive rating to each translation. The evaluation was completed by filling in a questionnaire of general queries about the translations and technology.

We developed a web-based format for the evaluations, as shown in Fig. 6. In order to develop an evaluation interface that is both functional and facilitates usability, we used the "Eight Golden Rules for Interface Design" (Shneiderman 1998).

Intelligibility and fidelity scores along with the answers to the questionnaire were collected from the monitors. Bar charts detailing the intelligibility and fidelity scores are shown in Figs. 7 and 8, respectively. In both figures the x-axis shows the number of sentences and the y-axis the score given; different colours represent the different monitors. We can see from these charts that 82 % of animations were considered intelligible by the monitors and 72 % were considered good-to-excellent translations. While the manual evaluations of this work show a decidedly positive assessment of the animation avatars, a larger number of monitors would give a broader representation of how good the translations were. In addition, it must be remembered that due to the nature of the task, monitors—while in no way affiliated with us—may have attributed a more positive evaluation in a willingness to encourage the development of such technology. We refer the reader to Sect. 5.1 earlier in this paper, where we indicate

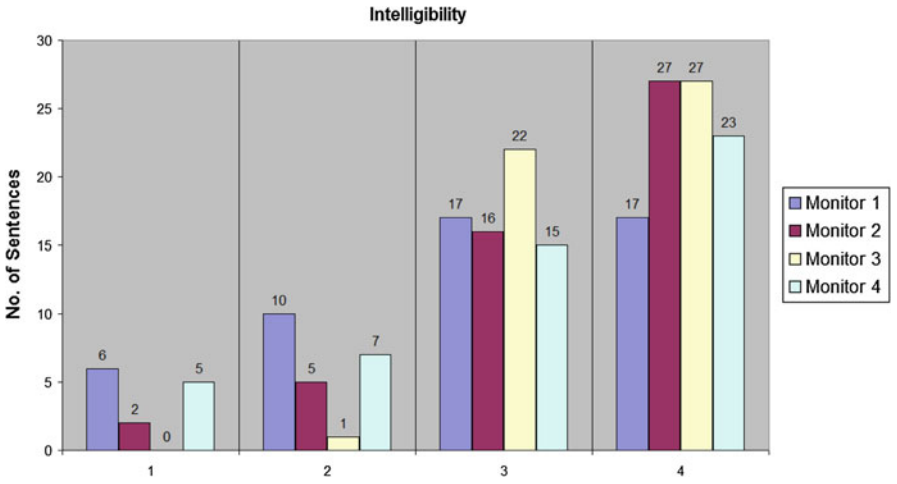


Fig. 7 Manual Evaluation: Intelligibility Scores

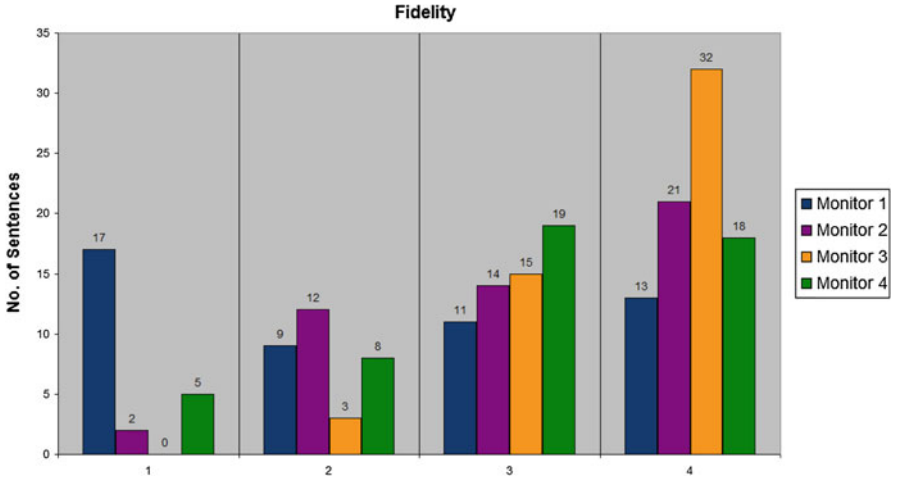


Fig. 8 Manual Evaluation: Fidelity Scores

the kind of information missing from the gloss annotation with the absence of NMFs (grammatical and semantic information, for example the non-inclusion of a head-nod could transform the meaning of a sign from its original negative intention to a positive one through omission of this important NMF, and thereby change the meaning of the sign or sentence). There is the likelihood that the human monitors compensated for loss of NMFs or other details missing from the animation translations and potentially provided a more favourable score than was deserved.

Given that the monitors were only provided with content from our own experiments for these evaluations, it must be noted that there is no defined baseline against which to compare the results. The results should be seen more as a rough guide on which to build future work and further such manual evaluations will be more comprehensive and scientifically valid.

5.2 Translation using the medical appointment corpus

Subsequent and more recent work on SL MT has focussed on the creation of a multimedia parallel corpus 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 have constructed a corpus representing the dialogue between the two parties when scheduling an appointment.

A multitude of literature (e.g Jones and Gill 1998 and many others) confirms that the primary obstacle to healthcare for non-native speakers is the language barrier. Traditionally interpreters are used to address this problem, and rightly so. However, in the area of appointment scheduling where the hire of an interpreter for such a short interaction is expensive, technology may have a solution. Ongoing 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 2003; 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).

The corpus is a multimedia four-way parallel corpus[27] consisting of:

- audio recordings of the original material
- written English transcription of the audio
- ISL video recordings of English
- HamNoSys transcription

In order to create the initial corpus we required dialogue. Matters of confidentiality and other ethical issues preclude the collection of genuine data in situ. While 'standardised patients' (trained actors simulating the situation) are used in medical teaching scenarios, training such actors is a major undertaking and outside the scope of our research project, so we chose to instead engage the expertise of a medical receptionist. Following the receptionist's guidance, we role-played a number of scenarios including general appointment scheduling with a GP, scheduling specific activities (i.e. vaccinations) and changing or cancelling an appointment. 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 comprising 7.04 words with a standard deviation of 3.82.

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

[27] A translation of the corpus was also made into Bangla, a low resourced Indian language for parallel research into minority language MT, but will not be discussed further here. Interested parties can find out more by consulting (Dandapat et al. 2010).

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 order to make the ISL machine-readable, this time we chose to use HamNoSys transcriptions based on the premise that using glosses entails describing one language in terms of another, thereby potentially misrepresenting or under-representing the SL (Pizzuto et al. 2006). To facilitate transcription of the ISL videos, we used the eSignEditor developed at the University of East Anglia (Kennaway et al. 2007). This software tool facilitates the transcription of SL videos using a library of HamNoSys symbols. Part of the transcription involves assigning a reference to the sign being transcribed, this takes the form of a gloss, akin to that described in the previous set of experiments. Each sign is also attributed a numerical sign ID code unique to that version of that sign. The tool also provides a facility whereby the user can highlight transcriptions and have them signed via an animated avatar. This process converts the HamNoSys symbols into an XML script called SiGML (Signing Gestures Markup Language). In the following experiments we compare sign IDs, SiGML code and glosses as transcription methods for this data set. All experiments described below are English to ISL and are evaluated automatically using the same evaluation metrics as in the previous set of experiments. No animation or manual evaluations have taken place so far.

5.2.1 Experiments on medical dialogue data

We tested various representation methods for this data set. 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 while the test and training data are mutually exclusive at sentence level, all vocabulary in the test data is present in the training data.

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 in (3). 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):

(3) < 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 >

This example of HamNoSys shows the gloss and Sign ID contained within a tag (denoted by opening and closing angle brackets: < and >) as part of the code. Additional tags represent various body parts used, with the bulk of the text, namely words beginning with the prefix 'ham-' representing the individual phonemes of a sign that can later be articulated using an animated avatar.

5.2.2 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 unique ID code attributed to it within the eSignEditor system. Based on the argument of spoken-language glosses potentially misrepresenting signs, 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. The average sentence length in ID numbers (corresponding to signs) is 5.66 with a standard deviation of 3.28. An example of the sign ID format for the phrase 'goodbye' is shown in (4a). The second example in (4b) shows the sign ID sequence in (3) matching the phrase 'When do you want to come in?'.

(4) a. `160`
b. `15 27 17 18 15`

Using sign IDs allows for the detailed description of the sign (provided by the associated HamNoSys as indicated in (3)) to remain intact. The MT output produced would also take the form of sign IDs and these can then be looked up 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.

5.2.3 English glosses

Given that by using eSignEditor we had access to a gloss-based representation of each sign, for comparative purposes, we extracted the uppercase 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. The average sentence length (counted in glosses terms) for this data is 5.8 with a standard deviation of 3.41. Corresponding glosses for the above examples are shown in (5a) and (5b), respectively.

Table 5 Evaluation scores for medical dialogue corpus EN-to-ISL experiments

	System	BLEU	WER	PER
ID tags	Test^R	3.45	119.35	115.32
	Test^S	14.98	84.86	77.83
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

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

For the purposes of MT, all glosses are converted to lowercase. The suffix '_AS_1' refers to an alternative sign for 'come'. If there is more than one distinctly different sign in the database with the same 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.

5.2.4 SiGML code

SiGML code provides us with the HamNoSys tags that directly correspond to the HamNoSys symbols used to describe the phonetic features of 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' was shown in (3). Given the verbose nature of the HamNoSys tags, for reasons of space we will not show an example of a full sentence. In order to 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. This gives us an average sentence length (measured in 'signs') of 5.67 with a standard deviation of 3.31. It is worth noting that the type-token ration is very low for this data at 1.4:1 in comparison with the English (3.13:1) and the other transcription formats (2.4:1 for glosses and 3.12:1 for ID codes). This could be accounted for by the already mentioned verbose nature of the SiGML format and the unlikelihood that tag sequences are exactly repeated within the text without small differences.

5.2.5 Evaluation results

Each of the above versions of the corpus was tested with corresponding test sets. The output of each was evaluated against one reference sentence using BLEU, METEOR and WER and PER. Table 5 shows the evaluation results for the six experiments carried out. Random and special testsets are indicated respectively by Test^R, Test^S.

5.2.6 Discussion of results

From the above results we see that using ID numbers gives by far the lowest evaluation scores across the board.[28] This may be in part related to the use of number codes to represent the signs. Gloss and SiGML datasets achieve more comparable scores with SiGML achieving an improvement of almost 15 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 probably accounts for this.

Although it appears on first sight 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 slightly different suffixes for the 'ham-' codes and different ID numbers and gloss words. The repetition of tags and code greatly outweighs the differences resulting in the inflated BLEU scores. 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 whether the translation quality is good or not.

Although glosses are not considered to be 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.[29] However, we must bear in mind the caveat that all these means of representation may not appropriately encapsulate inflectional and other linguistic information that may be integral to the understanding of a signed utterance. Until the output is produced via a signer or signing avatar, it is difficult to ascertain what information is present, and indeed this is one of the arguments for the inadequacies of transcription methods for SLs, namely that the complex gestural-spatial performance nature of SLs and the linguistic artefacts 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.

Although we have not yet scheduled manual inspection and evaluation of the signed output for any of the above experiments, we performed some checks on the HamNoSys tag output in order to assess how much of the HamNoSys content remains intact through the translation process. Missing information or information in the incorrect order will result in the eSignEditor rejecting the text and no animation will be produced.

[28] 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.

[29] 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.

Fig. 9 A Sample of the Avatar provided in the eSignEditor tool

We took the HamNoSys tag output produced by the MT system and post-processed it in order to restore supplementary tags and spacing that was removed in the pre-processing phase. These rudimentary efforts to restore the data and produce an animated output using eSignEditor proved unsuccessful. The input was not animated, suggesting that the eSignEditor parser did not accept the SiGML data. Further work is required here in order to appropriately prepare the MT output for animation, but in such a way that the MT output itself is not post-edited so as to affect or inadvertently improve the translation.

For the purpose of comparison with previous efforts, an image of the signing avatar used in the eSignEditor is shown in Fig. 9.

6 Discussion on main translation issues: transcription and evaluation

6.1 Transcription methods

It is well documented that data-driven MT requires a sizable corpus from which to glean the required information to proceed with translation. It is also well documented that there is no one agreed standard text-based representation for SLs where the signed utterance is considered to be completely represented. These two facts raise interesting challenges for SL MT. We know that a corpus is a highly valuable linguistic resource for MT whether it seeds the training of data-driven approaches, or provides SL linguists with a resource from which to study the language and derive source and target information for a rule-based approach. However, the primary problem is not what to do with the data once it we have it, but how to get it, get enough of it and in both an appropriate format to fully represent the SL and in a format that is suitable for MT processing. The obvious solution to data scarcity and varied formats is to create a central data repository and to develop an agreed standard for representation through consultation with Deaf people, SL linguists and NLP developers and researchers. In the absence of such a repository and standards, researchers much make do with creating their own data (usually resulting in exceptionally small datasets by mainstream MT standards), sharing data (which can be difficult if there are IP constraints) or sourcing

external data (where the domain is typically wide and unrestrained making MT even more challenging).

If we are to build our own corpora, do we take advantage of the glossing methodology and how well it facilitates MT, knowing that the SL is possibly misrepresented, incorrect and at the very least disambiguated meaning it becomes more human-aided MT? Currently this seems to be the most accepted approach until something more suitable is available.

A further issue with transcription is the amount of information the target SL representation must contain. If we consider the representation as a transcribed, text-based annotation, the less detail provided the easier the translation (as evidenced in the experiments carried out in Sect. 5.2). In contrast, the more detail there is, the better the animation as there is more information to seed it. This issue could be overcome through the use of ID tags or glosses where the information necessary for animation is encapsulated in the tag or gloss and referenced through a lexicon.

On the other hand, if we consider the target representation to be animation via avatar, the animation must be articulate, human-like and competently able to use non-manual features and the signing space. However SL animators are then faced with the issue of the 'uncanny valley': how realistic can an avatar be before it gets disconcertingly real? However, this is a matter for SL animators.

Both these output representations do raise the question as to whether perfect output necessary? Mainstream MT puts a lot of stock in its ability to 'gist' and that is quite often sufficient, not perfect but still helpful, either to the post-editor or to the general public. SL MT is certainly not trying to replace interpreters, in the same way as MT is not trying to replace human translators, but there is a place for SL MT and surely gisting is better than nothing at all. The matter here is to define where the line of acceptability lies.

6.2 Evaluation methods

In the section on related research, we noted that many of the previous approaches to SLMT do not detail or did not carry out evaluation either using automatic MT evaluation metrics or manual evaluations with Deaf people. This is understandable where systems were developed before mainstream MT evaluation metrics were prevalent. If we first take automatic evaluation, we know that is an objective process whereby the output is compared against some 'gold standard'. It must also be borne in mind that it is not yet possible to automatically evaluate avatars, so in order to objectively evaluate SL MT output the raw text-based output can be evaluated by adopting the metrics used in mainstream MT. But what exactly is being evaluated here? We cannot say that we are evaluating SL output, because it is not SL, but rather a transcribed form. More often than not, only one reference translation is available for comparison with the MT output, so in this case we are further restricting the evaluation and potentially missing accurate translations where synonyms or different sentence structures are used.[30] But this is not to say that automatic evaluation metrics are not useful. On

[30] Some automatic evaluation metrics now process synonyms, such as METEOR.

the contrary, they can be quite effective at ascertaining the internal progress of an MT system with a previous instantiation following some tweaking of the engine. They can be quite successful at indicating some degree of improvement, but this will not be known for sure until the actual SL output is observed and evaluated.

This takes us on to human evaluation. In a complete MT system producing any SL it is imperative to evaluate the actual signed output. In the absence of automatic animation evaluation, human evaluation is a must.[31] But this too has its problems. Humans are subjective and are likely to be influenced by factors outside of the actual translation quality or accuracy, such as their opinion of the avatar and how it looks and their opinion of animated signing in general. Furthermore, 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.

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 to assess the quality of the output as best we can until something more objective is developed.

7 Conclusions and future work

In this article we explored the application of the MaTrEx data-driven MT system to ISL and DGS, and addressed a number of related issues including SL linguistics, data representation, data-driven translation to and from SLs, supplementary modules to the translation process and evaluation.

Our exploration of SL linguistics discussed phenomena that are interesting for MT, such as classifiers and NMFs. We also discussed notational representations for SLs, including video annotation, which we chose for our own experiments. Overviewing past approaches, we showed that SL MT is still a fledgling field of research, yet with a wide variety of both rule-based and data-driven approaches taken.

Within the context of our own work, we described two different corpora on which we carried out multiple experiments. With the ATIS corpus, we demonstrated the first data-driven MT system to make use of example-based sub-sentential information for SL MT, namely the MaTrEx system. Through sets of experiments translating English and German into DGS and ISL, as well as ISL and DGS into English and German, we showed that the MaTrEx system produces twice as good results compared to annotation alone. Experiments changing the distortion limit also proved successful, with an allowance of 10 jumps creating the best result and improving scores across the board.

We also detailed supplementary modules such as recognition and speech synthesis that were incorporated with our MT system to facilitate a prototype SL-to-spoken language translation system. While recognition requires further training, the addition of a speech synthesis module was successful and added functionality to the system.

[31] Note that human evaluation of the output transcriptions is a somewhat artificial exercise given that the evaluator must use their intuition and SL knowledge as to how accurate that transcription would be if signed by a human or avatar.

We concluded our set of experiments with the addition of an animation module to complete the English–ISL translation task and to demonstrate what a fully functioning system might look like. Using Poser animation software we manually created 50 signed sentences that were manually evaluated by 4 ISL monitors with positive results.

In the second set of experiments, carried out on our own purpose-built MediCorpus, we demonstrated that while SiGML and HamNoSys are perhaps a more acceptable way of representing the SLs, annotation proved the most successful for MT processing.

Following these experiments we included a discussion on the two main issues for SL MT as we see them, namely transcription and evaluation.

In terms of future avenues of investigation, the area of SL MT affords many options. For the most part, further development of current methods is a priority and that development should be carried out on both the MT as well as the output animations concurrently, given how intrinsic they are to each other. It is our intention to examine more closely the linguistics of the SLs we are working with[32] and, similar to the current trend towards more linguistic-driven statistical models of mainstream MT, we plan to investigate this path for SL MT. We believe that given the complexities of SL production, a linguistic approach is imperative.

In addition, the development of SL MT technology, particularly with the acquisition of data in multiple SLs, such as ISL and DGS corpora described in this paper, opens the doors for an SL–SL MT system, allowing translation between different SLs to bridge communication barriers within different Deaf communities. In short, given that SL MT is still a fledgling area compared to spoken language translation, there are plenty of possibilities for future development that could potentially provide us with a better understanding of SLs as translatable natural languages.

Acknowledgments We would like to express our gratitude to Kevin G. Mulqueen and Mary Duggan for their assistance in the creation of the ISL ATIS corpus, Suzanne Lindfield for her medical receptionist assistance, Shane Gilchrist for his work translating and signing the ISL Medical Dialogue corpus, Ronan Dunne for his HamNoSys transcription, Ege Karar and Horst Sieprath for their linguistic knowledge and input for German Sign Language and the four anonymous ISL monitors who performed manual evaluations on ATIS data. This research was part-funded by the Embark Initiative at the Irish Research Council for Science, Engineering and Technology (IRCSET; http://www.ircset.ie) and IBM (http://www-927.ibm.com/ibm/cas/sites/dublin/). Current 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

Bauer B, Nießen S, Heinz H (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

Bungeroth J, Stein D, Ney H, Morrissey S, Way A (2008) The ATIS Sign Language Corpus. In: Proceedings of the 3rd Workshop on the representation and processing of sign languages at the 6th international conference on Language Resources and Evaluation (LREC-08), Marrakech, Morocco

Cahill P, Carson-Berndsen J (2006) The Jess Blizzard Challenge 2006 Entry. In: Proceedings of the Blizzard Challenge 2006 Workshop, Interspeech 2006, Pittsburgh, PA

[32] In a similar fashion to the work on SL linguistics being carried out by Radboud University, Nijmegen as part of the SignSpeak project.

Chomsky N (2000) New horizons in the study of language and mind. Cambridge University Press, Cambridge

Conroy P (2006) Signing in & signing out: the education and employment experiences of deaf adults in Ireland. Research report, Irish Deaf Society, Dublin

Conway A, Veale T (1994) A linguistic approach to sign language synthesis. In: Cockton G, Draper SW, Weir GRS (eds) People and computers IX: Proceedings of the Human Computer Interface conference (HCI), Glasgow, Scotland, pp 211–222

Crasborn O, Zwitserlood I (2008) The Corpus NGT: an online corpus for professionals and laymen. In: Proceedings of the 3rd workshop on the representation and processing of sign languages at the 6th international conference on Language Resources and Evaluation (LREC-08), Marrakech, Morrocco, pp 44–49

Dandapat S, Morrissey S, Naskar SK, Somers H (2010) Statistically motivated example-based machine translation using translation memory. In: Proceedings of the 8th International Conference on Natural Language Processing (ICON 2010), Kharagpur, India

Dreuw P, Rybach D, Deselaers T, Zahedi M, Ney H (2007) Speech Recognition techniques for a sign language recognition system. In: Proceedings of Interspeech 2006, Antwerp, Belgium, pp 2513–2516

Dreuw P, Ney H, Martinez G, Crasborn O, Piater J, Miguel Moya J, Wheatley M (2010) The signspeak project—bridging the gap between signers and speakers. In: Proceedings of the international conference on language resources and evaluation, Valletta, Malta, pp 476–481

Fourie J (2006) The design of a generic signing avatar. Technical Report, University of Stellenbosch, South Africa

Grieve-Smith AB (1999) English to American sign language machine translation of weather reports. In: Proceedings of the Second High Desert Student Conference in Linguistics (HDSL2), Albuquerque, NM, pp 13–30

Hanke T (2002) ViSiCAST Deliverable D5-1: Interface Definitions. Manuscript, Hamburg, Germany

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), Lisbon, Portugal, pp 1–6

Hanke T, Storz J (2008) iLex—a database tool integrating sign language corpus linguistics and sign language lexicography. In: Proceedings of the 3rd Workshop on the representation and processing of sign languages at the 6th international conference on Language Resources and Evaluation (LREC-08), Marrakech, Morrocco

Hassan H, Ma Y, Way A (2007) MaTrEx: the DCU Machine Translation System for IWSLT 2007. In: Proceedings of the International Workshop on Spoken Language Translation (IWSLT), Trento, Italy, pp 69–75

Hemphill C, Godfrey J, Doddington G (1990) The ATIS Spoken Language Systems Pilot Corpus. In: Proceedings of the workshop on speech and natural language, Hidden Valley, PA, pp 96–101

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), Baltimore, MD, pp 65–74

Huenerfauth M (2006) Generating American sign language classifier predicates for English-to-ASL machine translation. PhD thesis, University of Pennsylvania, Philadelphia, PA

Jones D, Gill P (1998) Breaking down language barriers. Br Med J 316(7127):1280–1476

Kanthak S, Vilar D, Matusov E, Zens R, Ney H (2005) Novel reordering approaches in phrase-based statistical machine translation. In: Proceedings of the ACL workshop on building and using parallel texts at the 43rd annual meeting of the Association of Computational Linguistics (ACL-05), Ann Arbor, MI, pp 167–174

Kennaway J, Glauert J, Zwitserlood I (2007) Providing signed content on the Internet by synthesized animation. ACM Trans Comput-Hum Interact 14:1–29

Kneser R, Ney H (1995) Improved backing-off for n-gram language modelling. In: Proceedings of the IEEE international conference on acoustics, speech and signal processing, vol 1, Detroit, MI, pp 181–184

Koehn P, Och F, 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), Edmonton, Canada, pp 48–54

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, 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), Prague, Czech Republic
Leeson L (2001) Aspects of verbal valency in Irish sign language. PhD thesis, University of Dublin, Trinity College, Dublin, Ireland
Leeson L (2003) Sign language interpreters: agents of social change in Ireland?. In: Cronin M, ÓCuilleanáin C (eds) Languages of Ireland. Four Courts Press, Dublin pp 148–164
Leeson L, Saeed J, Macduff A, Byrne-Dunne D, 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, Carlow, Ireland
Leusch G, Ueffing N, Ney H (2006) CDER: efficient MT evaluation using block movements. In: Proceedings of the 11th European Chapter of the Association for Computational Linguistics (EACL-06), Trento, Italy, pp 241–248
Liddell S, Johnson RE (1989) American sign language: the phonological base. Sign Lang Stud 64:195–277
Ma Y, Tinsley J, Hassan H, Du J, Way A (2008) MaTrEx: the DCU system for IWSLT 2008. In: Proceedings of the International Workshop on Spoken Language Translation (IWSLT 2008), Honolulu, HI, pp 26–33
Marshall I, 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, 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, Sapporo, Japan, pp 113–116
Matusov E, Zens R, Vilar D, Mauser A, Popovic M, Ney H (2006) The RWTH machine translation system. In: Proceedings of the TC-STAR workshop on speech-to-speech translation, Barcelona, Spain, pp 31–36
Matusov E, Leusch G, Banchs RE, Bertoldi N, Dechelotte D, Federico M, Kolss M, Lee Y-S, Marino JB, Paulik M, Roukos S, Schwenk H, Ney H (2008) System combination for machine translation of spoken and written language. IEEE Trans Audio Speech Lang Process 16(7):1222–1237
Morrissey S (2008a) Assistive technology for deaf people: translating into and animating Irish sign language. In: Proceedings of the 11th International Conference on Computers Helping People with Special Needs (ICCHP) Young Researchers' Consortium, Linz, Austria
Morrissey S (2008b) Data-driven machine translation for sign languages. PhD thesis, Dublin City University, Dublin, Ireland
Morrissey S (2011) Assessing three representation methods for sign language machine translation and evaluation. In: Proceedings of the 15th annual meeting of the European Association for Machine Translation (EAMT 2011), Leuven, Belgium, pp 137–144
Morrissey S, Way A (2005) An example-based approach to translating sign language. In: Proceedings of the workshop in example-based machine translation (MT Summit X), Phuket, Thailand, pp 109–116
Morrissey S, Way A (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 the Languages and Resources Evaluation Conference (LREC 2006), Genoa, Italy, pp 91–98
Morrissey S, Way A (2007) Joining hands: developing a sign language machine translation system with and for the deaf community. In: Proceedings of the conference and workshop on assistive technologies for people with vision & hearing impairments (CVHI), Granada, Spain
Morrissey S, Way A, Cahill P, Carson-Berndsen J (2007a) A complete Irish sign language to speech translation system. In: IBM CASCON Dublin Symposium, Dublin, Ireland
Morrissey S, Way A, Stein D, Bungeroth J, Ney H (2007b) Combining data-driven MT systems for improved sign language translation. In: Proceedings of machine translation Summit XI, Copenhagen, Denmark, pp 329–336
Naqvi S (2007) End-user involvement in assistive technology design for the deaf–are artificial forms of sign language meeting the needs of the target audience?. In: Proceedings of the conference and workshop on assistive technologies for people with vision & hearing impairments (CVHI), Granada, Spain
Neidle C, Sclaroff S, Athitsos V (2001) SignStream^{TM}: a tool for linguistic and computer vision research on visual–gestural language data. Behav Res Methods Instrum Comput 33(3):311–320
Ó'Baoill D, Matthews PA (2000) The Irish Deaf Community (Volume 2): the structure of Irish sign language. The Linguistics Institute of Ireland, Dublin, Ireland

Och FJ, Ney H (2000) Improved statistical alignment models. In: The 38th annual meeting of the Association for Computation Linguistics (ACL-00), Hong Kong, China, pp 440–447

Och F, Ney H (2003) A systematic comparision of various statistical alignment models. Comput Linguist 29(1):19–51

Papineni K, Roukos S, Ward T, 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), Philadelphia, PA, pp 311–318

Penkale S, Haque R, Dandapat S, Banerjee P, Srivastava AK, Du J, Pecina P, Naskar SK, Forcada ML, Way A (2010) MATREX: The DCU MT System for WMT 2010. In: Proceedings of the joint fifth workshop on statistical machine Translation and MetricsMATR, Uppsala, Sweden, pp 149–154

Pierce J, Carroll J, Hamp E, Hays D, Hockett C, Oettinger A, Perlis A (1966) Language and machines: computers in translation and linguistics. Technical Report: Automatic Language Processing Committee, National Academy of Sciences, National Research Council. Washington, DC

Pinker S (1994) The language instinct: how the mind creates language. Penguin, London

Pizzuto E, Rossini P, 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), Genoa, Italy, pp 0–13

Prillwitz S, Leven R, Zienert H, Hanke T, Henning J (1989) HamNoSys, Version 2.0. Hamburg notation system for sign languages: an introductory guide. Hamburg Signum, Hamburg

Rabiner LR, Juang BH (1989) An introduction to hidden Markov models. IEEE ASSP Mag 4(1):4–16

Sáfár E, Marshall I (2002) The architecture of an English-Text-to-Sign-languages translation system. In: Proceedings of the international conference on Recent Advances in Natural Language Processing (RANLP-02), Tzigov Chark, Bulgaria, pp 223–228

San-Segundo R, Montero JM, Macias-Guarasa J, Córdoba R, Ferreiros J, Pardo JM (2007) Proposing a speech to gesture translation architecture for Spanish deaf people. J Vis Lang Comput 19:523–538

San-Segundo R, López V, Martín R, Sánchez D, 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, Valetta, Malta, pp 208–211

Shneiderman B (1998) Designing user interfaces. Addison Wesley, Menlo Park, CA

Somers H (2006) Language engineering and the pathway to healthcare: a user-oriented view. In: HLT/NA-ACL-06 medical speech translation, Proceedings of the Workshop, New York, NY, pp 32–39

Somers H, 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, Hungary, pp 41–49

Speers D (2001) Representation of American sign language for machine translation. PhD thesis, Georgetown University, Washington, DC

Stein D, Bungeroth J, Ney H (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), Oslo, Norway, pp 169–177

Stein D, Dreuw P, Ney H, Morrissey S, Way A (2007) Hand in hand: automatic sign language to English translation. In: Proceedings of the 11th international conference on Theoretical and Methodological Issues in Machine Translation (TMI-07), Skövde, Sweden, pp 214–220

Stein D, Schmidt C, Ney H (2010) Sign language machine translation overkill. In: Federico M, Lane I, Paul M, Yvon F (eds) International workshop on spoken language translation, Paris, France, pp 337–344

Stein D, Schmidt C, Ney H (2012) Analysis, preparation and optimization of statistical sign language machine translation. Mach Translat 26:325–357

Stokoe WC (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

Stokoe WC (1972) Semiotics and human sign languages. Mouton & Co. N.V. Publishers, The Hague

Stroppa N, Way A (2006) MaTrEx: DCU Machine translation system for IWSLT 2006. In: Proceedings of the International Workshop on Spoken Language Translation (IWSLT), Kyoto, Japan, pp 31–36

Sutton V (1995) Lessons in sign writing, textbook and workbook (2nd edn). The Center for Sutton Movement Writing Inc., La Jolla, CA

Tinsley J, Ma Y, Ozdowska S, Way A (2008) MaTrEx: the DCU MT system for WMT 2008. In: Proceedings of the third workshop on statistical machine translation, ACL 2008, Colombus, OH, pp 171–174

Traxler C (2000) The Stanford Achievement Test, 9th Edition: national norming and performance standards for deaf and hard-of-hearing students. J Deaf Stud Deaf Educ 5(4):337–348

van Slype G (1979) Critical methods for evaluating the quality of machine translation. Technical Report: BR-19142, European Commission Directorate General Scientific and Technical Information and Information Management, Bureau Marcel van Dijk, Brussells, Belgium

van Zijl L, Combrink A (2006) The South African sign language machine translation project: issues on non-manual sign generation. In: Proceedings of South African Institute for Computer Scientists and Information Technologists Conference (SAICSIT 06), Cape Winelands, South Africa, pp 127–134

van Zijl L, Olivrin G (2008) South African sign language assistive translation. In: Proceedings of IASTED international conference on assistive technologies, Baltimore, MD

Veale T, Conway A, Collins B (1998) The challenges of cross-modal translation: English to sign language translation in the Zardoz system. Mach Transl 13(1):81–106

Vilar D, Stein D, Huck M, Ney H (2010) Jane: open source hierarchical translation, extended with reordering and Lexicon models. In: ACL 2010 joint fifth workshop on statistical Machine Translation and Metrics (MATR), Uppsala, Sweden, pp 262–270

Ward W (1991) Understanding spontaneous speech: the phoenix system. In: Proceedings of the IEEE international conference on acoustics, speech and signal processing, Toronto, Canada, pp 365–367

Weaver W (1949) Recent contributions to the mathematical theory of communication. In: The mathematical theory of communication. The University of Illinois Press, Urbana IL, pp 94–117

Wu C-H, Su H-Y, Chiu Y-H, Lin C-H (2007) Transfer-based statistical translation of Taiwanese sign language using PCFG. In: ACM Transactions on Asian Language Information Processing (TALIP), vol 6, No. 1

Zens R, Ney H (2008) Improvements in dynamic programming beam search for phrase-based statistical machine translation. In: Proceedings of the International Workshop on Spoken Language Translation (IWSLT 08), Honolulu, Hawaii, pp 195–205

Zhao L, Kipper K, Schuler W, Vogler C, Badler N, Palmer M (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