International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Text-to-ISL Gloss Translation: Bridging Language Accessibility Through NLP

Geetha Ramani R [1], Shanmukhaa M S [2]

[1] Professor, Department of Information Technology, MIT-Anna University.
[2] Student, Department of Information Technology, MIT-Anna University.

Abstract

The lack of representation of Indian Sign Language (ISL) in technological innovation hinders communication for India's Deaf community. This research addresses this problem by introducing a system for transcribing English sentences into ISL gloss, a formalized textual representation of ISL grammar. The proposed approach utilizes a Sequence-to-Sequence (seq2seq) model with a Bidirectional LSTM (BiLSTM) encoder and an LSTM decoder to capture the mapping of English syntax to ISL gloss structure. Due to the absence of public English-ISL gloss datasets, a specialized corpus of 10368 sentence pairs was created in collaboration with ISL experts. The system's performance was evaluated using BLEU and ROUGE-L F1 scores on a test set of 2074 sentences, yielding encouraging results under various training settings. This work provides a foundation for future studies in ISL video generation and its integration into accessible education and communication devices, ultimately aiming for enhanced digital accessibility and inclusive technology.

Keywords: Indian Sign Language, Gloss Translation, Seq2Seq, BiLSTM, NLP, Accessibility

1. Introduction

Language plays an essential role as a medium for communication, acquisition of information, and interaction between people. In the lives of those Deaf or hard of hearing, the use of sign language comes naturally as an access point for basic communication. A large proportion of the Indian population that uses the Indian Sign Language (ISL) finds no representation even after its role becomes crucial for creating inclusive access of resources and avenues of communication.

ISL is a complete visual language with its own syntax, grammar, and semantic organization, different from English as spoken or written. The ISL gloss is the written notation for ISL and acts as a bridge from text input to sign language output. ISL gloss is often verb-final in construction, in the present tense, and it excludes inflectional markers, unlike English. These variations imply that word-for-word translation from English to ISL is impossible. Rather, sentence structures have to be subjected to grammatical conversion to fit the linguistic conventions of ISL.

Notwithstanding the extensive use of machine translation (MT) methods in closing communication gaps between spoken languages, their application in sign language processing is still in the initial phases of research and development. Converting English text into ISL requires two important steps: (1) transforming the text into ISL gloss, and (2) interpreting gloss into relevant sign language video. This paper deals with the first step — generation of gloss — that plays a pivotal role in creating precise and grammatically correct sign language representations.

One of the main difficulties here is that no publicly available English-ISL gloss datasets exist, which has tremendously hindered work in ISL-related machine translation systems. For this purpose, we partnered with ISL educators and linguists to design a tailored dataset suitable for the linguistic characteristics of ISL. This data was then employed to train a Sequence-to-Sequence (seq2seq) neural model that makes use of a Bidirectional Long Short-Term Memory (BiLSTM) encoder and an LSTM decoder for the task of grammatical transformation from English to ISL gloss.

The primary goal of this research is to build a system that can generate ISL gloss representations from English input, thus facilitating the development of sign language translation software, educational materials, and accessible communication portals. The envisioned system does not just assist Deaf people in retrieving textual content more autonomously but also facilitates non-signers' understanding of the structural component of ISL. In addition, this work sets the foundation for further building end-to-end English-to-ISL video translation systems, towards the ultimate objective of digital accessibility and inclusive technology.

2. Related Work

In comparison to other languages, ISL has some distinct characteristics. Vasishta et al. [1] established this through linguistic studies within four main Indian cities. The works by Zeshan et al. [2] also shed light on the grammatical features of ISL. Simple English sentences adhere to SubjectVerb- Object format, while ISL follows Subject- Object- Verb format. Also, ISL sentences contain words in their root form, and it avoids using linking verbs, inflection, etc. Question words and negations are generally placed at the end of the sentences. Typically, ISL sentences will be shorter in length, and words will be in the present tense. Further details on the structure of ISL sentences can be found in [3]. A few example sentences are shown in Table I.

The process of utilising computers to automatically convert one natural language to another is referred to as machine translation. A detailed review of machine translation systems for generating sign language from text is presented by Kahlonet al. in [4]. They highlighted various machine translation techniques, including rule-based, corpus-based, neural machine translation-based, etc. Rule-based approaches are classified as direct, interlingua-based, and transfer-based. The earlier systems were based on sign language grammar rules, and they were mostly domain specific. Rule-based approaches are still popular, as there is a need for more publicly available datasets and other resources needed for performing data-driven and neural machine translation approaches in many sign languages.

Corpus-based approaches depend on bilingual parallel corpus. Corpus-based or data-driven techniques are further subdivided into example-based, statistical, and hybrid machine translation. Neural Machine Translation techniques are based on artificial neural networks that translate one language to another with the help of large datasets. For speech to ISL conversion, a frame-based system using ISL grammar rules was developed by Anuja et al. [5]. Prerecorded motion caption data was used by them to render a 3D avatar after passing the sentences through a phrase reordering module. Dasgupta et al. [6] also developed a rule-based system for transforming English to ISL.

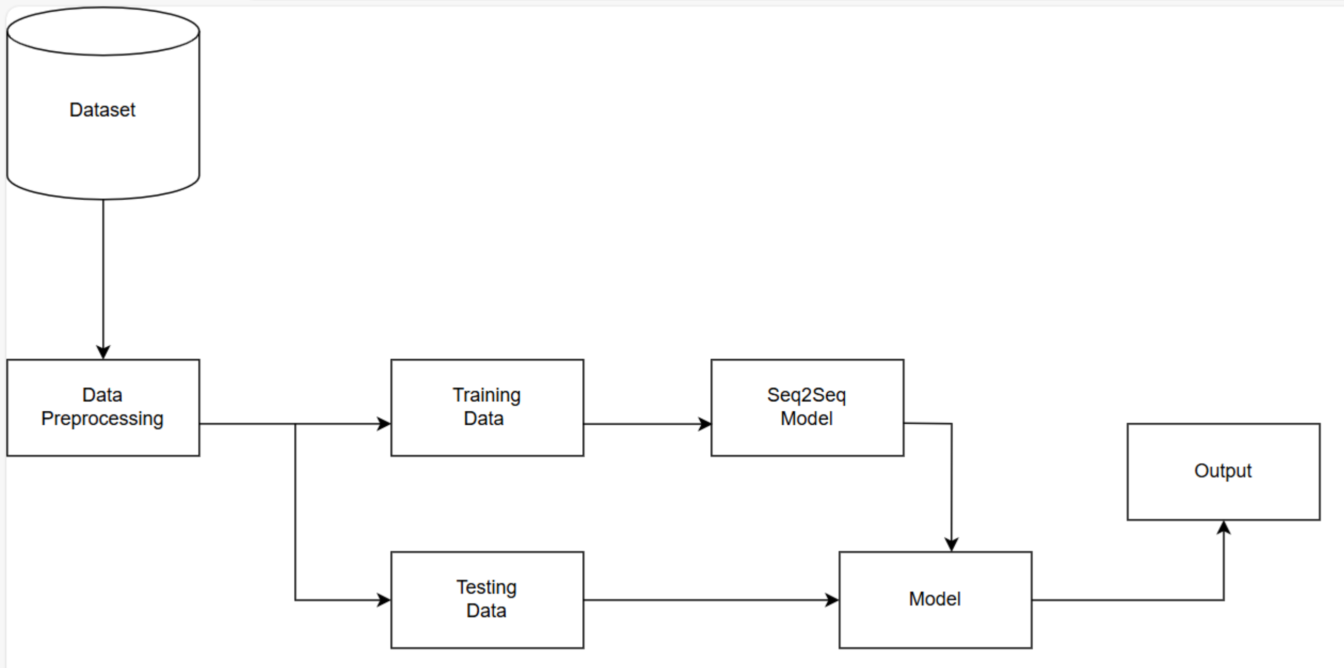
Transfer grammar rules were used for structure conversion and pre-recorded videos were used to generate the sign output. INGIT [7] was developed to translate Hindi to ISL in the railway reservation domain. For ISL gloss generation, they used an ellipsis resolution module and Fluid Construction Grammar (FCG). Kumar et al. [8] proposed a system for converting Hindi and English to ISL. They used ISL grammar rules and mapped sentences to ISL structure and then mapped each word to its HamNoSys [9] representation. Signing Gesture Markup Language(SiGML) [10] files were generated from HamNoSys and were used for driving the Avatar controller. Sign4PSL [11] was developed as a system to convert text to Pakistani Sign Language. They also used HamNoSys-based animation generation.

With the advancements in computational resources and the availability of data on a large scale, Neural Machine Translation(NMT) techniques are also gaining popularity in recent times. There are different types of NMT models, such as Recurrent Neural Networks(RNNs), Long Short-Term Memory(LSTM), and Gated Recurrent Unit(GRU) based models [12] [13]. The use of neural networks for sequence to sequence learning was first explored by Sutskever et al. in [14]. It paved the way for end-to-end approaches to machine translation. A feed-forward back-propagation artificial neural network was used by Brour et al. in [15] to create an Arabic text to Arabic sign language translation system. An attentionbased sequence-to-sequence approach was used in [16] for translating ASL Gloss to English. A bidirectional system for English to ASL gloss and vice-versa was proposed in [17]. They also depended upon attention-based encoder-decoder neural network model. In [18] Vasani et al. presented a system for ISL generation using generative adversarial networks.

3. System Architecture

Fig. 1 depicts the proposed system architecture. The following subsections explains the dataset used for the task and the proposed methodology.

Figure 1: System architecture



3.1 DATASET PREPARATION

Constructs an English–ISL gloss parallel dataset adhering to ISL grammar.

Input:

- English sentences from the Tatoeba Project
- Manually crafted ISL gloss translations

Process:

1. Select 10,368 English sentences
2. Translate using ISL grammar rules (SOV, root word simplification)
3. Store as tab-separated sentence pairs
4. Validate with ISL experts
5. Incorporate expert feedback and corrections

Output:

- Parallel dataset of 10,368 English–ISL gloss pairs

3.2 DATA PREPROCESSING

Prepares the dataset for Seq2Seq model training by cleaning and tokenizing.

Input:

- Raw English–ISL gloss sentence pairs

Process:

1. Convert text to lowercase
2. Remove punctuation and digits
3. Expand English contractions
4. Clean whitespace and tokenize
5. Append SOS and EOS tokens to gloss sentences

International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

6. Convert tokens to integers using TensorFlow tokenizer
7. Pad sequences to equal length
8. Calculate vocabulary sizes
9. Split into 80% training and 20% testing sets

Output:

- Tokenized, padded training and test data

3.3 TRAINING DATA PROCESSING

Trains the Seq2Seq model to learn English-to-ISL translation using supervised learning.

Input:

- Tokenized training set (~8,294 sentence pairs)

Process:

1. Feed English sequences into BiLSTM encoder
2. Initialize LSTM decoder with encoder's final state
3. Feed ISL gloss (with SOS) to decoder
4. Apply teacher forcing
5. Compute loss using categorical cross-entropy
6. Use Adam optimizer for weight updates
7. Enable early stopping based on validation loss

Output:

- Trained Seq2Seq model capable of English-to-ISL translation

3.4 SEQ2SEQ MODEL ARCHITECTURE

Defines the encoder-decoder architecture with attention for gloss generation.

Input:

- Tokenized and preprocessed training sequences

Process:

1. Use embedding layer to convert tokens to 1000-d vectors
2. Process embeddings with BiLSTM encoder
3. Initialize decoder with encoder's final state
4. At each decoding step:
    a) Embed previous token
    b) Pass through LSTM
    c) Apply attention for context
    d) Combine context and hidden state
    e) Predict gloss token using softmax
5. Apply cross-entropy loss
6. Use early stopping

Output:

- Final trained Seq2Seq model with attention

3.5 INFERENCE AND GLOSS PREDICTION

Generates ISL gloss from new English sentences using the trained model.

Input:

- Tokenized English sentence
- Trained encoder-decoder model

Process:

1. Encode input sentence to get context vector
2. Start decoder with SOS token
3. At each step:
   a) Predict gloss token using softmax
   b) Feed token into next decoding step
4. Stop at EOS token or max length
5. Convert token IDs to gloss words

Output:

- Final ISL gloss sentence for avatar rendering or educational tools

4. Results

To convert English text into grammatically valid Indian Sign Language (ISL) gloss using a transformer-based sequence-to-sequence model trained on a parallel English–ISL dataset.

Result Metrics

The performance of the gloss translation model was evaluated using standard sequence generation metrics such as BLEU Score, ROUGE-L F1 Score, and training/validation cross-entropy loss. Results were compared against a baseline LSTM model to validate improvements. These metrics provide insight into the model's accuracy in reproducing target glosses and its ability to retain long-range dependencies and ISL-specific grammar.

Observations

The model was trained on a curated dataset containing 10,368 English–ISL gloss pairs, and its effectiveness was measured using BLEU and ROUGE metrics. The system achieved a BLEU score of 78.04%, which indicates strong alignment between predicted and reference gloss sentences. Compared to the baseline LSTM model, the transformer-based Seq2Seq architecture consistently outperformed in preserving long-form structure and semantic clarity.

Over the span of three training epochs, the model demonstrated stable convergence. The training loss reduced from 0.0014 in epoch 1 to 0.0002 by epoch 3, while the validation loss declined from 0.000411 to 0.000306, as shown in Table 5.6. The final average validation loss of approximately 0.00025 indicates minimal overfitting and good generalization. These values are further summarized in Table 5.5.

The system also showed strong adherence to ISL grammar—particularly the subject-object-verb (SOV) structure and the omission of auxiliary verbs. Despite the modest dataset size, the model performed well across various domains, successfully translating formal English phrases from departments like Finance, Education, and Culture into semantically accurate gloss. This makes the model well-suited for real-time applications such as ISL avatar rendering and government communication accessibility.

Table 1: Performance Metrics for Text-to-ISL Gloss Translation Module

Metric	Value
Training Loss	0.0015 → 0.0003
Validation Loss	~0.00025
BLEU Score	78.04%

Table 2**:** Training and Validation Loss across Epochs

Epoch	Training Loss	Validation Loss
1	0.001400	0.000411
2	0.000400	0.000328
3	0.000200	0.000306

5. Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this research. The research work titled *"Text-to-ISL Gloss Translation: Bridging Language Accessibility Through NLP"* was conducted independently without any external sponsorship or influence from any organization or individual. The design, methodology, results, and interpretations presented are solely the outcome of the authors' academic effort and were not affected by any external parties.

6. Acknowledgement

The authors would like to express their sincere gratitude to the organizers of Tamil Nadu Niral Thiruvizha for providing a valuable platform that encouraged technological innovation. The event served as a catalyst for exploring real-world challenges and fostering impactful research directions. The authors also acknowledge the support and encouragement received from their institution and faculty members, which played a significant role in the successful completion of this work.

References

1. Woodward J., Wilson K., “Sign Language in India: Regional Variation within the Deaf Population”, Indian Journal of Applied Linguistics, 1978, 4 (2), 66–74.
2. Zeshan U., “Indo-Pakistani Sign Language Grammar: A Typological Outline”, Sign Language Studies, 2003, –, 157–212.
3. Nair M.S., Idicula S.M., “English to Indian Sign Language Gloss Conversion Using a Rule-Based Approach”, ICCCT 2022, Springer, 2022, –, 737–744.
4. Kahlon N.K., Singh W., “Machine Translation from Text to Sign Language: A Systematic Review”, Universal Access in the Information Society, 2021, –, 1–35.
5. Anuja K., Suryapriya S., Idicula S.M., “Design and Development of a Frame-Based MT System for English-to-ISL”, NaBIC 2009, IEEE, 2009, –, 1382–1387.

6. Dasgupta T., Basu A., Bhowmick P.K., Mitra P., “A Framework for the Automatic Generation of Indian Sign Language”, Journal of Intelligent Systems, 2010, 19 (2), 125–144.
7. Kar P., Reddy M., Mukherjee A., Raina A.M., “Ingit: Limited Domain Formulaic Translation from Hindi Strings to Indian Sign Language”, ICON, –, 52, 53–54.
8. Kumar P., Kaur S., “Indian Sign Language Generation System”, Computer, 2021, 54 (3), 37–46.
9. Hanke T., “HamNoSys—Representing Sign Language Data in Language Resources and Language Processing Contexts”, LREC 4, 2004, –, 1–6.
10. Elliott R., Glauert J., Jennings V., Kennaway R., “An Overview of the SiGML Notation and SiGMLSigning Software System”, signlang@LREC 2004, 2004, –, 98–104.
11. Sanaullah M., et al., “A Real-Time Automatic Translation of Text to Sign Language”, Computers, Materials and Continua, 2022, 70 (2), 2471–2488.
12. Sherstinsky A., “Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network”, Physica D: Nonlinear Phenomena, 2020, 404, 132306.
13. Hochreiter S., Schmidhuber J., “Long Short-Term Memory”, Neural Computation, 1997, 9 (8), 1735–1780.
14. Sutskever I., Vinyals O., Le Q.V., “Sequence to Sequence Learning with Neural Networks”, Advances in Neural Information Processing Systems, 2014, 27, –.
15. Brour M., Benabbou A., “ATLASLang MTS 1: Arabic Text Language into Arabic Sign Language Machine Translation System”, Procedia Computer Science, 2019, 148, 236–245.
16. Arvanitis N., Constantinopoulos C., Kosmopoulos D., “Translation of Sign Language Glosses to Text Using Sequence to Sequence Attention Models”, SITIS 2019, IEEE, 2019, –, 296–302.
17. Amin M., Hefny H., Ammar M., “Sign Language Gloss Translation Using Deep Learning Models”, International Journal of Advanced Computer Science and Applications, 2021, 12 (11), –.
18. Vasani N., Autee P., Kalyani S., Karani R., “Generation of Indian Sign Language by Sentence Processing and Generative Adversarial Networks”, ICISS 2020, IEEE, 2020, –, 1250–1255.
19. Lin C.Y., “ROUGE: A Package for Automatic Evaluation of Summaries”, Text Summarization Branches Out, 2004, –, –.
20. Papineni K., et al., “BLEU: A Method for Automatic Evaluation of Machine Translation”, Proceedings of the 40th Annual Meeting of the ACL, 2002, –, 311–318.