Hindi To Indian Sign Language: Rule-Based Translation System

Yashvir Singh Nathawat Kartik Jain Aditya Katare

231110059, 231110023, 231110005

{ysnathawat23,kartikj23,kaditya23}@iitk.ac.in

Indian Institute of Technology Kanpur (IIT Kanpur)

ABSTRACT

Sign language is a crucial mode of communication for the deaf community. Researchers around the world have taken on the linguistic challenge of developing systems to facilitate understanding natural language through sign language. This project discusses a model developed primarily to translate Hindi text into Indian Sign Language. The system utilizes a dependency parser and part-of-speech tagger to accurately categorize the input words into their syntactic forms. Subsequently, the WordNet knowledge base is employed to enhance the system's efficiency. The project highlights the pivotal role of WordNet in expanding the scope of the dictionary, which serves as an integral component in constructing the Indian Sign Language system. By harnessing natural language processing techniques like dependency parsing, part-of-speech tagging, and leveraging the WordNet database, this system aims to bridge the gap between written Hindi and Indian Sign Language, enabling more effective communication for the deaf community.

Keywords: Hindi, Indian Sign Language , Translation, ISL, WordNet

1 INTRODUCTION

Sign languages are really important for people who can't hear well. There are over 300 different sign languages all over the world, and each one is made to fit the language and culture of the area where it's used. In India, there are about 4 million people who are deaf and 10 million who have trouble hearing. About 1 million deaf adults and 0.5 million deaf kids in India use Indian Sign Language (ISL) to talk. But there are around 2.5 million deaf and hard of hearing people in India who don't have any way to learn sign language, so it's hard for them to communicate. This makes it tough for them to learn things and get information, and about 80% of people with hearing problems face these challenges.

Sign language is a way of talking that uses hand shapes, facial expressions, and body movements instead of words. Even though it's visual, sign languages have their own grammar and rules. In India, ISL is the main sign language used.

Using computers to translate sign language has become a popular way to help people communicate, especially for languages like ISL that don't have a lot of resources or books. These computer systems use rules and language structures to make sure the translation is accurate and easy to understand. This makes it easier for deaf and hard of hearing people to get information and learn new things.

Making these rule-based translation systems for sign languages is a big step in helping deaf and hard of hearing people. By using technology, we can break down communication barriers and make sure everyone has the same chance to learn, get information, and have opportunities, no matter if they can hear well or not.

2 RELATED WORK

A lot of research is going into Natural Language Processing (NLP). Not enough work has been done in converting Hindi to Indian Sign Language. But there is some work for converting English to Indian Sign Language, some of them are (Agarwal et al., 2015) and (Mishra, 2019) which are given in reference section

3 PROPOSED IDEA

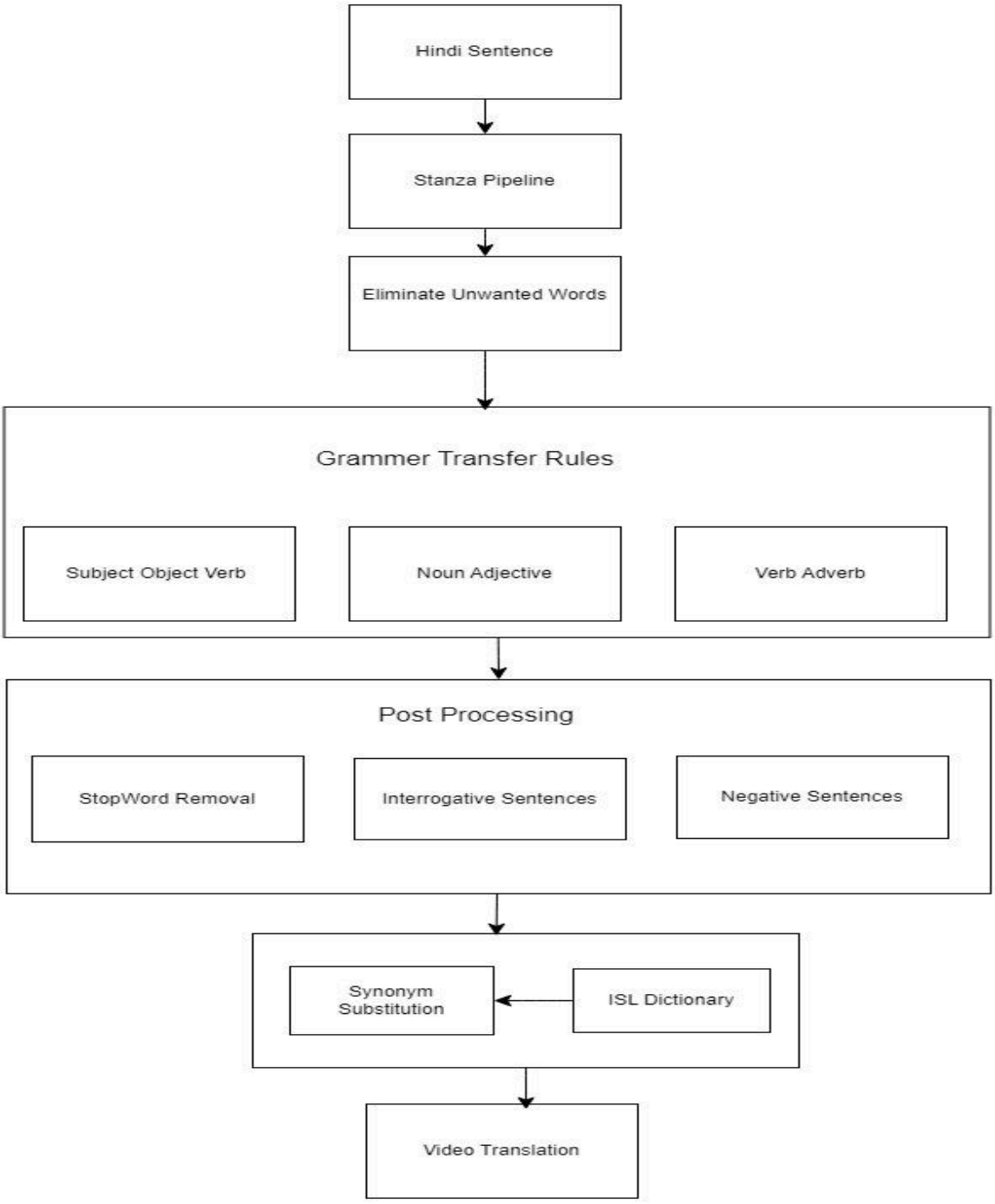
We're planning to create a system that translates Hindi text into Indian Sign Language. The process begins with segmenting sentences based on Part-of-Speech (POS) tagging using the stanza pipeline. Next, the segmented words are arranged according to grammar rules, removing stop words and lemmatizing the remaining words to simplify the vocabulary.

After arranging the sentence, each word will be matched with a video showing how to sign that word in sign language. If there's no video available, we'll try to find similar words in a database, if that also doesn't work, we'll spell out the word using finger signs.

After all words are mapped, a generated video will appear on screen showing signs in order based on grammar rules of ISL.

4 METHODOLOGY

4.1 Development of Hindi to ISL Translation The translation based system that we built uses following architecture:



4.2 POS Tagging and Dependency Parser - Stanza

First we use Stanza Hindi Pipeline for doing Tokenization, POS Tagging, Dependency Parsing.

4.3 Eliminate Unwanted Words

We remove some words with the following POS tags as they have no meaning in ISL Dictionary. Some of these words are:

- VAUX (Auxiliary verb) A verb that accompanies the main verb to convey grammatical distinctions.
- CC (Coordinating conjunction) - A word that connects words, phrases, or clauses of equal importance.
- SYM (Symbol) - A character or sequence of characters that represent a concept, idea, or entity.
- PSP (Postposition): A word that comes after its complement to express grammatical relationships.

4.4 Grammer Transfer Rules

Below are the Grammar Rules that we used to reorder the given hindi sentences :

Grammar Rule	Transformation	Example
VERB(root) SUBJECT OBJECT	SUB OBJ VERB	तुम पुस्तक पढ़ते हो। → तुम पुस्तक पढ़ते
VERB(any) → ADV (advmod)	VERB ADV	मैं जल्दी आऊँगा। → मैं आऊँगा जल्दी
NOUN(any) → ADJ(amod)	NOUN ADJ	सीता चमकता तारा है। → सीता तारा चमकता

4.5 Post Processing

From above stages we have a nearly stable ordered sentence which follows Indian Sign Grammer. Now we handle some special cases which are as follows :

4.5.1 StopWord Removal

Following the aforementioned processing, it is necessary to refine the sentence. ISL omits certain words such as articles and other functional terms that lack significant meaning. Through iterative examination of the sentences, we have compiled a definitive list of stopwords. This process is conducted manually, assessing the utility of each word.

At last, the total number of stopwords is 72.

4.5.2 Interrogative

Interrogative sentences are indicated in sign language by initially expressing the signs of the sentence in imperative form, followed by the inclusion of a question word. These question words typically lack specific dependency roles and can manifest in various relations depending on the sentence type. Additionally, discerning whether a wh-word in a sentence functions as a question or not can pose significant challenges.

Example : तुम कितने अच्छे हो. (Not interrogative)

कितने आदमी थे? (Interrogative).

It is difficult to handle such cases, what we have done is we just pick it up and place it at the end of sentence.

4.5.3 Negative Cases

Negative sentences are sentences which have नहीं (nahīn) or ना (nā). Same strategy is followed for negative i.e We take the negative word and place it at the end of the sentence.

Sentence : वह उस परीक्षा में सफल नहीं हुआ।

ISL Translation : वह उस परीक्षा सफल हुआ नहीं

Interrogative sentences and negative sentences are handled the same way in the ISL. But if both are included in the sentences like sentence -

Sentence : क्या तुम्हें यह नहीं पसंद है?

ISL Translation : तुम्हें यह पसंद नहीं क्या

4.6 Synonym Substitution

Synonym Substitution is one of the crucial steps of the whole system. We are mapping each filtered word of the ISL to the given list of ISL dictionary.

4.6.1 ISL Dictionary Creation : We extracted all the words of ISL videos and translated each word to its corresponding Hindi word.

4.6.2 Word to ISL Dictionary Mapping : For each hindi word(hindi_word) we have translated to english word(english_word) using (*Google Translation Api*). Also we have used *spacy lemmatizer* for lemmatizing the english word (lemmatized_english_word). We have made following cases :

Case 1: Word is Proper Noun then FingerSpell it

Case 2: Hindi Word in ISL Dictionary

Case 3: Synonym of Hindi Word in ISL Dictionary

Case 4: English Word in ISL Dictionary

Case 5: Lemmatized English Word in ISL Dictionary

Case 6: Synonym of English Word in ISL Dictionary

Case 7: FingerSpell

4.7 Video Translation

After mapping each word to its corresponding ISL entry, we combine all the videos into a single video and then run it. To merge the videos, we utilize MoviePy editor in Python.

For words that are not mapped to any ISL entries (which are very few, thanks to our Synonym Substitution step), we handle them through fingerspelling.

Fingerspelling is a technique used in sign languages where individual letters of a word are spelled out using handshapes corresponding to each letter. It's often used for proper

nouns, technical terms, or words for which there is no established sign. In the context of our sentence, if there are words not represented in the ISL dictionary, they would be fingerspelled in the video to ensure complete communication.

5 DATASET USED

We have used Indian Sign Language (ISL) videos obtained from Google Drive for video translation. These videos demonstrate sign language interpretations for different words.

Own Crafted Hindi Letters Videos: To illustrate the Hand Mapping Task further, we've crafted our own set of videos. These videos showcase the signing of individual characters in Hindi. Using a laptop webcam, we've recorded 50 videos, each dedicated to a specific character in the Hindi alphabet, encompassing both consonants and vowels. These videos are instrumental in mapping hand gestures for words lacking ISL video representations or their equivalents.

Own Crafted Dataset: For testing, we've assembled a corpus containing various sentences with different grammatical structures. This corpus consists of 150 categorized sentences along with approximately 70 additional random sentences. We analyze our code against all these sentences for testing purposes.

Own Crafted English Letters Videos: For Hand Mapping we also recorded 26 videos corresponding to each English alphabet.

6 RESULTS

CATEGORIES (Number of Sentences)	ADV + ADJ (13)	ONE SUB+ VERB (20)	SUB + OBJ+ VERB (19)	VERB +ADV (15)	ONE VERB ONE ADJ (20)	SUB OBJ ADJ (14)	SUB+ OBJ+ ADV (10)	INTERROGATIVE SENTENCES (20)	NEGATIVE SENTENCE (19)
Accuracy(%)	76.92	100	100	93.33	90.0	85.71	100	100	94.73
No. of FingerSpells Required	8	3	1	5	4	3	7	12	6

Table 1. Performance based on database built of 150 categorized sentences

RANDOM SENTENCES

Out of the 67 random sentences sourced from Kaggle, our evaluation indicates that the code performs effectively on 58 sentences. However, there are 9 sentences where the intended meaning is altered due to the code's performance. Additionally, in 16 sentences, we need to resort to finger spelling

ACCURACY:86.56%

NOTE:ALL SENTENCES AND THEIR CORRESPONDING OUTPUTS ARE GIVEN IN ANOTHER PDF FILE NAMED test_result.

7 Discussion and Future Work

1) If a bigger list of Hindi root words is present, more word-lemma mappings with context can be generated to create a bigger dataset using similar techniques as mentioned in the report.

2) The rule based lemmatizer developed here can act as a baseline using which more complex Deep Learning based techniques can be explored.

3) Using animations instead of videos allows for a larger dataset, reducing the need for finger spelling. This approach offers flexibility in gesture customization and seamless integration into interactive applications, enhancing sign language interpretation and education.

4) Finding a reliable Hindi lemmatizer remains a challenge. However, with an effective Hindi lemmatizer, synonym substitution can be significantly improved in terms of accuracy.

5) Interrogative Sentence : wh-words sometimes signify a question and maytimes a not. Deciding when it signifies will be helpful in translation task

6) ISL Videos : ISL is a growing community. Adding more videos to the system will significantly increase the translation task.

7) Grammar Rules : Adding more complex grammar rules such as how to multiple clauses and relative clauses.

8 References

[1] English To Indian Sign Language: Rule-Based Translation System Along With Multi-Word Expressions and Synonym Substitution.

[2] Sumeet R Agarwal, Sagarkumar B Agrawal, and Akhtar M Latif. 2015. Sentence formation in nlp engine on the basis of indian sign language using hand gestures. International Journal of Computer Applications, 116(17).

[3] William C. Jr. Stokoe. 1960. Sign language structure, an outline of the visual communications systems of american deaf. Studies in Linguistics Occasional paper, 8.

[4] Indian Sign Language As A Complete Language: Concept, Characteristics And Common Misunderstandings About Sign Language - https://www.nios.ac.in/media/documents/230-ISL/Ch-1.pdf

[7] Charlotte Brewer. 1993. The second edition of the oxford english dictionary. The Review of English Studies, 44(175):313–342.

[8] Madan Vasishta, James Woodward, and Susan De Santis. 1980. An introduction to Indian sign language:(Focus on Delhi). All India Federation of the Deaf

[9] Indian Sign Language which is made public by ISLRTC (Indian Sign Language Research And Training Center) link is given as follows: https://drive.google.com/drive/folders/1U-Pr4r1-cupgNOOq9NH_uTsQnPSVEKco

[10] Gouri Sankar Mishra. 2019. English text to indian sign language translation system. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 8(10S):460–467