Literature Review: Analysis of Sign Language Notations for Parsing in Machine Translation of SASL

Submitted in partial fulfilment
of the requirements of the degree of

Bachelor of Science (Honours)

of Rhodes University

Jessica Hutchinson

Grahamstown, South Africa
May 28, 2012

Introduction

Machine translation between Sign Language (SL) and spoken language occurs from sign to text, and text to sign, with most systems focusing on one or the other. In Sign-To-Text translation, there is an intermediate step between recognition of Sign Language and the translation of that Sign Language into text in which the Sign Language needs to be represented by some notation. However, there is as yet no standard notation for sign languages. A few of the major forms of representations that exist are: HamNoSys, Stokoe's notation, SignWriting and GLOSS notation. Most existing projects in the area of machine translation of SL focus on recognition of small datasets and do not give much consideration to the output annotation, nor the process of machine translation of the output in a manner that would scale to large systems.

Challenges in Translating Sign Languages

Sign Languages as Distinct From Spoken Languages

Sign Languages are not merely signed forms of spoken languages, as they are commonly perceived to be, but are complete natural languages with complex linguistic features [6]. The lack of a one-to-one relationship between a signed language and a spoken language is one of the challenges faced by Sign-To-Text or Text-To-Sign systems, because it means that there are essentially two steps to the process: recognition and 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 (i.e. text to text). Sign Language recognition would need to include both recognition and translation, functioning between two media (video and text) as well as two languages (the signed language and the spoken language).

Another challenge is that, unlike spoken languages, there are no large corpora available for Sign Languages, especially for South African Sign Language (SASL). This is partly due to the lack of a standard notation for Sign Languages; with no standard method of annotation, if data were to be collected it would not necessarily be usable. This lack of usable data means that methods such as example-based translation, which work very well for speech recognition, cannot so easily be used in Sign Language recognition/translation.

Phonological Structure

Research has been conducted in the advantages of phonemic modelling of signs for translation. Languages are rule-based systems [10]; using rules, an infinite number of phrases can be built from a finite number of words; an infinite number of words can likewise be built from a finite number of components known as phonemes [13]. In spoken languages, a phoneme is the smallest contrastive unit of sound; in Sign Language a similar concept exists (also denoted phoneme), with a visual rather than auditory contrastive component. A description of the concept of a phoneme in Sign Languages is given by Christian Vogler in [11]. Vogler also explains that a phonological approach to translation is more scalable than an approach which looks at whole signs or entire phrases at a time. The important idea here is that the number of phonemes is finite, whereas the number of possible words is infinite [3]. The challenge in such an approach lies in the fact that Sign Language Phonology is not well defined and has not been widely researched [11]. Getting a computer to recognise a phoneme requires a clearer understanding of what exactly constitutes a phoneme in a Sign Language.

Vogler [11] goes on to state that a better approach to Sign Language Translation would be on a phonetic level, rather than a phonological level. A phonological unit is related to meaning; a phoneme is "the smallest unit that can distinguish morphemes (units of meaning)" [11]. Phonetics, in comparison, describes the physical structure of a sign, not only the features with contrastive meaning. Either way, it is important to note that breaking down signs into smaller units aids in the scalability of the translation process. A dictionary-like approach, basing recognition on whole words or phrases of signs is not scalable, whereas using a finite set of smaller units for recognition is [13].

The Simultaneous Nature of Sign Language

Speech recognition systems can model speech as a sequential process, but Sign Languages are both simultaneous and sequential [3]. This increases the complexity of the process of sign recognition. The two possible solutions given by Vogler [11] to this problem are to

- model the dependencies and interrelationships between features
- decouple the simultaneous events

Vogler researches the second option, decoupling the events by dividing them into different channels.

Classifier Predicates

Sign Languages make use of three-dimensional signing space and are visual-gestural, in contrast to spoken languages which are oral-aural [8]. This gives rise to what is known as classifier predicates; classifiers are morphological units which indicate the class to which an item belongs, i.e. they classify that item. Classifier predicates in sign languages function pronominally, i.e. the classified item will function as a pronoun [1]. Pronouns in Sign Languages are positioned as objects within the signing space that the signer can refer to; the term “classifier predicate” refers to this pronominal construct which a signer can manipulate to indicate various meanings such as size, location, movement, and other properties [7]. This adds greater complexity to the translation process.

Notations

There are a number of notations that can be used to represent Sign Languages, but there is as yet no standard notation [6]. As stated before, a representation is necessary for the intermediate step between recognition and translation from Sign To Text. Some of the existing notations are briefly discussed below.

Stokoe

The Stokoe system, like HamNoSys, is phonetically based. Ceveloped by William Stokoe, it was the first notation to be developed for Sign Language. Most other notations are based on Stokoe's work[1].

Stokoe describes signs using what he calls 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)

[1] http://www.signwriting.org/forums/linguistics/ling006.html

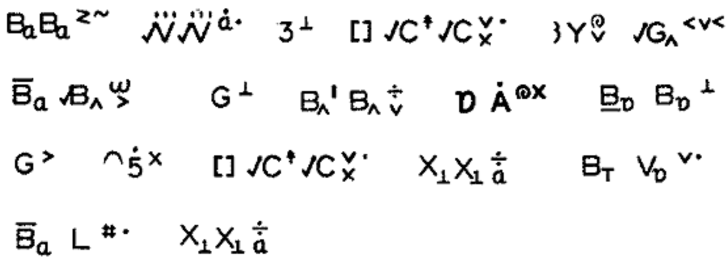

Figure 1: An Example of the Stokoe system - an extract from Goldilocks and the Three Bears

HamNoSys

The Hamburg Notation System (HamNoSys) is a phonetically based system that was developed by Siegmund Prillwitz. 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. Also, this notation was developed for a linguistic description of signs, not to be used in any form of communication [4].

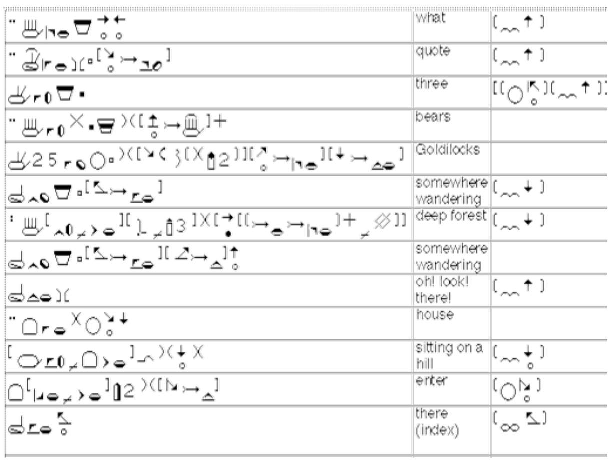

Figure 2: An Example of HamNoSys using the same passage

SignWriting

SignWriting (SW) is a system developed by Valerie Sutton, a dancer who first developed Sutton DanceWriting to annotate movement[2]. SignWriting was developed for communication purposes rather than linguistic purposes. It is a pictorial notation system and can describe non-manual features. It was also developed to be appropriate for any sign language.

HamNoSys and the Stokoe system have been criticized for being impractical due to their technical format, whereas SignWriting is considered by some to be a more readable notation [8].

Figure 3: An Example of SignWriting, with the same passage

GLOSS

GLOSS notation represents a sign using a word from English (or some other language) [2]. The implication of this notation is that, if used in a translation system, it would require the translation system to take a dictionary-based approach.

Parsing - Visual Language Parser

The notation that gets chosen for the recognition process is likely to affect the type of parser that will be used to process that notation. Helm et al [5] discuss the use of

[2]http://www.signwriting.org/forums/linguistics/ling004.html

visual language parsers for diagrams and handwritten text. As mentioned before, text or spoken language is organised sequentially. Paper [5] makes the statement that, though it is possible to sequentialise input, there is "no one natural order in which people draw a complex diagram." This is also true for Sign Language. The paper proposes parsing visual languages in sets rather than sequential data, which might be a plausible approach for parsing Sign Language, particularly SignWriting notation.

The Process of Sign-To-Text Translation

There are two sides of the translation process - translating from Sign Language to English text (Sign-To-Text or STT), and translation from English text to Sign Language (Text-To-Sign or TTS). Most existing systems focus on translating from text to sign, rather than from sign to text. STT systems are mostly small-scale, functioning on limited datasets and focusing on the image processing component.

Machine Translation

Machine Translation can occur on three different levels, described in [6] and depicted in Figure 1, also taken from [6].

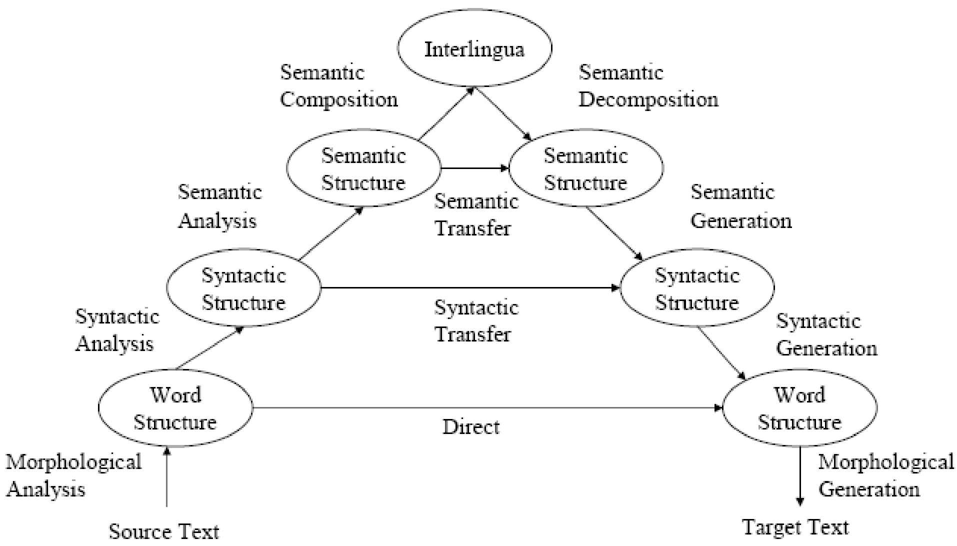

Figure 4: Machine Translation Architectures

Translation can occur on any of the three levels, the depth of linguistic information increasing with each level. The three different levels - direct, transfer and interlingua - require different levels of analysis of the source language, as shown by the labels on the left of the figure (i.e. morphological, syntactic, and semantic analysis), as well as different levels in the process of synthesizing the target language [6].

Text To Sign Systems

Some of the major systems that translate English text into American Sign Language (ASL) are discussed in [6] and briefly summarised below. Text-to-sign systems do not form part of the focus of this research, however, and are mentioned here as background only, therefore not with much depth.

TEAM

This system makes use of a synchronous lexicalised Tree Adjoining Grammar (TAG), storing pairs of words and signs in a dictionary. TEAM translates using a transfer architecture on a syntactic level.

ZARDOZ

This system's approach to translation is to use a language-independent event schema for translation, which is essentially an interlingua, the highest translation level in the pyramid (see Figure 1). The higher the translation level, however, the more restricted the domain of the translation system.

ViSiCAST

The ViSiCAST system describes two major stages of TTS: semantic analysis of English text, producing some semantic-based representation, and translation from this to a representation that can be used to render an animation [9]. The tools used in this process are: a CMU Link Parser, Discourse Representation Theory, and a Head-Driven Phrase Structure Grammar.

ASL Workbench

This system uses a lexical-functional grammar and a sophisticated phonological model, but requires user input at most stages of translation.

A phonological approach to translation would function on the direct level, falling under morphological analysis, unless further analysis is applied.

Sign To Text

The process of Sign To Text (STT) translation is depicted in Figure 2, taken from [11]

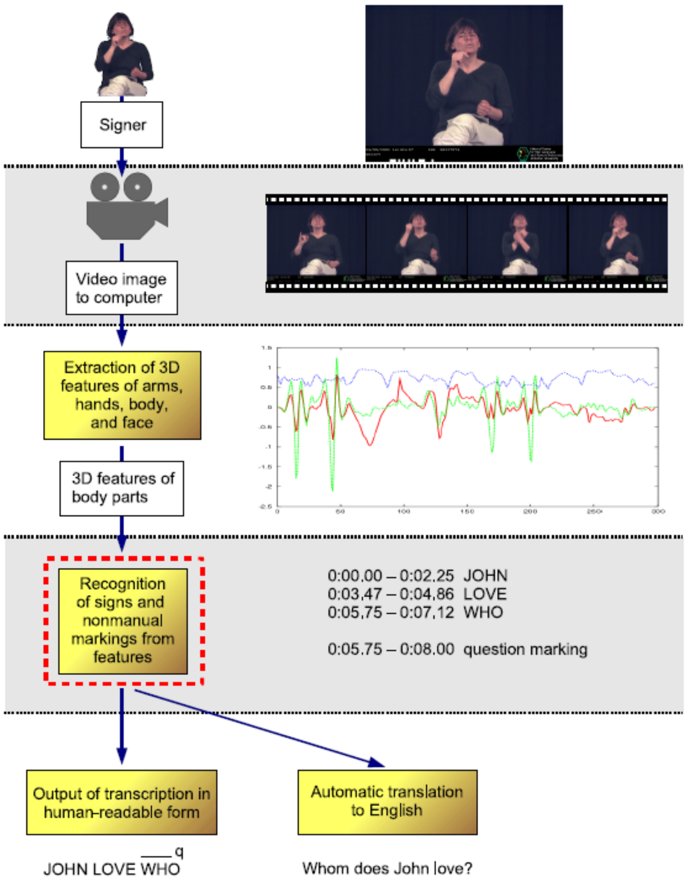

Figure 5: Stages of STT Translation Process

The five stages are:
(1) and (2) Getting video data from the signer and feeding it into the computer
(3) Feature Analysis
(4) Recognition
(5) Translation

Between stages four and five a decision needs to be made about what notation or model will be used to represent the recognised sign features. Natural Language Processing (NLP) techniques can then be applied to this output notation to translate it into text.

Previous work in the area of STT translation has focused mostly on feature recognition using Hidden Markov Models (HMMs). HMMs allow segmentation of the data stream, and have been used successfully for feature recognition [13]. Christian Vogler has done research in the analysis of features based on a phonemic model [12, 13, 11, 14]. He suggests using Parallel Hidden Markov Models (PaHMMs) and multiple channels to adequately recognise phonemes. His research states that the use of Parallel HMMs will enable translation systems to scale to larger and more complex systems by using a separate HMM for each phoneme [13].

Conclusion

A notation system for representing Sign Languages is necessary in the process of Sign-To-Text translation. The different existing notation systems should be analysed with regards to their use with a phonemic model, and their usefulness with regards to the process of translation as a whole, especially with regards to scalability. Research into different techniques for parsing these notations should also be considered.

Bibliography

[1] Crystal, D. *Dictionary of Linguistics and Phonetics.* The Language Library. John Wiley & Sons, 2011.

[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] Efthimiou, E., Fotinea, S., and Sapountzaki, G. Feature-based natural language processing for gsl synthesis. *Sign Language &# 38; Linguistics 10*, 1 (2007), 3–23.

[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] Helm, R., Marruitt, K., and Odersky, M. Building visual language parsers. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology* (1991), ACM, pp. 105–112.

[6] Huenerfauth, M. A survey and critique of american sign language natural language generation and machine translation systems. Tech. rep., Technical Report MS-CIS-03-32, Computer and Information Science, University of Pennsylvania, 2003.

[7] Huenerfauth, M. *Generating american sign language classifier predicates for english-to-asl machine translation.* PhD thesis, University of Pennsylvania, 2006.

[8] Moemedi, K. *Rendering an Avatar from Sign Writing Notation for Sign Language Animation.* PhD thesis, University of the Western Cape, 2010.

[9] Sáfár, É., and Marshall, I. The architecture of an english-text-to-sign-languages translation system.

[10] Valli, C., and Lucas, C. *Linguistics of American sign language: An introduction.* Gallaudet Univ Pr, 2000.

[11] Vogler, C. American sign language recognition: reducing the complexity of the task with phoneme-based modeling and parallel hidden markov models.

[12] Vogler, C., and Metaxas, D. Parallel hidden markov models for american sign language recognition. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* (1999), vol. 1, IEEE, pp. 116–122.

[13] Vogler, C., and Metaxas, D. Toward scalability in asl recognition: Breaking down signs into phonemes. *Gesture-Based Communication in Human-Computer Interaction* (1999), 211–224.

[14] Vogler, C., and Metaxas, D. Handshapes and movements: Multiple-channel american sign language recognition. *Gesture-Based Communication in Human-Computer Interaction* (2004), 431–432.