Brigham Young University
BYU ScholarsArchive

All Theses and Dissertations

2015-06-01

English to ASL Gloss Machine Translation

Mary Elizabeth Bonham
Brigham Young University - Provo

Follow this and additional works at: http://scholarsarchive.byu.edu/etd

Part of the Linguistics Commons

BYU ScholarsArchive Citation
Bonham, Mary Elizabeth, "English to ASL Gloss Machine Translation" (2015). *All Theses and Dissertations.* Paper 5478.

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu.

English to ASL Gloss Machine Translation

Mary Elizabeth Jouett Bonham

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

Deryle Lonsdale, Chair
Alan Melby
Norman Roberts

Department of Linguistics and English Language

Brigham Young University

June 2015

Copyright © 2015 Mary Elizabeth Jouett Bonham

All Rights Reserved

ABSTRACT

English to ASL Gloss Machine Translation

Mary Elizabeth Jouett Bonham
Department of Linguistics and English Language, BYU
Master of Arts

Low-resource languages, including sign languages, are a challenge for machine translation research. Given the lack of parallel corpora, current researchers must be content with a small parallel corpus in a narrow domain for training a system. For this thesis, we obtained a small parallel corpus of English text and American Sign Language gloss from The Church of Jesus Christ of Latter-day Saints. We cleaned the corpus by loading it into an open-source translation memory tool, where we removed computer markup language and split the large chunks of text into sentences and phrases, creating a total of 14,247 sentence pairs. We randomly partitioned the corpus into three sections: 70% for a training set, 10% for a development set, and 20% for a test set. After downloading and installing the open-source Moses toolkit, we went through several iterations of training, translating, and evaluating the system. The final evaluation on unseen data yielded a state-of-the-art score for a low-resource language.

Keywords: machine translation, ASL, sign language gloss

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation and deepest thanks for having Dr. Deryle Lonsdale as my thesis chair. His enthusiasm for this thesis, his expertise with the subject matter, his willingness to share his knowledge, and countless hours of his assistance as I went through the ups and downs of a Master's thesis are just a few of the reasons I was able to complete this thesis. I am also grateful for my other committee members, Dr. Alan Melby and Dr. Norman Roberts, for their valuable feedback and support. I am especially thankful for the parallel corpus received from Stephen Richardson, head of the translation group of The Church of Jesus Christ of Latter-day Saints.

My time at BYU has been joyful. I have learned so much from professors and fellow students and have felt buoyed by their encouragement and kind words. I want to express my appreciation for them. I wish to also express sincere thanks for LoriAnne Spear and the Linguistics office staff, who were always there to answer my questions; LoriAnne was a wealth of knowledge, encouragement and action on the path to the fruition of my thesis.

So many of my friends have encouraged me, for whom I am grateful. A very special thanks for my dear friend, Judy Francisco, who literally talked me through this experience, giving me confidence to stay the course.

A heartfelt thank you for my wonderful husband, Bob, who has been the 'wind beneath my wings' throughout our marriage and has been my rock during these thesis years. Thank you to our three daughters, Dianna, Heidi and Kelli, for being my personal cheerleaders and never losing faith that I could succeed. Thanks to my son-in-law, Todd, for his support in helping me whenever and with whatever I needed. Thanks to our six grandsons for their love, support and understanding when grandma was busy doing 'homework.'

In all of this, I will be eternally grateful to my Heavenly Father for guiding me to BYU, allowing me to partake of this wonderful opportunity, and always answering my prayers.

TABLE OF CONTENTS

LIST OF TABLES vi

LIST OF FIGURES vii

Chapter 1: Introduction 1

Chapter 2 - Literature Review 3

2.1. Machine translation 3

2.2. Sign Languages and Glossing 6

2.2.1. LDS Church translation and ASL gloss 13

2.3. Sign Language and Machine Translation 16

2.3.1 SMT and The Church of Jesus Christ of Latter-day Saints 21

Chapter 3 – Methodology 23

3.1. Toolkit installation and use 25

3.2. System refinement 27

3.3. Final evaluation 28

Chapter 4 – Results and Evaluation 32

4.1. System evaluations 32

4.2 Development and dev test phase 34

4.3 Final evaluation 37

4.4 Analysis of sample translations 38

Chapter 5 – Conclusions and Discussion 45

5.1. Future work 45

References 47

LIST OF TABLES

Table 1: Corpus partitions for various spoken language text/sign language gloss MT projects .. 26

Table 2: Baseline scores on the training set (seen data; 9,992 sentence pairs)..... 33

Table 3: Evaluation scores on dev and training+dev sets 37

Table 4: Final evaluation of unseen data of the test set (2,855 sentence pairs)..... 37

LIST OF FIGURES

Figure 2: ASL linguistic features that should be found in the target language..... 12

Figure 3: Examples of sentences in both English/ASL gloss 13

Figure 4: List of glossing techniques used by the LDS Church translation group 15

Figure 5: Series of procedures implemented in this thesis 23

Figure 6: Source and target sentences with hand-inserted SPLIT markers 24

Figure 7: MT and evaluation steps for the current thesis..... 29

Figure 8: Procedure for training and evaluation of the thesis..... 30

Chapter 1: Introduction

Translating from one language to another has become ever more important with increased communication across many languages worldwide. People frequently want to disseminate or assimilate information in their own language and in other languages. The traditional translation industry does not have the means to satisfy the vast needs for translation throughout the world. In an attempt to find a way to translate more material quicker and easier, researchers have developed computer machine translation systems, which have become tools to aid the translation process (Craciunescu et al. 2008).

There are several methods of machine translation (MT), one being statistical machine translation (SMT), the avenue in which the current thesis will proceed. SMT is performed using high volumes of sentences in a parallel corpus with which to train a computer program to translate sentences from the one language—the source language (SL)—to another language, the target language (TL) (Koehn 2005; Lopez 2008). Though researchers have been working with MT for many decades using text of spoken languages, sign language MT is less known. Sign languages do not have a widely accepted written form in which to create a parallel corpus, thus sign languages were largely excluded from MT research until the twenty-first century. Today, many researchers worldwide are investigating sign language MT (SLMT) using specialized textual glosses for representing sign languages.

The purpose of the current thesis is to develop and evaluate an SLMT system that translates from English text to American Sign Language (ASL) gloss.

For this thesis, our parallel corpus data was provided by The Church of Jesus Christ of Latter-day Saints (the LDS Church). We curated the raw corpus by removing the computer

markup language and then aligning the corpus sentence-by-sentence. The result is apparently the largest parallel corpus of text/sign language gloss to date, having over 14,000 sentence pairs. We developed an SMT engine using the Moses open-source toolkit (Koehn et al. 2007). We used the aligned phrases and words to create language models and execute the Moses decoder for the translation stage. To assess the quality of the MT output from the system, we used the state-of-the-art Bilingual Evaluation Understudy (BLEU) scoring protocol.

This thesis is organized as follows. Chapter 2 will be a review of general machine translation with a focus on: statistical machine translation, sign languages, American Sign Language, sign language notation systems, sign language and machine translation, and ASL MT for the LDS Church. Chapter 3 will describe the methodology used for the current thesis. Chapter 4 will present and discuss the results of the thesis. Chapter 5 will discuss the limitations of the thesis and possible future work.

Chapter 2 - Literature Review

Translation between languages is not a new phenomenon, but with development and continued expansion of the World Wide Web, the need for translation between languages has grown. Human translators can be expensive and there are not enough translators to satisfy all translation needs (Craciunescu et al. 2008). It would be impossible to translate all the information available on the Web into even one language, let alone many of them. To aid in the translation process, researchers have developed ways of automating translation. Computer machine translation is one such approach.

2.1. Machine translation

Machine translation (MT) is automatic translation from one language to another by computer. Researchers have been studying the process of machine translation for over 6o years. Today, machine translation paradigms include (Koehn 2010):

- rule-based (RBMT) which uses extensive handwork of human language experts in writing and applying linguistic rules for the translation process in the MT system (Morrissey 2008) .
- example-based (EBMT) which tries to match the input sentence with similar phrases and sentences in the corpus, retrieving their translation for the output.
- phrase-based (PBMT) which translates small sequences of words or phrases.
- statistical (SMT) which analyzes huge amounts of a parallel corpus and finds the highest probability of any given translation.
- hybrid MT which is a combination of multiple MT approaches.

The time for MT is now, claims Way (2013). He states that, depending on the requirements for the outcome, MT output can be used solely, or 'raw,' or it can be integrated with either minor or major human editing. He lists a number of people and companies using successful MT, including Stephen Richardson, head of the translation department of The Church of Jesus Christ of Latter-day Saints (henceforth LDS Church).

When MT is available for a particular language and particular domain, it can cut the working time of a translator by 50% (Richardson 2012). Human translators frequently express concern that a machine that can translate between languages will replace them; in reality, MT can provide a translator better tools, which would allow even more translation by increasing productivity and improving quality (Craciunescu et al. 2008).

Lopez (2008) describes how statistical machine translation (SMT) works with probabilities, using a learning algorithm that computes the probability that a specific word or phrase in the target language (TL) would be the best translation of a word or phrase in the source language (SL). The higher the probability, the better is the chance that the match will be correct. SMT was first attempted by IBM in the late 1980s, adding additional features successively, which created a foundation for further SMT research and development (Koehn 2010).

Researchers found that a large parallel corpus─or bitext─of source language sentences and human-translated target language sentences could be used to train statistical models for MT (Brown et al. 1990). As computers increased in capability of storage, so did the use of larger corpora. Brown et al. (1990) developed a bigram approach, where the computer searched for two contiguous words to see which words were most likely to appear together and matched them with the most frequent translation. They eventually developed a trigram model-based on three contiguous words─which provides even better translations than the bigram.

In early SMT research, it could take a researcher, if starting from scratch, nine months to develop MT algorithms to create a statistical MT system. Koehn and Hoang (2012) developed an open-source machine translation system called Moses. They explain that their system uses a parallel corpus to analyze and 'learn' the two languages to be used in the translation process. Moses is an SMT system and does not require the user to input translation rules about the languages. Through annual workshops, researchers worldwide contribute to the development of Moses, which has become a tool of choice for students and researchers. For details, see http://www.statmt.org/moses.

The current thesis focuses on development and evaluation of an SMT system based on the Moses toolkit to translate from English to American Sign Language (ASL) gloss. The current version of the system uses a trigram model.

Steps in SMT system development include training the Moses system using a parallel corpus in a specific limited domain. The parallel corpus provides data for Moses to create a source language model. The source language for this thesis is English. Moses creates an English language model from the parallel corpus, a model of correct, grammatical, fluent English in this domain. Moses also creates a target language model from the parallel corpus. The target language for this thesis is ASL gloss, and the model represents typical fluent ASL gloss in this domain. Moses also creates a translation model. The translation model shows us how English/ASL gloss should be translated. When parallel corpus lines are loaded into the system, Moses analyzes and learns models for the two languages, within the specific domain, and determines how the languages relate to each other. Use of narrower domains with limited vocabulary in the parallel corpus achieves better results in the output. Building language models with a larger monolingual corpus within the specific domain of the MT is also helpful (Corbett

2006). Moses uses the target language (TL) language model to try to simulate the TL when generating translation output. System parameters can be evaluated and readjusted to improve the output. A reordering model resequences the words in a source language (SL) phrase, word-for-word, and puts the words into their proper order for the TL. If words are reordered too much, though, it can result in TL unintelligibility.

For the current thesis, a quality human-translator-produced parallel corpus, or bitext, is of utmost importance in preparing the various statistical language models: the English language model, the ASL language model, and the English/ASL translation model. The current parallel corpus was created by translating from English to ASL gloss. The translation was done by a team of human translators who were skilled in both languages. When ten skilled human translators translate a sentence into a TL, the result could be ten different sentences (Koehn 2010). All the sentences would be considered good sentences, as there are more ways than one in which to translate a sentence and still retain the concept of the original sentence.

Most MT research deals with spoken languages and mention access to tens of millions of bitext lines of the SL and TL for statistical model development. As sign languages do not have large parallel corpora with which to work, researchers view them as low-resource languages. Working with MT for sign languages, most research is being done with only several hundred bitext lines. By contrast, this thesis involves tens of thousands of bitext lines of English/ASL gloss.

2.2. Sign Languages and Glossing

A sign language is a visual language. As trying to read lips, discern speech, and socialize with individuals who can hear and speak is a difficult, tedious, tiring, and often an impossible

task for many deaf people, sign languages have naturally developed throughout the world. There is no one universal sign language.

Some deaf people have difficulty reading and fully comprehending the majority written and/or spoken language, such as written American English. Often deaf people use sign language interpreters for important occasions where communication must be clear, but this is not always feasible. Due to the limitations in access to information, researchers in several different countries are investigating the possibility of using MT of sign language to aid communication within a variety of limited domains.

American Sign Language (ASL) is relatively new in the realm of world languages, having its beginnings less than 200 years ago. In the 1960s, a non-deaf English professor, William Stokoe, was hired by Gallaudet College, the first Liberal Arts college for the deaf. He was given a three-week crash course in 'the signs' the students used in the classroom. As Stokoe taught at the college, he began to recognize linguistic features in the signing used on campus. He posited that the signs the students used had grammar and syntax that were separate and unlike English, rather than just a representation of English on the hands. He created a sign language dictionary for ASL (Padden and Humphries 2005). ASL is now recognized as a full and natural language. Linguists consider ASL a non-IndoEuropean language[1]

Sign language poses a challenge for MT. Annotating all the facets of ASL is very time-consuming: facial expression, eye gaze, mouth movement, hand movement, hand shape, hand location, palm orientation, and other parameters require much time to represent orthographically (Morrissey 2008). I briefly describe other writing systems below to clarify why researchers of

---

[1] See www.ethnologue.com.

data-driven MT for sign language use glossing techniques (text) instead of writing systems currently in existence.

In 1960, Stokoe invented a writing notation for ASL called the Stokoe Notation System. This notation system indicated some of the parameters of the sign, including its location, hand shape, and movement. Since his invention, writing systems generally include the same three parameters; many systems include additional ones. The Stokoe Notation System was the first such system used in sign language linguistic academic research (Kyle and Woll 1988).

A team of hearing and deaf researchers in Hamburg, Germany developed another writing system in 1985 called the Hamburg Notation System (HamNoSys) (Prillwitz et al. 1989). It was also developed for research purposes, not for an everyday writing system. It consists of about two hundred symbols that could represent any sign language. Its symbols follow a standard in this order: Symmetry Operator, Non-manual Components, Handshape, Hand Position, Location, and Movement. It is difficult to notate the vital facial expressions (FE's), which are placed in a separate column below the notation symbols.

Sutton SignWriting (1981) is a system that uses drawings to show location, handshape, palm orientation and movement of a sign[2]. The script version can accommodate any sign language, has been acknowledged by the International Organization for Standardization (ISO) as a world script, and has been used in many countries.

Si5s is a writing system consisting of 80 characters, including the signing alphabet, for writing ASL[3]. The system, developed in 2003 by Robert Arnold, describes signs by using

---

[2] http://www.signwriting.org/

[3] http://www.si5s.org/ (8-5-2014)

specific drawings of handshapes, location, movement, palm orientation, and facial expressions.

Examples and brief explanations of various sign language writing systems can be found elsewhere[4]. Figure 1 shows samples of sign writing systems.

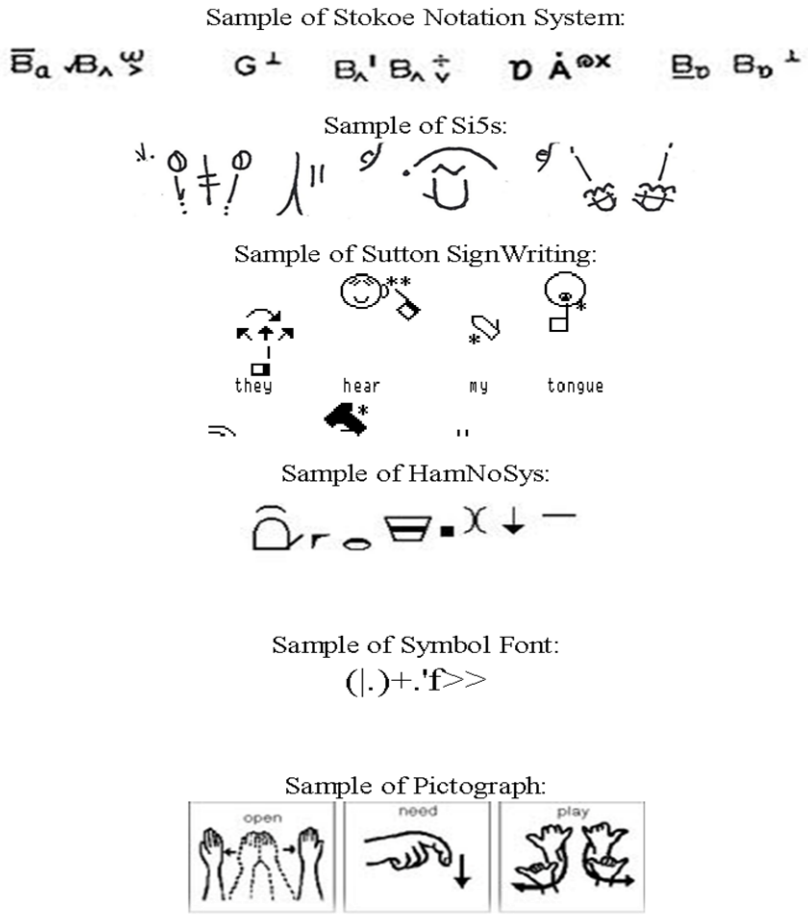

Figure 1: Samples of various sign writing systems for sign languages

[4] http://aslfont.github.io/Symbol-Font-For-ASL/ways-to-write.html.

Because of the difficulty in representing textually the various parameters of sign language, SLMT researchers use *glossing* systems. To *gloss* means using the written text of the spoken language to label and identify individual signs and concepts of the sign language.

Sign languages for English, Spanish, Italian, and other languages use glosses with their respective languages in approximately the same manner. For the purpose of the current thesis, I will discuss an English system called ASL gloss. To become proficient in ASL gloss, one would need to know ASL, English, and a variety of symbols that portray hand shape, location, movement of the signs, and non-manual markers, NMM's.

For transcription purposes, ASL gloss uses English words for each sign or phrase that can be labeled. ASL gloss is represented with small capital letters. The English word 'cat' would be transcribed as CAT in ASL gloss. The English declarative sentence 'The cat was bitten by the dog.' would be transcribed into ASL gloss as DOG BITE CAT.

An in-depth study of ASL linguistics gives insights on the many ways ASL is similar to and different from English (Valli and Lucas 2000). Figure 2 summarizes some basic ASL linguistic features. It includes examples in ASL gloss, which is explained later in this section. ASL gloss is written in small capital letters and distinguishes the gloss from the English text. A sematic concept can be expressed with one sign. 'What time is it?' = TIME? with appropriate facial expressions (FE's).

Unsurprisingly, English text to ASL gloss cannot be translated word-for-sign without loss of meaning (Valli and Lucas 2000). Ambiguity of an English word is addressed in the gloss by using the concept of the word (LDS Church 2012). For example, the English word 'will' can be used as a verb or noun, and each category has more than one sense. The sense of 'will' that means in the future is glossed as FUTURE. The sense of 'will' that means desire, is glossed WANT.

‘Want’ also has several senses, resulting in glosses such as NEED, POOR, TASTE, REQUIRE, etc. Conversely, the ASL gloss may differ in granularity from its corresponding English word. For example, the English sentence ‘My nose is running.’ is glossed as ‘MY NOSE CL:4 drip-from-nose.’

In gloss, transcription symbols express many types of linguistic information (Baker-Shenk and Cokely 1981):

- the topic of the sentence
- the type of sentence
- the hand-shapes
- the location of the signs
- fingerspelled signs
- compound or contraction signs
- emphasized signs
- repeated signs
- gesture signs
- signs made with one, both and/or alternating hands
- classifier signs
- miming concepts
- arc motion signs
- verb signs that move between objects
- the position of the eye gaze of the signer
- NMM’s show adverbial and adjectival information and provide important linguistic information to make the translation more clear

Fingerspelled words are represented with dashes, or preceded with a #. Fingerspelled loan signs are also preceded by a # symbol. Depicting verbs, or classifiers, are represented with a CL: and the hand shape that is used in the classifier. When a word is a compound word, it has a dash between the two signs, e.g., ‘really,’ and ‘really?’ as TRUE-BIZ. Depending on the NMM’s, this sign can be translated many different ways, as ‘actually,’ ‘Yes it did!’ ‘It did?’ and others. An English sentence of “It was great!” might be shown with the ASL gloss of FINE!

- One concept might generate many signs: 'weapons' = GUN, KNIFE, STICK, etc.
- Indicate tenses with a time marker. 'I went to the store yesterday.' = YESTERDAY, I GO. STORE.
- Discuss new concepts with lexical-visual paraphrases: Social Media = FB, TWITTER etc.
- ASL phonology involves the study of the parameters of handshape, location, palm orientation, movement and non-manual markers (NMM's) of the body and FE's.
- Plurals are shown by noun reduplication: 'brothers'= BROTHER, BROTHER .
- Reduplication also changes a verb to a noun: SIT (verb) – SIT,SIT (chair).
- Some numbers can be incorporated into the sign, called numeral incorporation: 5-WEEKS (signed as one sign.)
- Pronouns are shown by reference pointing, positioning them in 3-D space.
- When the handshapes perform an action, it is called a *classifier*: 'The car drove by last week.' = LAST WEEK CAR CL:3 *(drive-by*).
- Classifier predicates perform movement: 'The car drove by quickly.' = CAR CL:3 (*drive-by fast*).
- Aspect is shown while signing verb: an action can be once, continuous, intense, or drawn out.
- Derivational morphology is signed in a few ways: 'baker' = BAKE+AGENT, 'hotter' = MORE HOT, 'soften' = BECOME SOFT.
- Inflectional morphology is shown with movement of the sign: 'We talked for hours.' = CHAT-CHAT-CHAT.
- Past, present and future are shown in relation to the body: the body is NOW. PAST is over the shoulder.
- FUTURE is to the front and away from the body.
- Determiners can be shown by pointing or establishing the referent: 'that book' = BOOK CL:Y-hand>.
- Adjectives are placed after (and sometimes before) the noun: 'the red book' = BOOK RED or RED BOOK.
- Adverbs are made using non-manual markers (NMMs) and extending the sign: 'very tall' would be signed with an exaggeration of the sign TALL simultaneously performing specific NMMs.
- Auxiliary verbs, such as 'will' or 'should', are signed at the beginning and/or end of sentences.
- ASL has a few prepositions, but generally uses classifier predicates to describe them.
- Topicalization is at the beginning of the sentence SVO: FATHER LOVES CHILD.
- Topicalization can induce OSV with specific FE: CHILD, FATHER LOVES. ASL does not use passive voice: 'The cat was bitten by the dog.' = DOG BITE CAT.

Figure 2: ASL linguistic features that should be found in the target language

ASL gloss was not created for deaf people to use in everyday writing, and not all deaf people understand gloss. Therefore, gloss is not the end product in translation. Some sign

language researchers criticize gloss and express concern that using gloss will cause people to think the gloss is just a broken spoken language. Figure 3 shows examples of English/ASL gloss.

Declarative sentence:	"That is a beautiful cat."	//Cat-it\\ true-biz beautiful.
Yes-No Question:	"Is this your cat?"	//Cat-it\\ yours?
WH-Question:	"What is the name of your school?"	//Your school name\\ what?
RH-Question:	"We go to the temple to seal our family for eternity."	?We go temple for-for? Our family seal-circle always.

Figure 3: Examples of sentences in both English/ASL gloss

Though the use of gloss is standard for SLMT, glossing techniques have not been standardized. Researchers use many of the same basic gloss components, but each group of researchers must evaluate current glossing techniques and establish their own conventions for the gloss they use. Regardless of the concerns, SLMT researchers continue to use gloss because of convenience and ease of use.

Morrissey, a sign language machine translation (SLMT) researcher, asserts that it will take extensive effort to create and annotate an SLMT corpus. She suggested setting up a 'central repository with standards' (Morrissey 2011). She mentions current concerns with evaluation methods and contemplates how accurate MT output must be. Would a gist of the idea be enough? Would the output be helpful to translators? Her questions deserve careful consideration by MT researchers working on SLMT.

2.2.1. LDS Church translation and ASL gloss

ASL gloss helps document ASL in writing and is used academically for learning ASL grammar. However, there are times when a particular entity might use an adaptation of the generally accepted gloss format to meet its own needs. The LDS Church translation group is one

organization that extensively uses ASL gloss and has chosen to adapt it to meet specific needs. They have developed their own glossing technique that is a conglomeration of standardly accepted techniques. They have made custom adaptations mainly because their translation process includes the use of a teleprompter. The English message is translated into an ASL gloss, put into text files and loaded into a teleprompter, which will only accept keyboard symbols. A signer then reads the ASL gloss from the teleprompter screen, converts the message into ASL, and is filmed signing the message for dissemination.

It is critical that the signer, as he/she is watching the teleprompter, can quickly and easily read, understand and sign the transcription for filming. Even though typical ASL gloss is written in all capital letters, the translation department found that it is easier to read the transcription when it is in lower-case letters. For the ease of the signer, the gloss is created in lower-case letters, using capital letters at the beginning of sentences or for fingerspelled words. They also add punctuation such as periods, commas, and semi-colons to the gloss for signing clarity.

Each ASL translator is given an ASL Translation Guide (LDS Church 2012) to follow in the process of translating from English to ASL gloss. Figure 4 lists many of the transcription rules the English/ASL gloss translation department uses at the LDS Church.

The English/ASL corpus of LDS Church translation into ASL gloss contains translations done both prior to and after the establishment of the revised Translation Guidelines. Due to this discrepancy, existing translation data is not consistent.

For at least two decades the LDS Church translation department has used human translation of English/ASL in preparation for filming the LDS Church messages in ASL. Therefore, the translation department has hundreds of thousands of lines of English/ASL gloss parallel corpora. For the current project, we received approximately 20,000 lines of parallel

corpora, consisting of translations of several church magazine articles, translations of the *Preach My Gospel* book and translations of the 'Articles of Faith.'

- Although ASL gloss typically uses capital letters, the LDS Church uses lower-case letters.
- LDS ASL gloss does capitalize names, fingerspelled words, and honorific signs.
- Paired forward and backward slash marks delimit the topic of a sentence.
- Commands, imperative, and emphatic phrases begin and end with an '!'
- Signing 'finish' at conclusion of the sentence indicates past tense.
- The sign 'will' at the end of the sentence indicates future tense.
- The typical gloss for classifiers is CL:. The LDS Church shows a classifier with cl:.
- Directional and contracted signs are used when phrases of more than one English word can be signed with one sign. Help-you. Tell-me. Shout-to-him.
- Negative sentences are written with a '?' at the beginning of the affirmative sentence, followed by a '?' and then a wave-no.
- Yes/No questions and WhQ are shown with the topic and then the question and '?'.
- Rhetorical questions are two-part questions. The topic begins and ends with '?', followed by the answer to the rhetorical question. NMM's for RhQ are raised eyebrows.
- If a sign-gloss is ambiguous, translators add clarification. 'Run' would be glossed 'machine-run,' nose-run,' 'nylon-run,' 'run-for-office,' 'run-around,' etc.
- For pronominalization, the LDS Church does not use IX, which is used generally in glossing. Instead, the gloss would be simply 'he,' 'she,' or 'it.'
- Pluralization is done in several ways: child++, or car cl:in-a-line, or cookie cl:stack, or car#3, or children+group.
- Some NMM's are italicized and placed within parentheses. Example: //cat\\ (*cha)* fat! *Cha* is a mouth movement showing large size.
- Body shifts are glossed using arrows: > or < with slight shoulder shift following the arrows.
- rs signifies Role Shift – the signer becomes the character or signs a quote: *(rs:JS)*
- Deity is shown with a caret: ^He.
- Loan signs and name signs begin with '#' and are capitalized: Relief Society abbreviated #RS or a loan sign of #BANK.
- Compound signs are created with a +: true+work (which means 'really' in all its senses).
- Information to accompany certain signs is given in parentheses preceding the sign: (*1h*) one-handed sign, (*2h*) two-handed sign, (*alt*) alternate hands, (*not*) shake head while signing to negate the comment.
- Periods end a thought and are signed with a nod.
- Commas add a pause in the signing, with a slight nod. Commas replace 'and' or 'or'.

Figure 4: List of glossing techniques used by the LDS Church translation group

2.3. Sign Language and Machine Translation

Much of the research being done in sign language machine translation (SLMT) is based on Indo-European languages. SLMT research began approximately 21 years ago. One of the first attempts at an MT system involving sign language, the ZARDOZ system, (Veale and Conway 1994) was an MT prototype that would translate English text into three different sign languages: Irish Sign Language (ISL), American Sign Language (ASL) and Japanese Sign Language (JSL). The system generated output into a Doll Control Language program that controlled 'an on-screen animated doll' to sign the message for the deaf user.

Another early system for ASL MT of weather reports used Newkirk's (1989) Literal Orthography system. The corpus was tiny, just three lines from the National Weather Service forecast for January 20, 1999. The output was not analyzed by a native ASL signer (Grieve-Smith 1999).

In MT experiments of English to ASL 15 years ago, the researchers were adapting Stokoe notations of ASL for the computer (Speers 2001). The difficulty of learning this notation system has led many researchers to abandon this concept and to begin using a glossing system for MT. Speers (2001) introduced the *ASL Workbench* computer program prototype for his project. It used 'lexical functional transfer', which employs complex hand-engineered rules and requires a tremendous amount of human interaction.

Many researchers included computer avatars in their experiments. They annotate the MT gloss output into HamNoSys for generating a computer avatar using Sign Gesture Markup Language (SiGML), which controls an avatar that signs the translated message (Caridakis et al. 2005).

Bangham et al. (2000) experimented with MT and avatars, creating a very limited domain using speech of a post office clerk. After downloading a corpus of post office phrases, the program would find the exact phrase and apply it to an avatar that had been trained to sign the exact phrase word-for-sign in English (not ASL) word order.

Zhao et al. (2000) used glosses as the intermediate representation in their experiment with MT. They used Synchronous Tree-Adjoining Grammar to map English to ASL gloss, including linguistic, visual and spatial information. Their output fed into an avatar to sign the sentences.

Researchers in Spain developed SLMT systems for Spanish-to-LSE (Lengua de Signos Española) (San-Segundo et al. 2006). Looking for a specific, limited domain in which a deaf person might want an interpreter to assist with communication, they chose acquisition of a passport or other ID papers and collected 135 of the most common phrases used for that task. They expanded the corpus in 2010, collecting 1,360 Spanish sentences in the same domain. Two fluent LSE signers translated the sentences into LSE gloss (López-Ludeña et al. 2012). The researchers increased the Spanish sentences from 1,360 to 4,080 by using appropriate translation variations for each LSE sentence. The end result was a total of 820 LSE translation sentences for the 4,080 Spanish sentences (San-Segundo et al. 2010).

San-Segundo et al. (2012) divided the corpus into three parts: 75% of the corpus for training the language and translation models, 12.5% for development and 12.5% for testing the MT system. In this experiment, the researchers worked with 707 Spanish sentences in the driver's license renewal domain. The 707 Spanish sentences were then translated into LSE gloss. The researchers increased the corpus from 707 sentences to 2124 sentences by creating Spanish variations to the LSE sentences (San-Segundo et al. 2012).

López-Ludeña et al. (2013) also experimented with SMT. They developed a preprocessing module that replaced Spanish words with tags to improve the evaluation scores.

Porta et al. (2014) conducted a trial of a rule-based MT system for Spanish to LSE. They created a small corpus of 229 sentences, which they glossed using many of the same types of glossing techniques the LDS Church translation department uses. Porta et al. view automatic animation as a necessary step in MT. They state that an automatic evaluation can be used for evaluating the MT system output. However, measuring deaf people's comprehension of the animation of the MT output is necessary to have a true evaluation (Porta et al. 2014).

Weather reports are a popular domain for SLMT studies because the domain is small in size, making it easier to create a corpus by analyzing and annotating video segments of the signs made by the interpretation of the reports. Weather reports are what constitute the AEWLIS text (Lugaresi and DiEugenio 2013). AEWLIS (Atlas Extended Written LIS) is a custom serialization format in an Italian glossing system that produces LIS (Italian Sign Language or Lingua Italiana dei Segni) gloss from Italian text. The corpus size was 376 Italian/LIS sentences. The researchers converted the LIS into AEWLIS format. With such a small corpus, SMT would not be successful, so they used a rule-based method for this project.

From the RWTH Aachen University, Bungeroth and Ney (2004) received and used 1399 sentences in a parallel corpus with the DGS (Deutsche Gebärdensprache, German Sign Language) gloss. They tested their MT system on 200 German/DGS gloss sentences.

Forster et al. (2010) annotated their own sign language corpora with a form of gloss, adding linguistic information as a separate annotation for NMM's. They handle signs that are identical except for the change in meaning due to particular mouthing; they also handle synonyms, homonyms, compound glosses, and fingerspelling, give suggestions about how to

describe more features, such as referencing, classifiers, negation, confirmation, and incorporation, through the glossing of the signs.

Further research from RWTH produced an improvement to the DGS corpus (Forster et al. 2014). The vocabulary size is 1558 different glosses, which is an increase of 647 glosses.

Using the RWTH DGS Phoenix-Weather corpus and the NGT corpus, which is a Dutch Sign Language corpus of the Netherlands, the researchers then glossed the signed sentences, as these two corpora are not parallel corpora (Stein et al. 2012). The SL's for these experiments contained the DGS and the NGT video segments, which were translated by humans into a German gloss. The gloss is faster for MT, makes it easier to train the annotators, and can provide a larger corpus. The annotators placed some of the NMMs within the gloss through special tokens, as vital information would be lost without the NMMs.

Morrissey and Way (2005) discuss EBMT using a parallel corpus of English/NGT Dutch Sign Language text (with annotations) with 561 sentences from the ECHO project. A human evaluation judged output as *good, fair, poor* and *bad* (Morrissey and Way 2005). Unlike many of the articles in the field of SLMT, Morrissey (2008) discusses the importance of performing an automatic evaluation of the text output before programming an avatar, as the avatar process will inevitably bring with it additional errors. She also points out the lack of formal evaluation in many prior experiments with SLMT.

A prototype text-to-text system was created in Poland in 2002, with the goal of generating sign language from the translation (Suszczanska et al. 2002). Using the TGT-1, Text to Gesture Translator, a system created for translating Polish into Polish Sign Language, the researchers went through two phases: translating from text-to-text and then translating it into sign language.

A Tunisian group developed the WebSign system to broadcast important sports results in real time (Othman et al. 2010). It employs an avatar to use sign language by looking up words in a dictionary of SML (Sign Modeling Language) XML files, the results controlling the avatar.

Huenerfauth et al. (2007) performed research by producing a prototype MT ASL generator. In his article, Huenerfauth explains some difficulties with evaluation of ASL MT systems using automatic metrics. To run an MT system, researchers need a large parallel corpus. As a large English/ASL gloss corpus is not readily available, the corpus has to be created, which means a huge expense. Huenerfauth explains that researchers have concerns that, as ASL does not have an accepted written form, achieving a consistent gold standard candidate translation from human native signers for measuring against the MT output is not possible. In the Huenerfauth et al. study, a variety of signers, native and nonnative, produced the parallel corpus that was used in the MT experiment. With the inconsistency in human translation from English to ASL gloss, it remains to be seen if a gold standard was achieved in the translation portion of the corpus.

In 2011, Othman and Jemni developed an SMT system. They used a parallel corpus of 431 sentences, which consisted of 632 English words corresponding to 608 signs. They used a small 3-gram language model and IBM Models 1, 2, and 3 for alignment and string matching (Othman and Jemni 2011). With English as the input, ASL gloss was the output, which was then placed into the WebSign Tool for interpretation into an avatar. The authors claim the ‘best performing machine translation evaluated’ for English-to-ASL gloss. The article fails to mention the evaluation system for the output ASL gloss, or for the ASL generated by the avatar.

2.3.1 SMT and The Church of Jesus Christ of Latter-day Saints

The Church of Jesus Christ of Latter-day Saints (LDS Church) has an in-house translation group that translates over 100 million words annually with the help of human translators. It uses MT, with post-editing by human translators, in more than 20 languages, with several other languages in the planning stages to include in MT. Where MT is employed, the church produces up to 50% more translated text in the same amount of time it would take for a completely human translation. The LDS Church does not use MT for all Church materials, but would like to incorporate MT into many of the monthly magazine articles, and website translation tasks (Richardson 2014).

In 2012, the LDS Church began the process of using Microsoft Translator Hub[5] for translating English into 13 different languages. An additional 15 languages, which are not available on the Hub, were included in 2013. This was possible because the church continues to generate translation data (Richardson 2012). As the translation department has the need to translate many church materials into ASL, they have given us access to their substantial English/ASL gloss translations, which follow a modified version of standard ASL glossing standards. Our project involves developing an SMT engine for English/ASL gloss. Experts using their knowledge and intuition of English/ASL could then sign the TL output in ASL. Using MT to translate the English text SL into an ASL gloss TL shortens the process time of translation and increases the consistency of the translation.

Currently, the translation process for English to ASL gloss follows roughly the same process as the translation process for a spoken language. The English text to be translated is

---

[5] https://hub.microsofttranslator.com/SignIn?returnURL=%2FHome%2FIndex

assigned a team of trained translators, who are given a Translation Guide (TG) to follow. Much of the English text will be translated into ASL gloss in an OSV (object-subject-verb) word order. The TG explains how to translate time, location, topic, comment, tense, if-then statements, conditional clauses, when clauses, commands and imperatives, complex sentences, time sequences, negative sentences, and question structures; yes/no questions, the wh-questions and the rhetorical questions.

The translators render English SL sentences into ASL Gloss. The ASL gloss undergoes revision and refinement. When the translation supervisor approves the ASL gloss translation, an ASL expert, usually a deaf signer, is given the gloss to study and practice signing the message in ASL. A member of the translation team places the ASL gloss onto a teleprompter, and when the signer has prepared, they are brought into the filming studio at the Conference Center of the LDS Church. A film crew films the signer as they use the teleprompter to sign the message in ASL. The film is distributed online through an lds.org website at asl.lds.org[6]

In summary, researchers have worked with spoken language MT for several decades, with SLMT research beginning approximately 15 years ago. Lack of sufficient parallel corpora has kept all sign languages low-resource languages. Most SLMT researchers must create a small parallel corpus for their experiments. Because there is no standard writing system for sign language, SLMT researchers use glossing systems for the sign language portion of the parallel corpus. The current thesis is to attempt SLMT using the Moses toolkit. The LDS Church supplied the parallel corpus for the experiment, which uses gloss to represent the ASL portion of the corpus.

---

[6] https://www.lds.org/languages/ase?lang=eng.

Chapter 3 – Methodology

The development of our English to ASL MT system involved five basic steps (see Figure 5):

1. We curated a bitext of the LDS Church human-translated material consisting of English and ASL gloss.
2. We downloaded and installed the Moses toolkit and learned how to use it.
3. We loaded the parallel corpus into Moses and built the English, ASL, and translation language models.
4. After partitioning the corpus into three parts, (training set, development (dev) set, and test set), we trained the system and performed a baseline evaluation, tuned the system and made adjustments to improve the MT output.
5. Finally, using the test set, we decoded and evaluated the final output and compared the final results with the original baseline score.

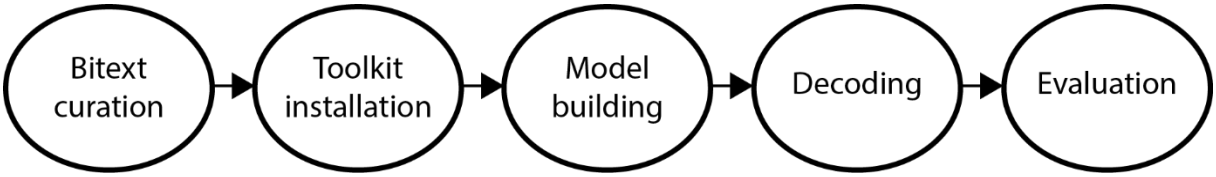

Figure 5: Series of procedures implemented in this thesis

Corpus curation, as mentioned in the previous section, proceeded as follows. The LDS Church supplied a religious parallel corpus of thousands of English and human-translated ASL gloss lines for this thesis. We downloaded and installed Olifant[7], an open-source translation memory editor. The received the raw text of the corpus in translation memory exchange (TMX) files. TMX format allows transferring translation memory between users and translation tools. The corpus was segmented into paragraphs interspersed with computer markup language, which were computer commands that followed standard markup language conventions.

[7] http://okapi.sourceforge.net/applications.html

We loaded the parallel corpus into Olifant. We edited the bitext by splitting the paragraphs into sentences and phrases of complete thoughts. Since Moses tools translate raw text, we also cleaned the corpus to remove the markup language from the text. We removed markup language from the bitext and split the paragraphs into sentence/phrases, working on both sides of the sentence pairs, in preparation for MT in Moses. This cleanup work required considerable time and effort. Figure 6 shows an example of one sequence that needed to be split into sentences, with hand-inserted split markers to divide sentences.

Source language, English:

If you are teaching a Melchizedek Priesthood or Relief Society lesson, you should not set this book aside or prepare lessons from other materials. [$SPLIT$]Prayerfully select from the chapter those teachings that you feel will be most helpful to those you teach. [$SPLIT$] Some chapters contain more material than you will be able to discuss during class time.[$SPLIT$] Allow good discussions to continue rather than trying to cover all the teachings.

Target Language, ASL Gloss:

//if you teach++ lesson\\ for melchizedek priesthood shrug relief-society, //use book\\ (head shake no) put-aside shrug focus other book++for purpose prior prepare not.[$SPLIT$] //chapter list\\ you pray can pick-from-list //that\\ will better help people you teach.[$SPLIT$]//some chapter++ inside\\ have more thing //during class time\\ enough time can discuss all not.[$SPLIT$]//you worry try teach everything\\ not //if have good discussion\\ go-ahead discuss.

Figure 6: Source and target sentences with hand-inserted SPLIT markers

Occasional errors in the markup language caused Olifant to display an error message giving the exact location of the mistake. To correct the error, we used Notepad++, an open-source source code editor.

As many of the translations we received in the TMX files had markup language and several sentences per sequence in the Olifant application, the work required more than an hour per 100 bitext chunks to divide the paragraphs into separate sentences. To indicate the beginning

and ending of chunks of text, we then wrapped the sentences in Standard Generalized Markup Language (SGML) code, which is a requirement of the NIST scoring tool for evaluation, and prepared the data for training the translation system.

For training and testing an MT system, the parallel corpus must be divided into two or three partitions of randomly selected bitext; the training set for training the MT system; the development (dev) set (optional) for use during development of the system; and the test set for the final evaluation of how well the system was able to translate. Many SLMT articles discuss the percentage chosen for each partition, which typically range from 64%-98% for the training set, 2%-12.5% for the dev set (when employed), and 2%-20% for the test set. **Table 1** summarizes corpus size, language, and partitions of SLMT researchers discussed previously. (Dreuw et al. 2008)

After studying how other researchers had partitioned their experiments, we determined our partition approach. Based on Morrissey and Way (2013), we partitioned the corpus into three sets as follows: 70% for the testing set, 10% for the dev set, and 20% for the test set. The corpus partitioned in this manner gives us a large training set with which to train the MT decoder, a sizeable dev set, and a substantial test set containing enough data to give a good indication as to how the MT system performed.

3.1. Toolkit installation and use

Information found in the Moses User's Guide (MUG) on the Moses website helped us establish our SMT infrastructure.

Table 1: Corpus partitions for various spoken language text/sign language gloss MT projects

Paper	Lang.	Quantity	Partition
Almohimeed et al. 2011	Arabic	203 sentence pairs	-
Bauer et al. 1999	German	100 signs	-
Bungeroth and Ney 2004	German	200 sentence pairs	84/16
Dasgupta et al. 2008	Indian	208 sentence pairs	-
D'Haro et al. 2008	Spanish	416 sentence pairs	75/12.5/12.5
Dreuw et al. 2008	English	843 sentence pairs	75/12.5/12.5
Forster et al. 2014	German	8767 sentences, (6861 DGS)	94/3/3
López-Ludeña et al. 2012	Spanish	3496 sentences, (820 LSE)	75/12.5/12.5
López-Ludeña et al. 2013	Spanish	12,741 signs	-
López-Ludeña et al. 2014a	Spanish	500 sentence pairs	75/12.5/12.5
López-Ludeña et al. 2014b	Spanish	1364 sentences, (443 LSE)	75/12.5/12.5
Lugaresi and DiEugenio 2013	Italian	374 sentences, (376 LIS)	-
Morrissey and Way 2005	English	561 sentence pairs	90/10
Morrissey et al. 2007	English	595 sentence pairs	-
Morrissey et al. 2010	English	350 utterances	-
Morrissey and Way 2013	Several	595 sentences	70/10/20
Othman et al. 2011	English	431 sentence pairs	-
Porta et al. 2014	Spanish	229 sentence pairs	15/85
San-Segundo et al. 2006	Spanish	135 phrases	-
San-Segundo et al. 2012	Spanish	1,413 sentences, (199 LSE)	75/12.5/12.5
Stein et al. 2012	German	3077 sentence pairs	83/17
Wray et al. 2004	English	500 phrases	-
This thesis	English	14,247 sentence pairs	70/10/20

The first step in using Moses was to download and compile the source code[8]. We went through the procedures step-by-step: we downloaded KEN-LM, a word alignment program called GIZA++, Moses, and Moses scripts. Following the commands for installing the packages, we built a baseline translation system and ran Moses. After the download, we ran the prepared data from the Shared Task link to test our installation of the MT toolkit.

[8] http://www.statmt.org/moses/

3.2. System refinement

To begin the development process, we trained Moses on the training set, had Moses translate the English portion of the training set with off-the-shelf settings, and evaluated the MT output. This provided a baseline score on the seen data in the training set. Next we established a second baseline by partitioning the training set into five equal parts and evaluating the unseen data using 5-fold cross validation. We then tuned the system and had it retranslate the English portion of the training set; we then evaluated the MT output.

We also performed a 'trial run' on the dev set as unseen data and with scored. We checked the output of the dev set to ensure we had not overtuned the system on the training set. We ran a couple of development iterations and analyzed the output to determine the MT issues and what we could do to improve the system to produce a better translation. Considerations for the adjustment process are: Does the reordering model need more weight, or a higher percentage, or does the translation model need a higher percentage? We adjusted the weights until the right balance and combination of weights produced a better translation (Koehn 2010). With each iteration of testing, the Moses decoder learned English text/ASL gloss further.

Using the tokenized English text and then the tokenized ASL gloss text, we trained truecase models for each language. Truecasing eliminates most capital letters, keeping only the capitals that are necessary to the translation. The English truecase model changed the capital letters at the beginning of sentences into lower-case letters. Capital letters of proper nouns remain capitalized. The truecase model for ASL gloss recognized which letters in the data should remain capitalized and which letters should be lowercased. Standard ASL gloss is made by capitalizing each gloss word. For the LDS Church, however, capitalization of all words is not the standard and only fingerspelled words are capitalized. Some of the older data we received from

the LDS Church was inconsistent with the current translation guidelines and contained capitals for each gloss word. We adjusted the parameters to help the truecasing model recognize only the necessary capitals.

3.3. Final evaluation

The best way to determine translation quality is to use human evaluators who are native speakers of the TL and skilled in the SL. (Of course, this is problematic for this thesis, since there are no native speakers of ASL gloss.) However, there can be great variation between evaluators. Some factors that affect translation and translation evaluation include the audience of the translation, the purpose for the translation, the need (or not) for complete concept correctness, and the resources available for the quality check on the translation. Human evaluators can recognize a good sentence, but may vary in word choices and evaluation scores: one might be more critically perfectionistic, whereas another might be more casual and accepting of mistakes; one might be quick in the process, while another might be slow (Papineni et al. 2002). In any case, human evaluators are expensive and in short supply. Due to the lack of human evaluators, the variance of opinions of evaluators, and the costs in time and money to use them, many researchers opt to use computerized metrics instead.

One such metric is BLEU, which automatically evaluates an MT output (candidate) translation by comparing it to a human translated (reference) translation (Papineni et al. 2002). The BLEU metric is language-independent, automated, (debatably) accurate, open-source and widely used as an industry standard in MT output evaluation. For these reasons, we used BLEU to evaluate the MT output frequently throughout the development cycle of the thesis and for the final evaluation. Figure 7 shows the basic steps taken in our system to obtain BLEU scores.

Figure 7: MT and evaluation steps for the current thesis

In assigning a BLEU score to candidate translations, an evaluation tool follows these guidelines:

1. Find and count *n*-gram matches.
2. Penalize sentences that are too short or too long.
3. Penalize candidate translation words that are used more frequently than the reference translation.
4. Penalize candidate translation when multiple translations of a word are used.
5. Prefer a candidate translation that matches the reference translation in length, word order, and word choice.

The similarity between the reference and the candidate translation gives a BLEU score ranging between 0.0 – 1.0. A score of 1.0 indicates that the candidate translation is an exact equal translation to the reference translation. This does not often occur, even when comparing human translations. BLEU scores correlate highly with human evaluation scores (Papineni et al. 2002).

Since BLEU measures *n*-grams, BLEU works best with high quantities of translation material (more reference and candidate sentence pairs) to improve the quality of the score. Having multiple reference translations per candidate translation sentence would also help boost the BLEU score, by adding possible matches to the candidate translation. When an MT system has been trained using a limited corpus of a small domain, the BLEU score tends to be higher (Papineni et al. 2002). Research with low-resource languages typically yields BLEU scores of

between 0.0 – 0.2 (Genzel et al. 2009; Ma et al. 2011; Irvine and Callison-Burch 2013). Even a small increase in a BLEU score of .06 is discernible by humans[9].

To summarize, Figure 8 shows the steps we completed during the training and evaluation of the SMT:

- Partitioning the corpus into three sets; training, development, and test set.
- Decoding the training set.
- Evaluating output to attain Baseline #1 BLEU score and 5-fold cross-validation to attain Baseline #2 BLEU score.
- Training and tuning the system on the training set for several iterations.
- System refinement on the training set.
- Training the system with the newly refined training set.
- Decoding the dev set.
- Evaluating output to attain the BLEU score and ascertain whether the system had improved.
- Training and tuning the system on the dev set through several iterations.
- Training the system with the newly refined training/dev set.
- Evaluating the system for the final BLEU score.

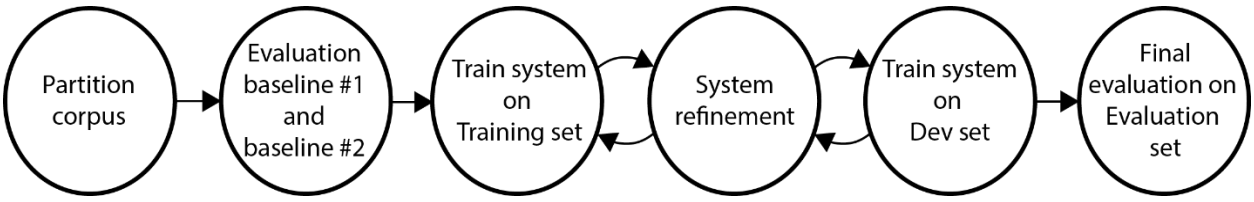

Figure 8: Procedure for training and evaluation of the thesis

Sign languages are considered low-resource languages and we had access to a small parallel corpus of English/ASL gloss. We curated the bitext, downloaded and installed Moses, aligned the sentence pairs, created the language models, and prepared Moses to translate. We began a cycle of training, evaluating, tuning, retraining, evaluating, refining, retraining, and

[9] Richardson personal communication.

evaluating. We did this cycle using two partitions, the training set and dev set. We then combined the two sets into the training+dev set.

In conclusion to the experiment, we translated the final test partition, which was completely unseen data. We evaluated the output using BLEU. The results were comparable to other low-resource language MT and will be shown in the next chapter.

Chapter 4 – Results and Evaluation

During the development phase of an SMT system, researchers frequently evaluate the target language output for determining whether the researchers' system modifications have improved the system.

As explained in the previous chapter, we followed Morrissey and Way (2013) in partitioning our corpus into three sets: 70% for the training set, 10% for the dev set, and 20% for the final test set. This allowed a large training set with which to train the Moses system and establish a baseline score, a good-sized dev set to run an evaluation and to refine the Moses system, and a substantial test set.

We thus partitioned the bitext by randomly selecting 9,992 sentence pairs for the training set, 1,427 for the dev set and 2,855 for the test set. After partitioning the corpus, we trained the Moses system on the training set using the Moses default settings.

4.1. System evaluations

We then had Moses decode the entire training set and ran an automatic evaluation of the candidate translation to establish a baseline score. We needed a baseline score to determine how well the training set, in and of itself, had trained Moses without any user adjustments. We expected that our subsequent improvements would increase the evaluation score at the conclusion of the thesis.

We used a scoring tool provided by the National Institute of Standards and Technology (NIST) (Papineni, et al. 2002), and compared the newly decoded candidate translation of the training set to the human reference translation of the training set. As the machine had been trained using the same bitext sentences, this constituted processing of seen data. Working with

ASL gloss is a challenge, such as not sufficient data, multiple translators and word order mismatch. We wanted to see how well the system decoded, so we evaluated the MT output. We achieved a BLEU score of 0.5104. This is an encouraging score, considering that the combination of English and ASL gloss was new to the Moses system. We titled this score Baseline #1.

As the Moses system had seen the bitext data prior to decoding, we decided to run a second baseline using a technique called 5-fold cross-validation on the training set. To run the cross-validation (CV), we randomly divided the 9,992 bitext lines of the training set into five partitions, with approximately 2,000 bitext lines in each partition. We trained Moses using default settings on the first four partitions and then had the Moses decoder translate the remaining fifth partition of unseen or unfamiliar data. We did this five times, each time choosing four different partitions to train and leaving one remaining partition of unseen or unfamiliar data for Moses to decode, until all five partitions had been translated.

Table 2: Baseline scores on the training set (seen data; 9,992 sentence pairs)

EXPERIMENTS	BLEU SCORE
Baseline #1	0.5104
After tuning	0.6260
Baseline #2: 5-fold CV	0.1223

The 5-fold cross-validation gave us five sets of candidate translations. We applied the NIST scoring tool to each translation, comparing the candidate translations to the reference translations. This gave us five separate BLEU scores: 0.1240, 0.1258, 0.1228, and 0.1198 and 0.1193. We averaged these five scores together to establish a BLEU score of 0.1223. We called this average Baseline #2. We expected Baseline #1 to be higher than Baseline #2 as we included

only four portions of the training set each time we ran the 5-fold cross-validation, and thus the fifth partition of the test would be unseen data for that round. Table 2 shows the baseline scores.

Moses has an automatic tuning feature. To tune automatically, Moses makes a copy of itself, checks the candidate translation against the reference translation, and determines where it can improve in translating by adjusting its own parameters. After establishing the two baseline BLEU scores, we automatically tuned the system using the full training set. This included:

- Automatically tuning the system
- Updating the system with the new tuning weights
- Retranslating the training set using the new weights
- Evaluating the output

The BLEU score after tuning was 0.6260, an increase of 0.1156.

4.2 Development and dev test phase

After the automatic tuning, we ran two rounds of hand refinements and adjustments to fine-tune the SMT system. By running iterations of evaluating, adjusting parameters and decoding, we expected to be able to improve the SMT system and increase the BLEU score further.

We loaded the candidate and corresponding reference translation into Notepad++ and hand compared the translations line-by-line. Then we analyzed and listed several issues in the machine translation. Prioritizing the issues, we determined which adjustments we would like to attempt; adjustments that we expected could substantially improve the translation system. Next we then refined the translations, adjusted the weights of the models, decoded, automatically evaluated the new candidate translations, and calculated new BLEU scores through both iterations.

For Round 1, we noticed a few tokenization issues that could be changed, such as:

- The ASL gloss for tithing is 1/10; the fraction is split up: + + 1 / 10 pay-to-Lord.
- The deity reference caret is separated from its word.
- The topic slash markers are separated from themselves and from the topic.
- Plural ++ sequences are separated.
- Fingerspelling hashtags are separated from the word, but fs- for fingerspelling is not.
- Classifiers cl: become separated.
- The NMM's or explanation words are placed within parentheses, which become separated.
- The ASL gloss uses > or <, which the tool does not accept. Instead, we used < – tokenized - & lt ;
- Compound signs joined with a + are separated.

We made several changes related to these tokenization issues to enable the reference and the candidate translations to be similarly tokenized. The changes we made were:

- We added spaces after specific characters, such as , / ? . + ( ) # and ! from their words.
- We added a space to separate topic markers // and \\.
- We space-delimited non-ASCII characters, such as exotic encodings for quote marks, ellipses, etc.
- We removed the space after the caret ^. Recall that the ^ mark is used to signify signing the attached word upwards, such as to signal deity or heavenward.

The candidate translation had more frequent capital letters in the data than did the reference translation. Because of the variation in upper and lower case, we did not include case sensitivity when using the scoring tool.

The first round yielded a higher BLEU score, reflecting improvement in the SMT system. Making the few adjustments explained above, we increased the BLEU score to 0.6726. When we eliminated case sensitivity, the BLEU score increased to 0.6749.

In Round 2 we corrected further tokenization discrepancies in the reference and candidate translations. The second round yielded a BLEU score of 0.6761. With the iterations of evaluation, adjusting parameters and decoding of the training set, the BLEU scores showed that

our training set refinements and parameter adjustments we made to the MT system had a positive impact on the candidate translation.

Table 3 shows the various BLEU scores we received when we evaluated the development (dev) set. After automatic tuning of the SMT system we ran a test of Moses decoding the English portion of the dev set. The dev set was 10% or, 1,427 sentence pairs of unseen data. The candidate translation of the dev set was evaluated by comparing it to the reference translation of the same set using the NIST scoring tool. We got a BLEU score of 0.1454 on the dev set. We then used the hand-refinements to test the dev set, applied the first set of refinements and got a BLEU score of 0.1677. We then applied the second set of refinements and got a BLEU score of 0.1678.

These values are substantially lower than the Baseline #1 previously cited. This is because the data in the dev set was unseen by Moses and by us. The 1,427 sentence pairs had been randomly selected prior to any training and set aside for translation and evaluation during the development phase.

We then combined the training and dev sets (training+dev), which totaled 80% of the data, or 11,419 sentence pairs. We did a baseline training and translation on the combined sets (seen data) and got a BLEU score of 0.5634. This represents a substantial improvement over Baseline #1.

Finally, using the hand-refinements we developed in the development phase, we had Moses tune its parameter on the training+dev set. This tuning took approximately 20 hours of processing time on a dedicated server. We then had Moses retranslate the training+dev set and we ran the scoring tool on the result, which yielded a BLEU score of 0.6666, which gave an increase of 0.1032.

Table 3: Evaluation scores on dev and training+dev sets

Experiments	# of sentence pairs	BLEU score
Baseline Dev test (10% of data, unseen)	1,427	0.1454
Round 1 (1st set of refinements)	1,427	0.1677
Round 2 (2nd set of refinements)	1,427	0.1678
Training+dev set (80% of data, seen)	11, 419	0.5634
After tuning on training+dev set (80% of data, seen)	11,419	0.6666

4.3 Final evaluation

When we had refined and adjusted the SMT system, thus completing the development phase, it was ready for final evaluation. This involved training Moses on the training+dev set (ie 80%) of the corpus, and then decoding on the last 20% of the corpus, the evaluation/test set. This was unseen data for us and for Moses. This consisted of 2,855 sentence pairs.

Table 4: Final evaluation of unseen data of the test set (2,855 sentence pairs)

Experiments	BLEU score
Evaluation of final test set (20% of data) (unseen data)	0.1765
Evaluation of final test set 5-fold cross-validation	0.1606

As before, our final evaluation compared the candidate translation of the test set with the human-produced reference translation of the same set. The results of our final evaluation are shown in Table 4. We again used the NIST scoring tool and received a BLEU score of 0.1765, which is as good as or even better than other low-resource languages using Moses.

Since the paucity of parallel corpora for low-resource languages, such as sign languages, typically yields BLEU scores of 0.08 - 0.12 in SMT research, our final BLEU score of 0.1765 was clearly a state-of-the-art result.

4.4 Analysis of sample translations

Following is an analysis of a few English sentences with their corresponding reference translations and candidate translations. The examples below have been labeled ‘poor’, ‘somewhat reasonable’, and ‘good candidate translations’. When looking at the ‘good candidate translations’ and comparing them to the reference translations, they would not do well on the BLEU score, even though they are good translations, due to lack of corresponding phrases. Occasionally, the ‘reasonable’ translation would receive a higher BLEU score, due to more corresponding phrases.

Poor candidate translations:

E: No one knows the exact time that the Savior will come again.

R: ?When time come-down when? None know.

C: one No know that prayer will .

The reference translation adds a rhetorical question delimited by ‘when’ and question marks, followed by the answer: ‘none know’, following ASL structure. The reference fails to mention the ‘Savior’. The candidate translation contains several words that match the English source (but only has ‘know’ in common with the reference). The system added the gloss ‘prayer’ to the candidate. Like the reference, the candidate fails to mention the ‘Savior’.

E: it is not sufficient that he answer the calls that our Heavenly Father makes from time to time through his servants for mission service;

R: Himself must become worthy, mission work preach can do++.

C: he (& gt ; index) not answer called , Heavenly Father inform them (alt) teach + + what ? His servants serve + + + .

The reference in this example fails to give the semantic equivalence of the English sentence. The only word that matches the English is ‘mission’ (‘he’ is translated as ‘himself’). The original alignment may have been off on this one, since they don’t seem to match conceptually at all.

The candidate translation has several words matching the English, though no words match the reference (to great detriment in BLEU-based scoring). The word ‘calls’ is in the candidate as ‘called’, and ‘service’ was translated as ‘serve+++’, which is OK for service. ‘Call,’ ‘calls,’ ‘called’ have the root sign of ‘call’ (as in church calling) in this instance. The morphology might be handled in a variety of ways, such as ‘calls’ being ‘call++’ or ‘call variety’; similarly for ‘serve’ and ‘service’. The candidate includes information in parentheses to reference the ‘he’, which is not in the reference translation. The candidate also adds ‘(alt)’ unnecessarily. The candidate translation includes a rhetorical question that is not necessary. The candidate has added words that are not included in the English concept, such as ‘teach’, which has been pluralized, maybe translated incorrectly for ‘mission service’ in the English. The reference mentions ‘preach’, and this also could have been translated from ‘mission service’ in the English. ‘Preach’ and ‘teach’ are what mission service entails, however they would not be the gloss for ‘mission service’.

E: “Yea, for food and for raiment, for taste and for smell, to strengthen the body and to enliven the soul”

R: Things there earth// for, clothes, things need++ , for food can taste smell, help body become strong, spirit uplift\\.

C: yes , / / things grow-grow there , food , & gt ; he smell for world & lt ; body , sink-into enliven . (End quote) .

The reference in this example placed the topic markers in the wrong part of the sentence. The English ‘strengthen the body’ is translated ‘help body become strong’ and ‘to enliven the soul’ is translated ‘spirit uplift’, which is reasonable. The reference adds information not in the English source, such as ‘things here earth’ and ‘things need++’.

The candidate translation has extensive lexical overlap with the English and the reference translation. Additional information such topic markers appear in the candidate, which signify body shifts to show a slight shift in the shoulders side to side while signing the different concepts, such as ‘food’, ‘he smell for world’, ‘body, sink into enliven’. The last two do not make sense in ASL. The English has quotes around the sentence. The candidate ends with ‘. (End quote) . ASL gloss should include quotation marks at the beginning and the end of the quotation and would be signed by the signer. The reference doesn’t address the quotes at all.

Somewhat reasonable candidate translations:

E: I know that Joseph Smith was a prophet of the Lord.

R: I know JOSEPH SMITH summary JS true+biz ^Lord ^His prophet.

C: me finish know JS true-biz prophet .

The reference adds ‘summary’ to signal that Joseph Smith will now be signed JS. It largely follows English except for the ASL structure ‘true+biz ^Lord ^His prophet’. The candidate adds ‘finish’ (unnecessary, but used frequently in the reference portion of the bitext). The candidate

only uses ‘JS’ instead of ‘Joseph Smith’, which is appropriate. It also does not include ‘of the Lord’ in the translation. ‘I’ is translated as ‘me’, a case mismatch with the reference.

E: They taught simple truths and invited their investigators to pray.

R: Both finish taught simple truth, invite them investigators go-ahead pray.

C: / / That ix-ghost \ \ finish taught simple truth , invite them investigators go-ahead pray .

The reference had ‘both’ (possibly information from the previous sentence) and ‘finish’ (superfluous). The reference followed the English source closely, though adding ‘go-ahead’; interestingly, which the candidate also added this aspectual marker. The source term ‘truths’ was translated in the singular in both the reference and candidate. ‘Truth’ and ‘truths’ could be signed in the same manner, or ‘truths’ could be signed ‘truth cl:5-list’. If the reference includes the ‘s’ and the candidate does not, it would affect the BLEU score, even though they would be equivalent translations.

The candidate added, with topic markers, ‘That ix-ghost’ [10] when there is no indication the English is talking about a ghost. The remainder of the sentence matches exactly the reference translation, except for whitespace, which impacts BLEU scoring.

E: The Savior made an infinite atoning sacrifice so that if we have faith in Him, repent, are baptized, receive the Holy Ghost, and endure to the end, we can receive forgiveness of our sins and enter and follow the path that will lead us to eternal life in God’s presence.

---

[10] ‘ix’ has been eliminated from the LDS Church translation guide and replaced with ‘he’ or ‘she’ or ‘it’. The parallel corpus we were given had some of the old gloss and the candidate used it at times

R: Savior ^His sacrifice related-to atonement benefit full all. Happen we have faith in ^Him, repent, baptized, receive holy ghost, endure end, we can forgive remove sin++ can enter follow path continue lead eternal life BACK present God.

C: / / Savior BACK present . happen we have faith in ^ Him , repent , baptized , receive holy ghost , endure end , we can forgive remove sin + + can enter follow path continue lead eternal life with God ^ His presence again .

The reference and candidate use 'happen' for 'if'. The candidate matches the reference for the entire middle section: 'happen … eternal life', except for whitespace around plural and deity tokens. The reference uses 'BACK' and the candidate also uses 'BACK' but in a different place, inserting 'again' where the reference has 'BACK'.

The candidate added half of a topic marker // at the beginning of the translation. The candidate glossed '/ / Savior BACK present', hence only one word—'Savior'—matched the beginning segment of the English and the reference.

Good candidate translations:

E: In the temple we receive sacred ordinances, including ordinances that bind families for eternity.

R: There temple, sacred ordinance+ we get, include what? ordinance+ family seal eternity.

C: / / Happen temple ordinance we receive sacred ordinance + that seal both heaven , family + + forever .

The reference glosses 'there temple', the English uses 'in the temple' and the candidate glosses '/ / happen temple'. The candidate shows half of a topic marker (an appropriate enough place, just lacking the closing topic marker). The candidate glossed 'family + +' for 'families', where the reference only glossed 'family'. The '+ +' were separated from the word and each other. The reference had a rhetorical question; the candidate did not. The English and the reference used

‘eternity’; the candidate used ‘forever’. ‘Forever’ and ‘eternity’ are signed the same. Commas are in different places in the sentence, with the commas in the candidate being separated from the words.

E: If we are living exemplary lives, our influence may encourage others to learn about the gospel.

R: Happen we live good+straight, our influence, >they, interested, want learn gospel.

C: suppose we true business exemplary life , / / other people \ \ \ \ influence can learn gospel .

The reference added the ‘>they,’ to show to point to the right for ‘they’. The candidate glossed ‘/ / other people / / / / ‘, and added topic markers. (Additional marker slashes on the left). The candidate used ‘exemplary’ from the English, whereas the reference used ‘good+straight’ as the gloss for ‘exemplary’. ‘If’ in the English, ‘happen’ in the reference’, and ‘suppose’ in the candidate are all correct, but none match the source. The candidate glossed ‘true business’, which is usually glossed ‘true-biz’, but is signed the same way. The candidate and the reference used the gloss ‘influence’ from the English. Neither translation used ‘encourage’.

E: How can I help the poor and needy?

R: ?Poor people their need, me help how?

C: how can help people group poor , how ?

The reference translation glossed the question with a ‘? ‘at the beginning and ‘how?’ at the end. The candidate glosses ‘how’ at the beginning and ‘how?’ at the end, for the question. (In the

bitext, some translators use the question word both at the beginning and the end.) The reference doesn't setup the 'poor people' into a group by pointing with the possessive palm for use in signing 'their need'. However, the candidate does gloss 'group' which could then be referenced with the possessive palm of the hand. The reference does include the first person pronoun; the candidate does not indicate who can help the poor. The reference includes 'their need', for the English 'needy', but the candidate just glosses 'poor' and not 'needy'.

Chapter 5 – Conclusions and Discussion

We trained a Moses MT system using an English text/ASL gloss parallel corpus within the religious domain of The Church of Jesus Christ of Latter-day Saints. We curated the parallel corpus, hand-separating multiple-sentence paragraphs into single sentences and phrases. We also hand-separated the extensive markup language from the parallel corpus. After weeks of cleaning the English/ASL bitext with the Olifant tool, we had created apparently the largest parallel corpus, thus far, of any sign language worldwide.

We then downloaded and installed the Moses toolkit on a dedicated computer on the BYU campus. Loading the bitext into the Moses toolkit, we trained and tuned the system through several iterations. Each time we evaluated the output using automatic scoring tools. Each iteration yielded increasingly better output results with increasing BLEU scores. We attained a BLEU score on unseen data that was comparable to, and even better than, typical BLEU scores for low-resource languages.

5.1. Future work

Continuing this research with additional quality data, if available, would improve the MT output.

MT is typically only one component in the overall process of document production. More thorough analysis of how our system might impact documentation production could be studied by others to investigate the viability of our system in a production environment.

Follow-on research could involve a thorough, systematic linguistic evaluation of the output results to consider ways the MT system might be improved, such as using constituent reordering models to compensate for syntactic differences between English and ASL gloss.

Instead of using humans to convert the MT output to sign language, extra steps could be taken that would include an additional intermediate representation of HamNoSys to the final output for programming a computer avatar that would sign the message in sign language.

This thesis includes automatic evaluation using BLEU. Future work might include human evaluation of the output. That would entail incorporating people skilled in reading ASL gloss and English to evaluate the output. Or the output can be converted to sign language and a target audience skilled in ASL and English could watch the ASL message and compare it with the English to analyze the output. These human evaluation options are beyond the scope of this thesis.

References

Almohimeed, Abdulaziz, Mike Wald, and Robert I. Damper. 2011. "Arabic text to Arabic sign language translation system for the deaf and hearing-impaired community." *In Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies* pp. 101-109.

Baker-Shenk, Charlotte, and Dennis Cokely. 1981. *American Sign Language: A Teacher's Resource Text on Grammar and Culture.* Washington, D.C. Gallaudet University Press.

Bangham, J Andrew, Stephen Cox, Ralph Elliott, John Glauert, Ian Marshall, Sanja Rankov, and Mark Wells. 2000. "Virtual signing: capture, animation, storage and transmission ─ an overview of the ViSiCAST project."

Bauer, Britta, Sonja Nießen, and Heinz Hermann. 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,* [no page numbers], Siena, Italy.

Brown, Peter F., John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Frederick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. 1990. "A Statistical Approach to Machine Translation." *Computational Linguistics* 16 (2). 79-85.

Bungeroth, Jan, and Hermann Ney. 2004. "Statistical Sign Language Translation." In *Proceedings from the LREC Workshop On Representation and Processing of Sign Languages.* 105-108.

Caridakis, George, Kostas Karpouzis, Golina Sapountzaki, Stravola-Evita Fotinea, and Eleni Efthimiou. 2005. "A Dynamic Environment for Greek Sign Language Synthesis Using Virtual Characters." *Proceedings of the 10th International Conference on 2D Web Technology.* 69-76.

Corbett, Christine. 2006. "23C3: Open source machine translation." *23rd Chaos Communication Congress,* presentation. https://www.youtube.com/watch?v=mnOrCnSh7z8

Craciunescu, Olivia, Constanza Gerding-Salas, and Susan Stringer-O'Keeffe. 2008. "Machine translation and computer-assisted translation: a new way of translating?" *Translators and Computers.* 8 (3).

Dasgupta, Tirthankar, Sandipan Dandapat, and Anupam Basu. 2008. "Prototype machine translation system from text-to-Indian sign language." *IJCNLP 2008: Proceedings of the Workshop on NLP for Less Privileged Languages.* 19-26.

D'Haro, Luis Fernando, Rubén San-Segundo, de Córdoba Ricardo, Jan Bungeroth, Daniel Stein, and Hermann Ney. 2008. "Language model adaptation for a speech to sign language translation system using web frequencies and a MAP framework." *Proceedings of the 9th Annual Conference of the International Speech Communication Association.* 2199-2202.

Dreuw, Philippe, Carol Neidle, Vassilis Athitsos, Stan Sclaroff, and Hermann Ney. "Benchmark databases for video-based automatic sign language recognition." *LREC.* 2008.

Forster, Jens, Christoph Schmidt, Oscar Koller, Martin Bellgardt, and Hermann Ney. 2014. "Extensions of the sign language recognition and translation corpus RWTH-PHOENIX-weather." *Proceedings of the 9th Language Resources and Evaluation Conference (LREC 2014).* 1911-1916.

Forster, Jens, Daniel Stein, Ellen Ormel, Onno Crasborn, and Hermann Ney. 2010. "Best practice for sign langauge data collections regarding the needs of data-driven recognition and translation." *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Langauge Technologies (CSLT 2010).* 92-98.

Genzel, Dmitriy, Klaus Macherey, and Jakob Uszkoreit. 2009. "Creating a high-quality machine translation system for a low-resource language: Yiddish." *Proceedings of the Twelfth Machine Translation Summit.* 41-48.

Grieve-Smith, Angus. 1999. "English to American Sign Language machine translation of weather reports." *Proceedings of the Second High Desert Student Conference in Linguistics.* 23-30.

Huenerfauth, Matt, Liming Ahzo, Erdan Gu, and Jan Allbeck. 2007. "Design and evaluation of an American Sign Language generator." *Proceedings of the ACL 2007 Workshop on Embodied Language Processing.* 51-58.

Irvine, Ann, and Chris Callison-Burch. 2013. "Combining bilingual and comparable corpora for low resource machine translation." *Proceedings of the Eighth Workshop on Statistical Machine Translation.* 262-270.

Koehn, Philipp. (2005). "Europarl: A Parallel Corpus for Statistical Machine Translation." *MT Summit* 5. 79-86.

Koehn, Philipp. 2010. *Statistical Machine Translation.* New York: Cambridge University Press.

Koehn, Philipp, and Hieu Hoang. 2012. "Open source statistical machine translation." *The Tenth Biennial Conference of the Association for Machine Translation in the Americas.* Presentation. 139 slides.

Koehn, Philipp, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, et al. 2007. "Moses: Open source toolkit for statistical machine translation." *In Proceedings of the 45th Annual Meeting of the ACL: Interactive Poster and Demonstration Sessions.* 177-180.

Kyle, Jim G., and Bencie Woll. 1988. *Sign language: The study of deaf people and their language.* Cambridge University Press.

Lopez, Adam. 2008. "Statistical machine translation." *ACM Computing Surveys.* 40 (3). 1-49.

López-Ludeña, Verónica, Carlos González-Morcillo, Juan Carlos López, Roberto Barra-Chicote, Ricardo Córdoba, and Rubén San-Segundo. 2014a. "Translating bus information into sign language for deaf people." *Engineering Applications of Artificial Intelligence.* 258-269.

López-Ludeña, Verónica, Carlos González-Morcillo, E. Ferreiro, Javier Ferreiros, Juan Carlos López, and Rubén San-Segundo. 2014b. "Methodology for developing an advanced communications system for the Deaf in a new domain." *Knowledge-Based Systems.* 240-252.

López-Ludeña, Verónica, Rubén San-Segundo, Juan Manuel Montero, Ricardo Córdoba, Javier Ferreiros, and José Manuel Pardo Muñoz. 2012. "Automatic categorization for improving Spanish into Spanish Sign Language machine translation." *Computer Speech and Language.* 149-167.

López-Ludeña, Verónica, Rubén San-Segundo, Carlos González Morcillo, Juan Carlos López, and José M. Pardo Muñoz. 2013. "Increasing adaptability of a speech into sign language translation system." *Expert Systems with Applications.* 1312-1322.

Lugaresi, Camillo, and Barbara DiEugenio. 2013. "Translating Italian connectives into Italian Sign Language." *Proceedings of the 51st Meeting of the Association for Computational Linguistics (ACL-2013).* 270-280.

Ma, Jeff, Spyros Matsoukas, and Richard Schwartz. 2011. "Improving low-resource statistical machine translation with a novel semantic word clustering algorithm." *Proceedings of the Thirteenth MT Summit.* 352-359.

Morrissey, Sara. 2008. *Data-Driven Machine Translation for Sign Languages.* DCU: Dublin City University, School of Computing, Doctoral Dissertation.

Morrissey, Sara. 2011. "Assessing three representation methods for sign language machine translation and evaluation." *Proceedings of the 15th conference of the European Association for Machine Translation.* 137-144.

Morrissey, Sara, and Andy Way. 2005. "An Example-based approach to translating sign language." *Proceedings of the Second Workshop on Example-based Machine Translation.* 109-116.

Morrissey, Sara, and Andy Way. 2013. "Manual labour: tackling machine translation for sign languages." *Machine Translation*. 25-64.

Morrissey, Sara, Harold Somers, Robert Smith, Shane Gilchrist, and Sandipan Dandapat. 2010. "Building a sign language corpus for use in machine translation." *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies.* 172-177.

Morrissey, Sara, Andy Way, Daniel Stein, Jan Bungeroth, and Hermann Ney. 2007. "Combining data-driven MT systems for improved sign language translation." *Proceedings of the Eleventh MT Summit* . 329-336.

Newkirk, Don. *SignFont Handbook.* Redmond, WA: Edmark Corporation, 1989.

Othman, Achraf, and Mohamed Jemni. 2011. "Statistical sign language machine translation: from English written text to American Sign Language gloss." *International Journal of Computer Science.* 65-73.

Othman, Achraf, Oussama El Ghoul, and Mohamed Jemni. 2010. "SportSign: a service to make sports news accessible to deaf persons in sign languages." *Computers Helping People with Special Needs.* 169-76.

Padden, Carol, and Tom Humphries. 2005. *Inside Deaf Culture.* Cambridge: Harvard University Press.

Papineni, Kishore, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. "BLEU: A method for automatic evaluation of machine translation." *40th Annual Meeting of the Association for Computational Linguistics (ACL).* 311-318.

Porta, Jordi, Fernando López-Colino, José Colás, and Javier Tejedor. 2014. "A rule-based translation from written Spanish to Spanish Sign Language glosses." *Computer Speech and Language.* 28. 788-811.

Prillwitz, Siegmund, Regina Leven, Heiko Zienert, Thomas Hanke, and Jan Henning. 1989. "An introductory guide to HamNoSys Version 2.0: Hamburg notation system for Sign Languages." *International Studies on Sign Language and Communication of the Deaf.*

Richardson, Stephen. 2012. "Using the Microsoft Translator Hub at The Church of Jesus Christ of Latter-day Saints." *Proceedings of the Tenth Conference of the Association for Machine Translation in the Americas.* 8 pp.

Richardson, Stephen. 2014. "Translation Systems at the Church of Jesus Christ of Latter-day Saints." *BYU Translation, Interpretation, & Localization Event.*

San-Segundo, Rubén, Roberto Barra-Chicote, Luis Fernando D'Haro, Juan Manuel Montero, Ricardo de Córdoba, and Javier Ferreiros. 2006. "A Spanish speech to sign language translation system for assisting Deaf-mute people." *Proceedings of the International Conference of Spoken Language Processing.* (Interspeech). 1399-1402.

San-Segundo, Rubén, Juan Manuel Montero, Ricardo de Córdoba, Valentin Sama, F. Fernandez, Luis Fernando D'Haro, Verónica López-Ludeña, D. Sánchez, and A. García. 2012. "Design, development and field evaluation of a Spanish into sign language translation system." *Pattern Analysis and Applications.* 203-224.

Speers, d'Armond. 2001. "Representation of American Sign Language for machine translation." Doctoral Dissertation, Department of Linguistics, Georgetown University.

Stein, Daniel, Christoph Schmidt, and Hermann Ney. 2012. "Analysis, preparation, and optimization of statistical machine translation." *Machine Translation.* 325-357.

Suszczanska, Nina, Przemyslaw Szmal, and Jarozlaw Francki. 2002. "Translating Polish texts into sign language in the TGT System." *Proceedings of the 20th IASTED International Multi–Conference on Applied Informatics.* 282-287.

Sutton, Valerie. 1981. *Sign writing for everyday use.* Newport Beach: Sutton Movement Writing Press.

Valli, Clayton, and Ceil Lucas. 2000. *Linguistics of American Sign Language: An Introduction.* Washington, D.C.: Gallaudet University Press.

Veale, Tony, and Alan Conway. 1994. "Cross modal comprehension in ZARDOZ, an English to sign-language translation system." *Seventh International Workshop on Natural Language Generation.* 249-252.

Way, Andy. 2013. "Traditional and emerging use-cases for machine translation." *Translating and the Computer.* 35.

Wray, Alison, Stephen Cox, Mike Lincoln, and Judy Tryggvason. 2004. "A formulaic approach to translation at the post office: reading the signs." *Language & Communication.* 59-75.

Zhao, Liwei, Karin Kipper, William Schuler, Christian Vogler, Norman Badler, and Martha Palmer. 2000. "A Machine Translation System from English to American Sign Language." I *Envisioning Machine Translation in the Information Future*. 54-67.