Analysis of Notation Systems for Machine Translation of Sign Languages

Jessica Jeanne Hutchinson
Department of Computer Science
Rhodes University
Grahamstown, South Africa 6139
Email: g0h2318@campus.ru.ac.za

***Abstract*—This project analyses sign language notation systems; what systems exists, what data is currently available, and which of them are best suited for machine translation purposes. Machine translation is aided by a textual representation, but no standard representation exists for sign languages. This research analyses notation systems and builds a corpus of parallel American Sign Language (ASL) and English data, assessing the limitations involved in building it and the feasibility of using such a corpus for machine translation. SignWriting proved to be a good representation with the most advantages for a translation system.**

***Index Terms*—Natural Language Processing, Statistical Learning, Communication Symbols**

I. Introduction

Sign languages as languages have not been studied as extensively as spoken languages, and there is still much to be learned about them. There are few standards for sign languages, both in terms of the languages used by signers within regions and dialogue groups, and in terms of the notations with which sign languages are represented in written form. The second factor affects the first: sign languages remain localised to small regions, as there is little technology that allows widespread use of a language or dialect. A standard textual representation of sign languages would provide an easier communication medium to sign language users and if more communication is happening across regions, language use and dialects also become more standardised [1].

A number of systems for translation between signed and spoken languages have been built or prototyped, with a range of focuses and using a variety of techniques, some successful and some not yet successful, all with their advantages and disadvantages. The current goal for this area of research would be to get to that point where the question of whether translation can be achieved no longer needs to be asked and where the best, most efficient and scalable techniques for translation are known.

This project aims to look into the field of sign language translation in terms of notation systems; what systems exists, what data is currently available, and which of them might be best suited for machine translation purposes. The question being asked is how using a textual representation of signs could aid machine translation, and which notation would best suit the task. To do this, the research will provide an analysis of the major notation systems with a focus on their suitability for translation by machine. It will also involve building a basic translation system from available data in order to practically test the limitations and strengths of that notation. The results can then be built upon and used to improve future translation systems.

A background into the three main fields will be provided; sign language linguistics, machine translation, and the combination: machine translation of sign languages. The research will then provide an analysis of notation systems and go into describing the process of constructing a corpus for translation, with reference to practical experience translating ASL data with the Moses translation system.

II. Background

Sign languages have not been studied as extensively as spoken languages, and the field of research is still fairly young [5]. A good understanding of how sign languages are used is necessary for creating a good translation system. Sign languages are vastly different from spoken languages; some of the methods used in spoken language translation/recognition systems can be applied, but not all of them are suitable. Sign languages are not all the same; each one will have differences in syntax, lexicon, etc. However, there are concepts referring to the nature of the production of signs that are unique to sign languages and that apply to all sign languages. What is known about sign language universals must be considered and used to adapt translation systems.

A. Sign Language

Sign languages exist in a visual medium and are three dimensional; spoken languages exist in an audio medium and are two dimensional. Sign languages function within both space and time, therefore requiring specific techniques for processing, and posing interesting challenges not encountered in the translation/recognition of spoken languages.

Furthermore, sign languages use the medium of time differently to spoken languages; signed utterances make use of both simultaneous and sequential processes to distinguish signs. The Movement-Hold model developed by Scott Liddell and Robert Johnson describes signs according to their sequential nature, as a series of movements and holds[9]. A sign may be described by a series of movements ending in a hold

(MMMH), or moving from a hold to a hold (HMH), etc. The phonetic components of signs, however, can occur simultaneously as well as sequentially.

There will always be some variation in the way that signs are performed, whether by a different signer or the same signer. Sometimes these differences may have meaningful components, such as making a sign smaller or bigger to indicate the size of the object being spoken of or to emphasise the word, etc. but even between signs with exactly the same meaning performed by the same signer, there are still statistical variations in the motion that is made [16].

B. Machine Translation

Machine translation is one of the oldest fields in computing and translation of spoken language has made significant progress by this point in time, but translation of sign languages remains in the early stages of development.

There are two main approaches to building machine translation systems: the rule-based approach, and the empirical approach. Early systems for spoken language translation were all rule-based; advancing research in artificial intelligence led to the incorporation of empirical methods[6]. The hybrid approach, combining empirical methods with a few grammatical rules, has proved to be the most successful[3].

The rule-based approach is not easily scalable or applicable beyond the language pair for which it is created, as rules and grammars need to be constructed specifically for that language pair and more rules need to be added when the language domain gets extended. The empirical approach was therefore taken in this research.

Empirical systems use data (examples) rather than predefined rules and can be either example-based or statistical[3]. Example-based translation compares the sentence/phrase to be translated with a knowledge base of previously translated sentences/phrases, building the translation from what it has encountered before. Statistical translation works with probabilities, using a corpus of parallel sentences to assign probability weights to words and phrases, and then using those weights to translate sentences/phrases[12].

An empirical system can be transferred from one language pair to another. The system can be reconstructed for another language pair by building a parallel corpus for that language pair and retraining the models. The problem that arises is the fact that large corpora for sign languages do not exist. TAUS labs recommends 800 000 sentence pairs as a minimum for a good translation system, and 1.2 million or more aligned sentences for difficult language pairs [3].

C. Translation of Sign Languages

With machine translation of sign languages there are essentially two steps to the process: the translation component, and either recognition or synthesis depending on the direction of translation. Speech recognition software functions within the same language but between two media (voice and text); language translation software, while creating a mapping from one language to another, usually functions within the same medium (text to text). A complete sign language translation would need to include both a natural language processing component and a computer vision component for recognition/synthesis of visual signs, functioning between two media (video and text) as well as two languages (the signed language and the spoken language). This process can be optimised if the translation occurs text to text, and the synthesis or recognition component functions separately. For this we need a textual representation.

III. Notation Analysis

There are a number of number of systems that can be used to represent signs in textual form, but there are no standards. Using a notation for sign language translation enables easy distribution of data and makes the translation system scalable. Current systems focus on limited domains, such as one topic domain [15] [12]. Use of a notation in an intermediary step may aid in making systems scalable to wider language domains, although it may depend on the notation used.

Each new project conducted by different researchers starts the entire process from the first step, gathering data for translation. It is difficult to share data between research projects due to the fact that each project builds its corpus with a particular research focus, using a particular language pair. This is further complicated by the lack of standards within the field. No standard notation means that each research project must decide on a notation system for that project. Gloss is the most common system used, but it has its own limitations. If a standard notation system emerges, it can aid in data sharing, and new projects can build on older projects, thereby enabling the projects and the field to progress more rapidly.

A. Stokoe

The Stokoe system was developed by William Stokoe in 1960[9]. It is phonetically based, and was the first notation to be produced for sign languages; most other notations are based on Stokoe's work[11]. Stokoe introduced the concept of segmenting signs into phones or phonemes, described with the following aspects:

- **Hand configuration**: determined by the active hand, and denoted designator (dez)
- **Place of articulation**: denoted tabula (tab)
- **Movement**: the action of the sign, denoted signation (sig)

The development of the Stokoe system advanced the field of sign linguistics significantly, but it is now outdated. The Stokoe system is seen as inadequate to represent sign language, as it has no way of indicating nonmanual features such as facial expressions. The system was created as a tool for linguists to describe signs, but is not necessarily useful for anything else and indeed is no longer used for its primary purpose, due to its lack of nonmanual feature representation. It is primarily a hand-written system, although an ASCII form has been developed [10]. This ASCII system is described, but if any electronic data using the ASCII system has been created, it was not found. Stokoe's system remains largely a hand-written system.

B. Gloss

Gloss notation represents a sign using a word stem from a spoken language[2]. Using gloss notation, one can describe signs with any degree of detail that one requires. This can be particularly useful for describing nonmanual components, emphasis, classifier predicates, etc. However, this also means that data described using gloss notation is variable. There are no standards for transcribing data using gloss and without guidelines each dataset will be transcribed differently if transcribed by different people. This has implications for statistical translation; greater variability will affect the weightings assigned by the translation system.

One advantage of using gloss for machine translation is that the transcribed signs will already be in words from the target language, making it easier to map between the two languages. However, this is only true for that particular language pair. To have this close match using the same sign language and a different spoken language would require the data to be transcribed again. It is possible to translate between sign and spoken language when the sign language data is transcribed with a different spoken language gloss, but a close match provides a more accurate and efficient translation[12].

C. Hamburg Notation System

The Hamburg Notation System (HamNoSys) is a phonetically based notation system that was developed by Siegmund Prillwitz in 1984[4]. This system, like most representation systems, was initially handwritten, but a machine readable font is available from the University of Hamburg[1]. An XML encoding of HamNoSys called Signing Gesture Markup Language (SiGML) is also available. It was developed for the ViSiCast project by Richard Kennaway[7].

Some advantages of this system are that it is international (can be used for any sign language), it is iconic, it is adaptable, has a formal syntax and can be stored in a computer database[4]. However, it does not provide any easy way to describe non-manual features, such as facial expressions. This notation was developed for a linguistic description of signs, not to be used in any form of communication[4]. This translation system has successfully been used for previous translation systems, but data was not easy to find when conducting this research.

D. SignWriting

SignWriting (SW) is a system developed by Valerie Sutton for communication purposes rather than linguistic purposes. The goal of SignWriting is to enable signers to be literate in their first language, not requiring them to learn another language in order to read and write[14]. SignWriting can be used for any sign language.

SW is a pictorial notation system and can describe non-manual features. It makes use of a set of symbols that can be combined to describe any sign. Though standardisation efforts are being put into place, the system is still flexible; if a language cannot describe a sign with the available symbols, it is possible to add more symbols to the set (with restriction).

The script is used extensively both on computer and on paper. A markup language exists to describe texts written in SignWriting (SignWriting Markup Language, or SWML). There are also a number of ways to represent individual signs that are machine readable. Stephen Slavinski has developed the Sign Writing Image Server (SWIS), which encodes the SignWriting script using a mathematical encoding known as Modern SignWriting (MSW)[13].

E. Summary of Notations

These are only a sample of the notation systems that have been and are being developed to describe sign languages. Each script was developed with a particular purpose in mind. The glossing system and the earlier scripts, Stokoe and HamNoSys, were developed to describe signs linguistically. SignWriting is the only one that was developed as a script for signers themselves to use as a writing system. The scripts that focus on describing signs linguistically can aid machine translation by the fact that they include relevant information. The scripts that focus on providing a written form for sign languages are also useful for machine translation, as they will describe signs enough to distinguish meaning. The major benefit of these representations is that if they are being used by signers then signs are already written down, reducing the enormity of the task of getting data into a transcribed form before using it to build translation systems.

With this analysis it can be concluded that the best notation system for translation is one that signers themselves are using. This rules out the Stokoe system completely. Glossing is problematic due to its lack of standards and its restriction of language pairing. There are other representations besides these main ones, but they have been developed for a particular project and are not used beyond that project.

It is not the aim of this research to decide on which notation system should be used by all signers and sign language research. Each script has its advantages and disadvantages, and most are being continuously changed and improved upon. The lack of existing data in certain scripts is seen as a disadvantage to machine translation, and it *is* the aim of this research to provide a recommendation on scripts to use for machine translation. However, the result will need to be continuously re-evaluated. Currently the most widely used representation by signers themselves is SignWriting. However, this is subject to change. If re-analysis of transcription systems results in a different system being recommended, then it is always possible to create data using that system, thereby providing existing data. However it would be most beneficial for all systems to focus on one representation, and to share their data.

SignWriting has proved to be the most accessible notation format for this research. There are a number of reasons to use SignWriting, one of them being that corpora are not only being added to by many people for various research reasons, but a corpus already exists. Moreover, SignWriting

[1] http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/hamnosys-97.html

is not only being used by linguists or computer scientists conducting research on sign languages, it is being used by signers themselves. SignWriting may have its limitations, but it is the most advanced and widely-used system that exists.

IV. Corpus Construction

In order to assess usability of notations with experiential as well as theoretical reference, a basic machine translation system was implemented. The task involved gathering textual data, building a parallel corpus, cleaning and preparing the data, and running it through the translation system. Complete success was not expected; the goal of this process was to gain experiential knowledge of the limitations of different notations.

A. Data Requirements

Transcribing data is time-consuming. The 595 sentences used in [12] took one person three months to transcribe using gloss notation. Most notations require familiarity with the system in order to use it. This means that building a translation system from scratch is no easy task and requires a great deal of work. The researcher needs to familiarise him or herself with the script as well as with the sign language, gather data, take the time to transcribe that data, build the data into a corpus, clean and prepare the data for translation, build the system to be used for translation (which will involve another fair number of tasks), run the corpus through the system, and finally analyse the results.

It would be more time-efficient if data from previous projects were accessible and usable by other researchers so that new research does not need to start from the ground up each time. More time to focus on other areas of the process means that more progress can be made. New systems need to build on what exists; the fact that the data has to be produced each time for specific research projects by the researchers themselves means that it very often restricts the focus of the project. This is particularly limited by systems that use gloss notation, as the same data would need to be re-transcribed into the relevant spoken language in order to be reusable.

If there is more than one reason to build a corpus, it is more likely to be built. A corpus that is constructed for one research project with a particular research focus might not have any relevance to other research projects, and therefore getting more people involved in the corpus construction is difficult. However, if the corpus is reusable, and constructed with a broader aim, the corpus can grow much more quickly with contributions from other people's research.

B. Acquiring Data

From the SignWriting website it was possible to find a large portion of the Bible translated into ASL and transcribed in SignWriting, easily downloadable from the ASL Bible SignPuddle[2]. This was the largest available dataset, and provides verse aligned parallel data.

[2] http://www.signbank.org/signpuddle2.0/export.php?ui=1&sgn=28

C. Data Cleaning and Preparation

The ASL Bible corpus was easy to obtain, and consists of an XML file with verses from the Bible encoded in MSW, aligned with an English translation. The data was extracted from the XML, aligned on a sentence level, and placed into two plain-text files in UTF-8 encoding. Sentence aligned data was required by the machine translation system used for this research. The ASL Bible is aligned by Bible verses, which may be complete sentences, but may also be fragments or may be more than one sentence on one line. This is not ideal for translation. Scripts have been written to automatically align parallel data by sentence using statistical methods, but these scripts are language specific; an ASL/English script would need to be written. The data was sentence aligned as well as possible, by removing any lines that contained sentence fragments or too many sentences in one line.

Ideally, the corpus should be cleaned better by going through it line by line, looking at both the English and the ASL and removing any unwanted data. If the corpus is very large, this is not possible. The corpus used in this research is not large, but there were further limitations due to time constraints and the fact that the researcher was not familiar with the source language. Errors will occur in corpus data; the best that can be done is to minimise the errors as far as possible.

The original XML file was 87950; after extracting and cleaning, the resulting corpus has 4275 parallel lines in English and ASL. Though this is more than the number of sentences used in previous translation systems (Morrissey used less than 600 [12]), it falls short of the recommended number: 800 000 aligned sentences for a good translation system, and upwards from 1.2 million aligned sentences for difficult language pairs[3].

D. Language model

For the language model, a monolingual corpus was built of English sentences. These sentences were also taken from the Bible, so that the language domain was consistent. This means that the monolingual corpus is also verse rather than sentence aligned. The data for this was acquired from the World English Bible[3].

E. Machine Translation

The machine translation system used in this research is Moses[8], which uses statistical methods to create a translation model from a sentence aligned parallel corpus of two different languages. The Do Moses Yourself (DoMY)[4] program was used, which automates the Moses installation process on Ubuntu 64-bit, and provides scripts which are wrappers around the Moses scripts. DoMY does no processing of its own; all the processing is done by the Moses system.

The translation system cleans the English data using predefined scripts, builds a language model from the parallel corpus,

[3] http://ebible.org/web/

[4] http://www.precisiontranslationtools.com

divides the corpus into a training set and a testing set, and then trains a model which is then used on the testing set for translation. The results are evaluated with a BLEU score.

The resulting BLEU score was fairly low, at 0.0845. On a second run the BLEU score was 0.0924, which is still fairly low. The translation output includes a number of words that are not translated, most likely due to words that only occurred once in the corpus and fell into the testing set, therefore did not have a mapping assigned to them. This would affect the BLEU score. Apart from the words that were not translated, however, the resulting sentences were readable, and even fairly close to the expected translation. The notation is adequate, the corpus however is not; if the corpus were improved it is likely that a much better BLEU score would be obtained.

V. Conclusions and Future Work

This research has focused on the intermediary step between sign recognition and spoken language synthesis, analysing sign language notation systems and how they can be used to optimise the translation process. It was found that SignWriting provides a good notation for translation; other systems could successfully be used for translation, but the conclusion from this research is that SignWriting would allow translation systems to progress more rapidly than other notations have.

The advantages of the SignWriting system are, in summary: it describes all the relevant features of signs, it is accessible to native signers, it is universal and is widely used, it has a machine readable format which is continuously being improved upon, data in SignWriting is easily available, and the available data is growing.

A parallel corpus of SignWriting data was constructed, and was then cleaned and translated using Moses, a statistical translation tool. The results obtained were fair and showed that the system works, though it requires improvement. The limitations were found to be the small amount of available data, the suitability of the data for accurate translation, and the availability (or lack of availability) of tools for cleaning SignWriting data. However, though there is currently only a limited amount of data available, it is growing, aided by SignWriting's ease of use and the fact that it is indeed used by a number of people for a variety of applications. The cleaning tools created for this project can be extended to be more generalised to other projects. Overall, the translation was not too difficult to put together; this is a positive result for future systems that can potentially get good translation results by building on this research.

There is still much to be done and great scope for improvement within every step of the translation of sign languages. A large corpus of data is essential, and future work can look at building such a corpus and standardising the format of the data that is gathered. This research has not included sign recognition to or sign synthesis from notation systems; further research could build a system which maps features extracted from video to SignWriting or another notation. Statistical translation involves a number of other processes for which scripts can be created to automate the process, for example: word segmenters, sentence aligners, etc.

References

[1] Crasborn, O., Mesch, J., Waters, D., Nonhebel, A., van der Kooij, E., Woll, B., and Bergman, B. Sharing sign language data online: Experiences from the echo project. *International journal of corpus linguistics 12*, 4 (2007), 535–562.

[2] Dreuw, P., Rybach, D., Deselaers, T., Zahedi, M., and Ney, H. Speech recognition techniques for a sign language recognition system. In *Eighth Annual Conference of the International Speech Communication Association* (2007).

[3] Haddow, B. Principles of machine translation. Online, July 2012.

[4] Hanke, T. Hamnosysrepresenting sign language data in language resources and language processing contexts. In *Workshop on the Representation and Processing of Sign Languages on the occasion of the Fourth International Conference on Language Resources and Evaluation, Lisbon, Portugal* (2004).

[5] Hoiting, N., and Slobin, D. Transcription as a tool for understanding: The berkeley transcription system for sign language research (bts). *Directions in sign language acquisition* (2002), 55–75.

[6] Hutchins, W. J. Machine translation. In *Encyclopedia of Computer Science*. John Wiley and Sons Ltd., Chichester, UK, pp. 1059–1066.

[7] Kennaway, R. Synthetic animation of deaf signing gestures. *Gesture and Sign Language in Human-Computer Interaction* (2002), 149–174.

[8] Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., et al. Moses: Open source toolkit for statistical machine translation. In *Annual meeting-association for computational linguistics* (2007), vol. 45, p. 2.

[9] Liddell, S., and Johnson, R. *American sign language: The phonological base*. Gallaudet University Press, Washington. DC, 1989.

[10] Mandel, M. Ascii-stokoe notation: A computer-writeable transliteration system for stokoe notation of american sign language, 1983.

[11] Martin, J. About the stokoe system.

[12] Morrissey, S. *Data-driven machine translation for sign languages*. PhD thesis, Dublin City University, 2008.

[13] Slevinski, S. E. Modern SignWriting. Online, February 2012. Available from: https://github.com/Slevinski/msw.

[14] Sutton, V. Why write sign language?

[15] Verlinden, M., Tijsseling, C., and Frowein, H. A signing avatar on the www. *Gesture and Sign Language in Human-Computer Interaction* (2002), 175–199.

[16] Vogler, C., and Metaxas, D. Asl recognition based on a coupling between hmms and 3d motion analysis. In *Computer Vision, 1998. Sixth International Conference on* (1998), IEEE, pp. 363–369.