Approaches to English to Sign Translation

Stephen Cox, Ian Marshall, and Éva Sáfár

School of Information Systems,
Norwich NR4 7TJ, United Kingdom
{sjc, im, es}@sys.uea.ac.uk

Abstract. We discuss the inherent difficulties of addressing the problems of translating English to Sign Language and outline approaches pursued in the ViSiCAST[1] project. The approaches of two subgroups of researchers in the project are presented. Firstly, an overview of the language-processing component of an English-text-to-sign-languages translation system is discussed focusing upon the inherent problems of the enterprise. Then, the architecture of the TESSA system is described. This system converts spoken language to a signed presentation for use within human transaction contexts.

1 Introduction

It is only since the mid 20th Century that sign languages have been recognised as 'natural' languages with their own phonology, morphology, syntax, semantics and pragmatics. Research over the past half century, however, has lead to greater recognition of sign languages and the Deaf communities in general[9, 10], and to an increased recognition as languages of interest to linguistic research[2, 18]. Arguably, it has also lead to a more sympathetic general public as witnessed in the UK by the popularity of British Sign Language courses amongst the hearing population. Nonetheless, there remains a shortage of sign language interpreters within the UK at a time when legislation and directives require greater provision of information and service access for deaf people[19].

Against this backdrop, research at the University of East Anglia, in conjunction with national and international collaborating organisations, investigates the potential of current computer natural language processing technology and computer graphics based 'virtual humans' for translating from spoken English or English text to sign language. The infancy of this enterprise is highlighted by the relative neglect of the reverse direction—sign language recognition and understanding—though we will briefly speculate on possibilities (section 3.2).

'Virtual human' (avatar) technology is reaching a stage where relatively realistic three dimensional characters can be generated with sufficient fidelity that

[1] *ViSiCAST* is an EU Framework V supported project which builds on work supported by the UK Independent Television Commission and Post Office. The project researches virtual signing technology in order to provide information access and services to Deaf people.

signed presentations are readable by skilled signers. The performance increases and the reduction in costs of this technology continue at a rate that one can expect increasingly realistic avatars to become commonplace.

Using this enabling technology, our research is concentrated on two approaches[5]. One is based upon use of motion capture technology, which permits the 'recorded' (motion captured) behaviour of a human signer to be stored and later replayed[4]. This forms the basis of the 'TESSA' system described in section 3. The other approach attempts to address the linguistic translation problems in a deeper and more fundamental way by constructing a semantic repesentation from which signing can be generated[16, 11]. This is discussed in section 2. In practice, these approaches are two extremes of a continuum of possibilities in which human motion may be analysed and used creatively in reconstructing new signed presentations.

2 Synthetic Generation

The aim of this research is the semi-automatic preparation of signed presentations from English texts, using a machine translation (MT) approach. Figure 1 illustrates the architecture of this approach for generation of synthetic signing[16, 11]. Research in computational linguistics and natural language processing has made significant progress during the past fifty years. However even the most effective automatic MT systems (e.g. the web-based 'Babelfish'[1]) produce target language output which is indicative of the source language content but below professional human translation quality[6]. Hence we assume an environment in which a human can intervene to volunteer information and correct automatic processing to enhance the quality of the final signed presentation.

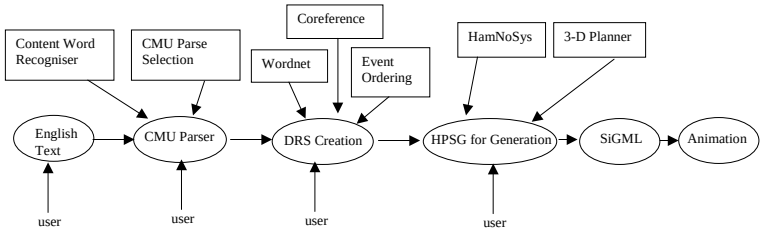

Fig. 1. Stages of English text translation to sign language

This approach seeks to exploit and extend natural language processing technology for English in order to generate an appropriate semantic representation from which sign language generation can begin. Discourse Representation Structures (DRSs)[7] are the basis for the underlying semantics. Other natural language processing components and techniques are integrated with each other to help construct the DRSs. To synthesise sign requires knowledge of sign language

syntax, morphology and phonology. The development of software to achieve this task is at the frontiers of sign language research, and poses a number of methodological and technical issues.

2.1 Methodological Issues

Sign language research (as with other minority languages) poses a number of methodological issues. The ideal situation would be the development of a technological environment in which native signers can develop lexicons and grammars and explore the use of notations for describing their own language. This reduces the design-implement-evaluate-revise cycle to as short a time period as possible, and it is possible that this will happen in the long term. However, the current situation is that use of notations (such as Head Driven Phrase Structure Grammar (HPSG) and HamNoSys—see below) to describe formal properties of sign language is a highly specialist activity. In addition, in the absence of large corpora stored in a suitable form to permit analysis of sign language, introspection leads to valuable insights.

Nonetheless, there remains a fundamental question of whose insights are to be sought and what value is to be placed upon them. Historically, signing research has frequently been carried out by hearing people using deaf informants and hence insights are typically second-hand. Additionally, the status of deaf informants themselves within the Deaf community raises a significant issue. Deaf people born to deaf parents are viewed as the genuine native signers who should act as informants, and who should be asked to identify the preferred manner of signing a proposition rather than what is merely acceptable signing[13]. Currently, we work within the limitations of using deaf informants with hearing researchers. Initial review is done by hearing signers, and more extensive review by deaf users of the generated signing to provide feedback and guide revision. Though this is far from ideal, it permits exploration of the use of the underlying formalisms prior to a more appropriate methodological framework.

2.2 Technical issues

The main obstacle to the development of a synthetic sign language generation software is the relative scarcity of research on the formal properties of signed languages. In our work, the phonological components of signing are described using an extension of *HamNoSys*[15]. HamNoSys is a notation for describing significant handshape, hand position, hand orientation and motion and our extensions include non-manual aspects of signing, especially facial expressions. Though HamNoSys has been used for analytical research of signers (particularly German signers), its use for such purposes for other sign languages and its use for synthesis has been much more restricted. Current developments using HamNoSys for synthesis are promising[8], though for lexicon development this requires significant expertise with the notation and attention to detail.

Fundamentally, this line of research requires development of an appropriate grammar to define well-formed signed sequences. Sign languages contain a

number of constructions which undermine the simplistic notion that signing is merely a process of concatenation of discrete events, each corresponding to an individual sign. For instance, the position of the eyebrows (neutral, raised or lowered) indicates either a declarative, a yes/no interrogative or a wh-interrogative proposition, respectively. Hence it is misleading to use motion-captured data from a declarative sentence as part of an interrogative sign sequence due to the potential misinformation conveyed in the facial expression. Another (and more subtle) example is given by directional (agreement) verbs. These require that one or more of subject/ object/ indirect object agree with the verb, either in terms of the start and end positions of the sign in signing space, or the handshape incorporated within the motion of the verb. For example, the sentence

```
He gave me the two red cups from the cupboard.
```

requires that the signing of 'give' starts at a position in signing space at which 'cupboard' is located and ends at the position associated with the speaker. In addition, it may optionally use a handshape within this movement that is appropriate for indicating a cup (at the very least it should not use a handshape inconsistent with the object given). Thus, appropriate signing involves parallel morphological components which is difficult to exploit with motion captured data.

Though these kinds of phenomena have been noted by sign linguists for a number of years, relatively few formal characterisations have been undertaken, with some notable exceptions[13, 3]. Hence a major undertaking within this work is the elicitation of precise descriptions of relevant phenomena to form the basis of computer programs. Currently we use a lexicalist HPSG [14] approach to formalising grammatical and lexical information for British Sign Language (BSL), as our German and Dutch colleagues are seeking to do for their own sign languages.

The HPSG implementation of a prototype BSL synthesis system has a lexicon of 50 signs, and a small number of grammar rules. However, these contain sufficient variation to allow investigation of a number of interesting sign linguistic phenomena within this framework.

3 TESSA—an Engineering Approach to Sign 'Translation'

The alternative approach to sign translation currently being pursued at UEA does not attempt to 'understand' the phrase presented for translation and to then synthesize it from a large number of primitive elements of sign language. Instead, the system has a database of complete phrases, stored together with their signed 'translations', and it looks up and replays the sequence of signs that represent the input phrase. These sign sequences are displayed by recording and storing the movements of a human signer using motion-capture technology, and then replaying these movements on an avatar[4].

This approach requires that the sign sequence representing the input phrase (or its semantic equivalent) has been pre-recorded and stored in the system's 'lexicon', which clearly imposes severe restrictions on the number of phrases that can be translated. However, the TESSA system described here (TESSA stands for TExt and Sign Support Agent) has been developed for a highly specific purpose, namely to aid a Post Office (PO) counter clerk in communication with a deaf customer. Many Post Office transactions can be accomplished using a small number of 'boilerplate' phrases, some of which need to be slightly modified to include variable quantities such as amounts of money, days of the week, postal destinations etc. By using a 'look-up' approach, we sidestep the difficult translation problems discussed in section 2. The advantage of this is that we are then able to investigate immediately matters such as the intelligibility, acceptability and usefulness to the deaf community of sign translation systems.

Because it has been designed to assist in transactions, the TESSA system translates speech rather than text to sign language. The PO clerk's speech is converted to text by speech recognition software, and the system then assembles the correct sequence of signs to translate the phrase uttered by the clerk.

3.1 Overview of the system

Figure 2 shows the structure of the TESSA system. The Post Office clerk wears

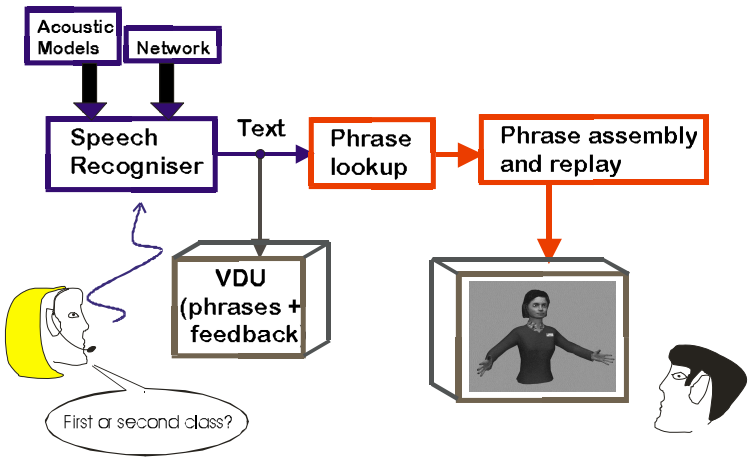

Fig. 2. The TESSA system

a headset microphone connected to the computer. The screen in front of the clerk displays a menu of topics available e.g. 'Postage', 'DVLA', 'Bill Payments ', 'Passports'. Speaking any of these words invokes another screen showing a list of phrases relevant to this category which can be recognised. However, this is only an *aide-memoire* to the clerk; all phrases can be recognised at any time, so

that switching between categories is seamless. The speech recognition software is always 'listening' for one of these phrases, and recognises the phrase when it is preceded and followed by a short pause. The result of the recognition output is fed to software that assembles the correct sequence of signs. In some cases, this is a single recorded sequence (e.g. 'Could you pop it on the scales, please?') and in other cases it is a sequence of signs that the software 'blends' together (e.g. 'That will be three pounds and thirty two pence, please', which requires four sequences to be joined together). Finally, the assembled sign sequence is sent to the avatar software and is replayed to the customer.

The phrases used The set of phrases used was chosen after examining transcripts of several hours of business at UK Post Offices. At the end of this analysis, a set of 115 phrases was prepared which we estimated would be adequate to cover about 90% of transactions performed. This set of phrases was changed and extended after trials with users and the total number of phrases currently available in the system is about 350.

The restriction to a small set of phrases has benefits for the input as well as the output of the system. By using a small set of phrases rather than the 50 000 or so words typically used in dictation software, we restrict the 'search space' for the speech recogniser and hence increase its accuracy. High recognition accuracy is very important for our system: the translation process is inherently slow because the avatar signs rather slowly to achieve maximum clarity, and any extra delay due to correcting mistakes made by the recogniser is likely to make the system unusable. Note also that, because there is no separation of speech and language decoding in this system, TESSA does not suffer from inaccuracies in the speech decoding process being forwarded to a language translation process that is also imperfect, an effect that can make more complex systems fail to translate correctly even quite simple phrases. By using pre-stored phrases, we in effect trade flexibility and range for accuracy.

The avatar and motion capture [2] The sign sequences are displayed by replaying 'motion-captured' data on on an avatar, rather than replaying video clips. There are several reasons for this:

- An avatar can be made to display any sequence of movements, and so offers the flexibility ultimately required for signing of unrestricted input text.
- Different figures and faces can be rendered onto an avatar's frame, so that a single set of recordings of signs can be used to drive different virtual humans.
- Conversely, multiple human signers can be used to generate the signed content of the system while using the same avatar for the output signing, making it easy to expand and update the signed content.
- Concatenation of signing is more fluent and controlled for an avatar than for video signing, as the exact positioning of the avatar can be easily manipulated.

[2] The avatar and motion capture software have been developed by Televirtual Ltd. (www.televirtual.com)

The movements of the signer are captured using a set of special sensors:

1. 'Cybergloves' are used to record finger and thumb positions relative to the hand itself.
2. Magnetic sensors record the wrist, upper arm, head and upper torso positions in three-dimensional space relative to a magnetic field source.
3. Facial movements are captured using a helmet-mounted camera with infra-red filters and surrounded by infra-red light emitting diodes to illuminate Scotchlight reflectors stuck onto the face. Typically 18 reflectors are placed in regions of interest such as the mouth and eyebrows.

Figure 3 shows this configuration in use. The outputs from the sensors are fed

Fig. 3. Data capture: face tracking camera with facial reflectors, Cybergloves for tracking the digits and Polhemus sensors taped onto the back of each hand, upper arm, body and head to track the body.

into a computer and processed by software into a form in which they can be recorded as a data file. When it is desired to replay the captured movements, another software package reads the data file and projects the recorded movements onto a wire-frame torso and face. The torso can be 'clothed' in any way desired and a large number of different faces are available.

3.2 Future Development of TESSA

The system described here is the first stage towards a more sophisticated system which will incorporate techniques used in 'speech-understanding' systems to enable a much wider range of transactions to be completed. In our current research system, we are experimenting with a speech recogniser that is not constrained to recognise only the phrases in the system, but can recognise any utterance, albeit with lower accuracy. The text output by this recogniser is fed to a language processor that decides what the appropriate pre-stored phrase is, if one is available. This has the benefit of allowing the clerk complete flexibility in what he or she says to the recogniser (as long as the words used are within the $\tilde{5}0\ 000$ word vocabulary of the recogniser) at the expense of requiring some language 'understanding' to determine the correct sequence of signs to be output. At time of writing, we do not know whether this system will be less accurate than the system that uses a network.

At present, TESSA is a one-way communication system and cannot respond to signs made back to her. However, we have begun to build a system that is capable of recognising a very limited number of signs. This system uses a camera to record the motions of a human signer, and software which firstly attempts to segment the image of the signer into its components (torso, arms, hands, fingers etc.) and then to find the sign that is the most likely origin of this sequence of movements of the limbs. In order to teach such a system about the range and variability of the movements associated with signs, we use a database of signs made by several different signers.

4 Conclusions

The two different approaches outlined above seek to address some common themes. The TESSA system aims to evaluate the quality of the technology supporting motion captured data and playback though an avatar decoupled from syntactic and morphological formulations of sign language. Conversely, the English-text to sign language approach requires a more sophisticated approach than a simple concatenation of motion captured signs will permit. Nonetheless, these are really two positions along a continuum. It would be possible to compare the user acceptability of purely synthetic sign by producing comparable HamNoSys descriptions of the motion captured data. Similarly it is possible to use motion-captured data to endow a synthetic signing based avatar with more human like motion.

Throughout the project, evaluation by the deaf community has played and continues to play a vital role in evaluating and improving the intelligibility and quality of signing, and the usefulness of the systems.

References

1. http://world.altavista.com

2. Brien,D. (Ed.): Dictionary of British Sign Language/English. London,Boston. 1992
3. Cormier,K.A.: Grammatical and Anaphoric Agreement in American Sign Language. Graduate School of the University of Texas at Austin. Master Thesis. 1998
4. Cox,S.J. et al.: The Development and Evaluation of a Speech to Sign Translation System to Assist Transactions. Submitted to International Journal of Human Computer Interaction 2001
5. Elliott,R., Glauert,J.R.W., Kennaway,J.R., Marshall,I.: The Development of Language Processing Support for the ViSiCAST Project. In: Assets 2000. 4th International ACM SIGCAPH Conference on Assistive Technologies. New York. 2000
6. Hutchins,W.J., Machine translation and human translation: in competition or in complementation? In: International Journal of Translation, vol.13, no.1-2. 2001
7. Kamp,H., Reyle,U.: From Discourse to Logic. Introduction to Model theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory. Kluwer Academic Publishers. 1993
8. Kennaway,J.R.: Synthetic Animation of Deaf Signing Gestures. In: The Fourth International Workshop on Gesture and Sign Language Interaction (GW2001). City University, London, UK. 2001
9. Klima E. and Bellugi U., *The signs of language*, Harvard University Press. 1979
10. Kyle J.G., Woll B., Sign Language: The Study of Deaf People and their Language, CUP. 1985
11. Marshall,I., Safar, E., Extraction of Semantic Representations from Syntactic CMU Link Grammar linkages. In: Recent Advances in Natural Language Processing (RANLP), G. Angelova et al (ed), Tzigov Chark Bulgaria, ISBN 954-90906-1-2. 2001
12. Miller,G.A., Beckwith,R., Fellbaum,Ch., Gross,D., Miller,K.: An Introduction to WordNet. An On-line Lexical Database. http://www.cogsci.princton.edu/~wn/. 1993
13. Neidle C, Kegl J, MacLaughlin D, Bahan B, Lee R.G.: The Syntax of American Sign Language. MIT Press. 2000
14. Pollard, C., Sag,I.A.: Head-Driven Phrase Structure Grammar. The University of Chicago Press, Chicago. 1994
15. Prillwitz,S., Leven,R., Zienert,H., Hanke,T., Henning,J., others: Hamburg Notation System for Sign Languages—An Introductory Guide. International Studies on Sign Language and the Communication of the Deaf, Volume 5. Institute of German Sign Language and Communication of the Deaf, University of Hamburg. 1989
16. Safar,E., Marshall,I.: Translation of English Text to a DRS-based Sign Language Oriented Semantic Representation. In: Conference sur le Traitment Automatique des Langues Naturelles (TALN) vol 2, pp297-306. 2001
17. Sleator,D., Temperley,D.: Parsing English with a Link Grammar. Carnegie Mellon University Computer Science technical report CMU-CS-91-196. 1991
18. Sutton-Spence,R., Woll,B.: The Linguistics of British Sign Language. An Introduction. University Press, Cambridge. 1999
19. Woll, B.: Language, culture and identity: insights from sign language and the deaf community. In: The Linguist, August/September 2001.