Text2Gloss: Translation into Sign Language Gloss with Transformers

Stanford CS224N Custom Project

Luke Babbitt
Department of Computer Science
Stanford University
lukeb23@stanford.edu

Jenna Mansueto
Department of Symbolic Systems
Stanford University
mansueto@stanford.edu

Abstract

Translation of text into its corresponding sign language gloss annotations is a key step in end-to-end sign language translation, as well as useful for sign language interpreters. We apply a transformer model to the text to gloss task, providing a baseline for further applications of transformers on this task. As sign language translation is a notably low resource translation task, we also explore methods of augmenting existing datasets to improve translation quality, including POS-tagging and multilingual datasets. We evaluate our approaches on the ASLG-PC12 and PHOENIX-14T datasets, achieving improved performance relative to baselines with multilingual corpus training.

1 Key Information

Mentor: Soumya Chatterjee
Team member contributions: Babbitt built the data loading pipeline, conducted data analysis, and ran the multilingual corpus training experiment. Mansueto implemented our preprocessing pipeline and model training/testing in Fairseq and ran the POS-tagging experiment. Both team members contributed to the report.

2 Introduction

This project will address the task of translating from a sentence in a spoken language to a sentence in its corresponding sign language gloss (sign language annotation). An example translation for English/American Sign Language (ASL) gloss is shown in Table 1.

English	the commission's role is limited to checking that there is no manifest error in the definition.
ASL Gloss	COMMISSION X-POSS ROLE BE LIMIT TO CHECK THAT DESC-RE BE NO DESC-MANIFEST ERROR IN DEFINITION.

Table 1: Example English-ASL gloss pairing from ASLG-PC12 dataset.

The opposite direction, Gloss2Text, has been proven as a helpful step in improving Sign2Text translation Camgoz et al. (2020b), which suggests that modeling Text2Gloss could potentially contribute to the development of Text2Sign translation. In addition to being an intermediary step, translation from spoken language to sign language provides for faster and more accurate interpretation of spoken languages. A live and accurate Text2Gloss model has the capacity to help sign language interpreters with live translation, a use case that further supports the necessity of our model.

Despite the importance of this task, previous work on sign language translation has primarily focused on translation in the opposite direction (from sign language to spoken language). Camgoz et al. (2020b) and Yin and Read (2020) have shown transformers to be effective in translating from sign language to spoken language, but to our knowledge, transformers have yet to be applied to text to gloss translation. In this paper, we aim to address this gap in the literature by providing a benchmark transformer for text to gloss translation.

While sign language translation is vitally important, it also presents a difficult neural machine translation (NMT) task due to the lack of parallel text/gloss corpuses. It is a classic low resource language translation task, as evidenced by the fact that the majority of previous work in this domain has been conducted on just two datasets (one English/ASL and the other German/DGS). Due to the low resources for all sign languages, our approach proposes multilingual corpus training for spoken language to sign language. We train our transformer on Text2Gloss across multiple languages in parallel, transferring learning between sign languages in real time.

3 Language Background

ASL and DGS have distinct origins and relationships with their respective spoken languages. Both ASL and DGS utilize spatial positioning, hand movements, gestures, and facial expressions to communicate meaning and indicate grammatical nuances Farrugia et al. (2016). These linguistic features can be expressed in GLOSS through the use of prefixes and indexes. Linguists have yet to reach consensus on annotating sign language, and any translation is subjective and imperfect. Given this, it's important to recognize that a gloss simplifies sign language and can be imprecise. Gloss attempts to represent with text multidimensional spatiotemporal signals, which presents an information bottleneck.

ASL emerged with influences from French Sign Language and Martha's Vineyard Sign Language, but its development was significantly shaped by interactions with English-speaking culture in the United States HDI (Human Development Institute) at the University of Kentucky (2023). As a result, ASL incorporates many signs and expressions from English. Nevertheless, it maintains its own grammar, syntax, and vocabulary, distinguishing it as a separate language NIDCD (2019).

Unlike ASL, DGS evolved independently from German, and possesses unique grammatical structures and syntax, establishing its distinctness from German Farrugia et al. (2016). The greater independence of DGS from German suggests that translation from text to gloss in German involves more complexities than for English translation.

ASL word order follows a subject-verb-object structure, whereas DGS employs a subject-object-verb arrangement Hosemann and Herrmann (2014). English and German both follow subject-object-verb as well. This implies that although DGS developed independent from German and ASL developed with influence of English, DGS and German have a more similar sentence structure than ASL and English. There are also similar and identical signs between ASL and DSG languages, both for internationally normalized signs and for concepts that rely on positional descriptions Farrugia et al. (2016).

4 Related Work

4.1 Sign Language Translation

While our task is translation from a language to its corresponding sign language gloss (Text2Gloss), previous work on sign language translation has primarily focused on the opposite translation direction (Gloss2Text) as an intermediate step in translating from sign language videos to text (Sign2Text). Previous findings by Camgoz et al. (2020b) showed that using a gloss as a mid-level representation in sign to text modeling improves performance on the Sign2Text task. For this intermediate Gloss2Text model, the authors trained an RNN-based encoder-decoder model with Gated Recurrent Units (GRUs) and report results on the PHOENIX-Weather-2014T dataset. Yin and Read (2020) also report results for a Gloss2Text model that uses the basic transformer architecture from Vaswani et al. (2023) tested on the PHOENIX-Weather-2014T dataset and the ASLG-PC12 dataset. These two sets of results, while from the opposite translation direction, nonetheless offer a baseline for our Text2Gloss model.

To the best of our knowledge, the only work on a Text2Gloss model comes from Stoll et al. (2018), who used an RNN-based encoder-decoder with GRUs. They evaluated their model on the PHOENIX-Weather-2014T dataset, and achieved performance comparable to that of Camgoz et al. (2020b) for the opposite translation direction. This RNN-based Text2Gloss model provides a useful baseline for our transformer Text2Gloss model.

4.2 Low Resource Language Translation

Given the lack of available training data for sign language translation, we look to previous work on low resource neural machine translation. While advances in NMT often depend on increasingly vast amounts of data, this subdomain of the field offers techniques for better leveraging small datasets. In particular, we experiment with two techniques:

4.2.1 Parallel Data Augmentation

Multilingual machine learning models like XLM and NLLB involve translation between many languages, and the use of language combination in training has specifically been applied for translations with low resource languages. Grave et al. (2020) and Sennrich et al. (2017) experiment with parallel text augmentation for low resource languages. They apply augmentation techniques to the existing dataset in parallel, meaning the transformer processes multiple data samples simultaneously. By doing so, parallel data augmentation aims to address the lack of diversity and coverage in the training data that results from limited resources.

We build on parallel data augmentation for low resource languages, experimenting with a new multilingual corpus training approach for sign language translation.

4.2.2 Part of Speech (POS) Tagging

The incorporation of linguistic features such as POS tags has been shown to improve the performance of neural machine translation systems (Chen et al. (2018), Pan et al. (2020), Hlaing et al. (2022)). This syntactic information can be included in the source/target sequences, word embeddings, or attention mechanisms, and previous work has demonstrated that POS tags can enhance low resource language translation models. Pan et al. (2020) used a dual-encoder transformer to separately encode source word sequences and their corresponding linguistic feature sequences, and achieved enhance performance on Turkish-to-English and Uyghur-to-Chinese translations compared to their baseline model. Hlaing et al. (2022) appended POS tags to words, and found that adding linguistic features in this way improved the performance of transformer-based models on low resource language translation.

5 Approach

5.1 Evaluation method

We used BLEU as our evaluation metric and report corpus BLEU and BLEU-1,2,3,4 scores. This was done with the sacrebleu library.

5.2 Baselines

Model	Corpus BLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Camgoz et al. (2020b) Gloss2Text		48.90	36.88	29.45	24.54
Yin and Read (2020) Gloss2Text		48.40	36.90	29.70	24.90
Stoll et al. (2018) Text2Gloss		50.67	32.25	21.54	15.26
Simple Sequence Copy	1.74	16.7	3.3	0.8	0.2

Table 2: PHOENIX-14T baselines.

Model	Corpus BLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Yin and Read (2020) Gloss2Text		92.98	89.09	85.63	82.41
Simple Sequence Copy	22.69	57.4	31.4	16.6	8.9

Table 3: ASLG-PC12 baselines.

We compare to the baselines for Gloss2Text and Text2Gloss established by Camgoz et al. (2020b), Yin and Read (2020) and Stoll et al. (2018). The Gloss2Text baselines are for the opposite translation direction, but they still provide a point of reference in evaluating our results.

Additionally, we compare our methods to a naive simple sequence copy baseline, in which we copy the input text and treat this as the translation. This approach is especially prudent for English text to ASL gloss and German text to DGS gloss because both involve the same language (i.e. English to English and German to German).

5.3 Transformer

For our experiments, we trained a 6-layer transformer as proposed in Vaswani et al. (2023), with 8 heads, word embedding size 512, shared encoder-decoder weights and learned positional encodings. To determine an appropriate number of layers, we trained 2, 4, and 6-layer transformers on the PHOENIX-14T dataset, ultimately achieving the best performance with 6 layers 4.

Layers	Corpus BLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
2	10.58	45.5	15.9	6.3	3.3
4	11.20	56.5	19.7	7.8	4.1
6	**11.89**	**60.3**	**21.6**	**8.7**	**5.1**

Table 4: BLEU Scores for PHOENIX-14T Transformer with Varying Encoder-Decoder Layers.

We employ cross entropy loss, Adam optimizer, a dropout rate of 0.1, and an early stopping mechanism with patience 5. We used Fairseq (Ott et al. (2019)) for training and testing, and conducted our experiments on a Google Cloud virtual machine with 1 x Nvidia Tesla P4 GPU.

6 Data

6.1 English-ASL Gloss Parallel Corpus 2012 Baker (2012)

The ASLG-PC12 dataset is derived from English texts sourced from Project Gutenberg, which have been converted into American Sign Language (ASL) glosses using a rule-based methodology. This corpus comprises 81,126 training pairs with 7,324 unique signs and 7,820 unique spoken words. As it is a synthetic dataset, it may be a simpler dataset for translation tasks.

6.2 PHOENIX-Weather-2014T Camgoz et al. (2020a)

The PHOENIX-14T dataset originates from weather forecast broadcasts aired on the German television channel PHOENIX. It comprises a parallel corpus featuring annotations at the gloss level with translations into spoken German language. It encompasses a vocabulary of 1,066 unique signs and 2,887 unique spoken words. In total, the dataset contains 8,116 sentence pairs, a much smaller corpus than ASLG-PC12.

The dataset comes from DGS interpretations of daily news and weather forecast airings of the German public tv-station. This limits the scope of the data to news/weather, and the context to the single tv-station. As DGS is a low-resource language, our conclusions could be improved/confirmed by testing on a larger variety of data sources.

6.3 Preprocessing and Prefixes

Both the ASLG-PC12 and PHOENIX-14T datasets come clean and processed for machine translation tasks. There are no missing values and the sentences are aligned, meaning they did not require extensive preprocessing. We applied Byte Pair Encoding (BPE) and Moses tokenization, and used an 80/10/10 train/val/test split with a fixed random seed to ensure reproducibility.

6.4 Data Analysis

Because many signs have the same name as their spoken word, its prudent to consider words that are identical across sign and text, and therefore do not require word-level translation. BLEU-1 scores demonstrate a 57.4% overlap between ASL and English words in ASLG-PC12 and a 16.7% overlap between DGS and German words. These results further confirm the similarities between ASL and English established by HDI (Human Development Institute) at the University of Kentucky (2023) as well as the differences between DGS and German established by Farrugia et al. (2016).

PHOENIX has a smaller (relative) vocabulary than ASLG-PC12 and exhibits a much higher rate of out-of-vocabulary token replacement (tokens not appearing in training) in test and validation splits. Across the ASLG-PC12 dataset, out-of-vocabulary token replacement was minimal, ranging from 0.0% to 0.0349%. In comparison, the PHOENIX-14T dataset exhibited slightly elevated out-of-vocabulary token replacement rates, ranging from 0.287% to 0.36% in the text and 0.151% to 0.323% in the gloss splits. Our model replaces out-of-vocabulary tokens in validation and test splits and imposes a penalty of 0.5 for the <unk> token appearing in translations.

Dataset	Average Word Length	Average Sentence Length
ASLG-PC12 Text	4.3436 characters	13.1240 words
ASLG-PC12 Gloss	5.1396 characters	11.7432 words
PHOENIX-14T Text	4.9060 characters	14.7722 words
PHOENIX-14T Gloss	5.6872 characters	7.6613 words

Table 5: Average Word and Sentence Lengths

Table 5 provides average word and sentence length for the datasets and languages. While ASL and English data is relatively similar in sentence length, sentence length differs greatly between DGS and German. As established by Farrugia et al. (2016), German and DGS have more differences overall than English and ASL, and sentence length appears to be one such difference. The large differences likely contribute to the poorer performance of PHOENIX-14T overall compared to ASLG-PC12.

As noted above, ASLG-PC12 contains 7,324 unique signs and 7,820 unique words, wherase PHOENIX-14T contains 1,066 unique signs and 2,887 unique spoken words. The difference in sentence length between DGS and German text likely comes from the higher number of words in the German vocabulary (i.e. DGS signers require fewer words than German speakers to relate the same sentence).

In conclusion, ASL and DGS are both low-resource languages, but these datasets are the best available for text to gloss translation. In particular, German Text2Gloss is challenged by the stark differences between DGS and German. These differences are reflected in the data, evident in sentence length and out-of-vocabulary words.

7 Experiments

We organize our experiments into three sections:

7.1 Basic Transformer

In this experiment, we trained and tested our transformer model on the ASLG-PC12 and PHOENIX-14T datasets separately.

7.2 Multilingual Corpus Training

In multilingual corpus training, we experiment with developing a sign language agnostic transformer model. Ott et al. (2019) establish that, for low resource languages, it is often beneficial to leverage data in similar but higher-resource languages, especially when they have similar vocabularies and/or structure. While both ASL and DGS are low resource, we have significantly more data and significantly better results for ASL via the ASLG-PC12 data. Consequently, we experiment with a combined language dataset that includes sentence pairs for both translations.

This approach expands on the parallel data augmentation process. Whereas parallel data augmentation involves generating augmented versions of the original data, we apply this techniques with completely separate datasets and languages (English/ASL and German/DGS). Our multilingual corpus training approach injects 8,116 randomly sampled English to ASL gloss sentence pairs from ASLG-PC12 into the 8,116 sentences pairs in PHOENIX-14T in hopes of improving results for German to DGS gloss translation. We also run a third control experiment where we inject 8,116 English to German and 8,116 German to English sentence pairs from the WMT 14 English-German Dataset Bojar et al. (2014).

In order to train a model on two different translations, we provided unique beginning of sentence (BOS) tokens for each language, which the model was trained to recognize as distinct between the two. We trained our transformer with the combined dataset, and tested on just German to DGS.

7.2.1 POS Tagging

To provide our model with additional syntactic information, we created POS-tagged datasets from the original ASLG-PC12 and PHOENIX-14T datasets. We used SpaCy (Honnibal et al. (2020)) to generate POS tags, and appended these tags to our source sequences such that each word was replaced with "word|POS". We then trained and tested our transformer model on each POS-tagged dataset.

8 Results

8.0.1 Basic Transformer

Dataset	Model	Corpus BLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
ASLG-PC12	Transformer	**93.5**	**98.0**	**94.7**	**92.3**	**90.0**
ASLG-PC12	Simple Sequence Copy	22.69	57.4	31.4	16.6	8.9
PHOENIX-14T	Transformer	**11.89**	**60.3**	**21.6**	**8.7**	**5.1**
PHOENIX-14T	Simple Sequence Copy	1.74	16.7	3.3	0.8	0.2

Table 6: BLEU Scores for ASLG-PC12 and PHOENIX-14T datasets compared to simple baseline.

Our basic transformer model showed significant improvement in BLEU-1,2,3,4 and corpus BLEU compared to our simple baselines on both datasets, as shown in Table 6.

Model	Corpus BLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Transformer	**93.5**	**98.0**	**94.7**	**92.3**	**90.0**
Yin and Read (2020) Gloss2Text		92.98	89.09	85.63	82.41

Table 7: BLEU scores for our transformer compared to ASLG-PC12 baseline.

Our model achieves higher BLEU-1,2,3,4 scores than the Gloss2Text model from Yin and Read (2020), which we believe could indicate that this translation direction (for this language/sign language pair) is an easier NMT task.

When comparing our model to Text2Gloss and Gloss2Text baselines on PHOENIX-14T in Table 8, we see that it outperforms all baselines for BLEU-1, but under performs on BLEU-2,3,4. Our

Model	Corpus BLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Transformer	11.89	**60.3**	21.6	8.7	5.1
Stoll et al. (2018) Text2Gloss		50.67	**32.25**	**21.54**	**15.26**
Camgoz et al. (2020b) Gloss2Text		48.90	36.88	29.45	24.54
Yin and Read (2020) Gloss2Text		48.40	36.90	29.70	24.90

Table 8: BLEU scores for our transformer compared to PHOENIX-14T baselines.

transformer and the Text2Gloss model from Stoll et al. (2018) show similar BLEU score distributions compared to those of the Gloss2Text models. The Text2Gloss models show higher BLEU-1 scores but lower BLEU-2,3,4 scores than the Gloss2Text models, which we believe is due to the shorter length of DGS sequences relative to German sequences. Interestingly, this trend is repeated when comparing our model to the Text2Gloss model from Stoll et al. (2018) – despite significant improvements in BLEU-1, our model shows significantly worse performance on BLEU-2,3,4, suggesting that our model struggles to correctly model DGS syntax relative to the RNN-based model.

8.0.2 Multilingual Corpus Training

Dataset	Corpus BLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
de/DGS	11.89	60.3	21.6	8.7	5.1
de/DGS + en/ASL	**14.35**	**61.5**	**23.2**	**10.0**	**5.2**
de/DGS + en/ASL + en/de	10.91	60.1	21.3	8.6	4.9

Table 9: BLEU Scores on German/DGS translation for single and combined datasets.

Multilingual Corpus Training with English to ASL showed a significant improvement in corpus BLEU and BLEU 1, 2, 3, 4 scores for German to DGS. The results indicate our model improved from training with both ASL and DSG translations present in the corpus.

The PHOENIX-14T dataset is also relatively small, so doubling the size of the dataset could simply have caused the model's improvement. However, we note that adding English to German and German to English translations did *not* improve the model's performance. This further points us to conclude that ASL and DGS's similarities can be attributed to the de/DGS + en/ASL model's improvement.

8.0.3 POS Tagging

Dataset	Corpus BLEU	BLEU-1	BLEU-2	BLEU-3	BLEU-4
ASLG-PC12	**93.5**	98.0	**94.7**	**92.3**	**90.0**
ASLG-PC12 (POS)	92.4	**98.5**	94.6	91.7	89.1
PHOENIX-14T	**12.4**	**59.6**	**21.2**	8.8	**5.0**
PHOENIX-14T (POS)	12.3	58.2	21.1	8.8	4.9

Table 10: BLEU scores for ASLG-PC12 and PHOENIX-14T datasets with and without POS tags.

Our POS tagging experiment reveals marginally better results without POS-tagging – adding POS information in this way did not significantly affect our results. This suggests that our model was not only unable to make use of syntactic information provided at the input level, but that the addition of this information slightly hindered its ability to recognize syntactic patterns (as evidenced by the associated decrease in BLEU-3,4 scores in particular). This does not entirely rule out the efficacy of POS-tagging in improving Text2Gloss translation, but indicates that this method of incorporating syntactic information is not effective. A dual-encoder transformer to separately encode POS-tags as proposed by Hlaing et al. (2022) could be a logical alternative.

9 Analysis

All of our PHOENIX-14T models struggled to overcome what Zhang et al. (2020) define as the repetition problem, in which the same words are mistakenly repeated. Some examples of repeated word transitions are: “NORD REGION MEHR KUEHL REGION UND UND UND UND UND REGEN” and “NORDWEST MEHR TROCKEN KUESTE KUESTE KUESTE TROCKEN”.

The PHOENIX-14T dataset has 1,066 unique signs across all gloss sentences. The de/DGS model produced only 183 unique signs. This lack of vocabulary is evident in our repetition errors. That being said, the de/DGS + en/ASL model produced 259 unique signs in the DGS gloss translations, an improvement from de/DGS. Overall, both models struggle to generate words that are uncommon, and mainly generate the most commonly used words (in this case, weather and news related).

We also noticed an improvement in the de/DGS + en/ASL model in handling sign language specific components. The de/DGS + en/ASL model appears to better handle spacial and possessive concepts. Take one sentence pair, a common sign off from German news stations, as an example: "ihnen noch einen schönen abend und machen sie es gut." (Have a nice evening and do well.). See translations below.

Translations across Models	
Source	Gloss
PHOENIX-14T Reference Translation	SCHOEN ABEND MACHEN GUT poss-EUCH
de/DGS Translation	SCHOEN ABEND MACHEN GUT
de/DGS + en/ASL Translation	SCHOEN ABEND MACHEN GUT poss-EUCH

Table 11: Translation Examples

The “poss-“ is the possessive prefix in DGS and “YOU” indicates who the subject is. In this context, it specifies for "you" (the viewer) to do well. While the de/DGS only model neglected to include many possessives such as these, the parallel de/DGS + en/ASL showed improvement in inclusion of these sign language specific concepts.

These qualitative improvements and our quantitative improvement in BELU scores with de/DGS + en/ASL point to the success of multilingual corpus training for sign languages. Farrugia et al. (2016) establish similarities between ASL and DGS (beyond them both being sign languages) that suggest why a model trained in parallel improved our translations to DGS gloss. These similarities include but are not limited to fingerspelling, spacial descriptors, iconicity (use of icons to represent concepts), and word prefixes. More sophisticated training on these sign language specific concepts in parallel likely improved the model’s ability translate from spoken German to DGS.

Much as in Grave et al. (2020) and Sennrich et al. (2017)’s success with parallel data augmentation, multilingual corpus training aims to increase the coverage of the training data. Because PHOENIX-14T represents a limited dataset for translation, the addition of ASLG-PC12 sentence pairs increases the diversity of training data and improves the model’s ability to create meaningful encodings.

10 Conclusion

Our research provides a baseline for transformer for text to gloss translation. Our results justify a multilingual corpus training approach for sign language translation. Training with the inclusion of the ASLG-PC12 corpus helped inform PHOENIX-14T translation specifically with concepts particular to sign languages. This speaks to the similarities between ASL and DGS, and the potential for multilingual corpus training across all sign languages. While we found success with multilingual corpus training, this does not rule out other approaches, such as dual encoding of syntactic information, pretraining, or transfer learning.

References

Charlotte Baker. 2012. English-asl gloss parallel corpus 2012 (aslg-pc12). *ResearchGate*.

Ondrej Bojar, Christian Buck, Christian Federmann, Barry Haddow, Philipp Koehn, Johannes Leveling, Christof Monz, Pavel Pecina, Matt Post, Herve Saint-Amand, Radu Soricut, Lucia Specia, and Aleš Tamchyna. 2014. Findings of the 2014 workshop on statistical machine translation. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pages 12–58, Baltimore, Maryland, USA. Association for Computational Linguistics.

Necati Cihan Camgoz, Oscar Koller, Simon Hadfield, and Richard Bowden. 2020a. Neural sign language translation. *Papers with Code*.

Necati Cihan Camgoz, Oscar Koller, Simon Hadfield, and Richard Bowden. 2020b. Sign language transformers: Joint end-to-end sign language recognition and translation.

K. Chen, R. Wang, M. Utiyama, E. Sumita, and T. Zhao. 2018. Syntax-directed attention for neural machine translation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Nicolas Farrugia, Irene Sperandio, Victoria J Bourne, and Michael R Ibbotson. 2016. Neural signatures of conscious face perception in an inattentional blindness paradigm. *Journal of Neurophysiology*, 115(2):993–1000.

Edouard Grave, Armand Joulin, Piotr Bojanowski, and Tomas Mikolov. 2020. Unsupervised multilingual word embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3407–3421.

HDI (Human Development Institute) at the University of Kentucky. 2023. A semi-brief fascinating history of american sign language: Resources for learning. *HDI at the University of Kentucky*.

Zaw Zaw Hlaing, Ye Kyaw Thu, Thepchai Supnithi, and Punnarai Netisopakul. 2022. Improving neural machine translation with pos-tag features for low-resource language pairs. volume 8, page e10375.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spacy: Industrial-strength natural language processing in python. If you use spaCy, please cite it as below.

Jana Hosemann and Annika Herrmann. 2014. Language typology: A multimodal investigation on the syntax of placement and removal events. *BMC Neuroscience*, 15:62.

NIDCD. 2019. American sign language (asl). *National Institute on Deafness and Other Communication Disorders*.

Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. 2019. fairseq: A fast, extensible toolkit for sequence modeling. *arXiv preprint arXiv:1901.07291*.

Yirong Pan, Xiao Li, Yating Yang, and Rui Dong. 2020. Multi-source neural model for machine translation of agglutinative language. volume 12, page 96.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2017. Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96.

Stephanie Stoll, Necati Camgoz, Simon Hadfield, and Richard Bowden. 2018. Sign language production using neural machine translation and generative adversarial networks.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2023. Attention is all you need.

Kayo Yin and Jesse Read. 2020. Better sign language translation with stmc-transformer.

Fuzi Zhang, Congying Xia, Bowen Liu, Yu Zhang, and Jie Tang. 2020. A theoretical analysis of the repetition problem in neural language generation. *arXiv preprint arXiv:2012.14660*.