# Generating Indian Sign Language Videos from Hindi using Rule-Based Translation

**Anonymous ACL submission**

## Abstract

Sign language serves as a vital mode of communication for the deaf community. Recognizing its importance, researchers worldwide have undertaken the linguistic challenge of bridging the gap between spoken and signed languages. This paper focuses on the development of a model designed to translate Hindi text into Indian Sign Language (ISL). The system takes Hindi sentences as input and find dependency relationship among words and use this dependency relationship to make rigid word order. To further improve the effectiveness of the system, we use synonym substitution module which uses English and Hindi WordNet and lemmatizer to get the most suitable substitute word whose corresponding video is in the ISL dictionary. Individual videos are merged to provide the final video.

## 1 Introduction

Sign languages are indispensable for individuals with hearing impairments, and serve as their primary mode of communication. Globally, there are over 300 distinct sign languages, each uniquely adapted to the linguistic and cultural contexts of their regions. In India, the situation is particularly pressing, with approximately 4 million people who are deaf and an additional 10 million who have hearing difficulties (Zeshan et al., 2005). Indian Sign Language (ISL) is utilized by around 1 million deaf adults and 0.5 million deaf children for communication. Despite this, about 2.5 million deaf and hard-of-hearing individuals in India lack access to sign language education, making effective communication a significant challenge for them.

*Indian Sign Language (ISL)* is a unique visual-gestural language characterized by hand shapes, facial expressions, and body movements. It comprises a dictionary of over 10,000 words, distinct from other sign languages worldwide. It has its own grammar rules. In ISL, adjectives follow nouns, such as in "आकाश नीला" (ākāśa nīlā) (sky blue), and linking verbs like **(हूँ, है, हैं, था, थे)** (hūm, hai, haiṁ, thā,thē) (am, is, are, was, were) are omitted. ISL also avoids word endings and suffixes, maintaining only present tense and omitting articles like **(एक, कुछ, को)** (ēka, kucha, kō) (a, some, to). Wh-questions appear at the end of sentences, such as "हम खाते कहां" (hama khātē kahāṁ) (we eat where). Expressions, including eyebrow and eyelid movements, facial expressions, and head and shoulder movements, play a crucial role in ISL communication. Importantly, ISL naturally developed within the deaf community, rather than being artificially created by hearing individuals.

The primary objective of this paper is to bridge the communication gap for millions of deaf individuals who lack knowledge of sign language by providing an innovative solution: *converting Hindi sentences into Indian Sign Language (ISL) videos.* This is the contribution of our work.

We take Hindi sentences as input and generate a video that combines individual ISL videos, each corresponding to a specific Hindi word. These videos are arranged according to ISL grammatical rules. By doing so, we aim to empower deaf individuals to access information more easily and learn ISL effectively. This system also serves as a vital medium for facilitating communication between deaf and non-deaf individuals, thereby fostering social inclusivity and accessibility to information. We adopt a rule-based approach for converting words into ISL videos.

1

## 2 Related Work

To the best of our knowledge, not much research has been done on Hindi to Indian Sign Language translation-based systems. Below are a few selected models. The system described in (Vij and Kumar, 2016) uses dependency parser, and then uses grammar rules. They have not addressed anything about adjective and adverb grammar rules. Also, this system uses a WordNet (Miller, 1992) based Synset and chooses one of the synonyms which may not be related to the original sentence. INGIT (Kar et al., 2006), a cross-modal translation system converting Hindi text to Indian Sign Language (ISL) for use at Indian Railways reservation counters. It utilizes a semantically mediated, formulaic framework for mapping Hindi to ISL, incorporating Construction Grammar to handle various input constructions. The system's effectiveness is validated through a small corpus tested by native ISL signers.

## 3 Methodology

### 3.1 ISL Videos

For the proposed system, we utilized Indian Sign Language (ISL) videos obtained from the Indian Sign Language Research and Training Center (ISLRTC) (http://islrtc.nic.in/) for video translation. These videos provide sign language interpretations for various words and represent hand mapping of various English words. When we have an English word, the corresponding video will be added. However, since we are working with Hindi, we translate all these English words (for which videos are available) into their Hindi equivalents with the same meanings.

### 3.2 Character Videos

Additionally, since the ISLRTC data did not include corresponding videos for English and Hindi alphabets, we decided to create our own set of videos. These videos demonstrate the signing of individual characters in Hindi. We recorded 50 videos, each dedicated to a specific character in the Hindi alphabet, including both consonants and vowels. These videos are crucial for mapping hand gestures for words that lack ISL video representations or their equivalents using finger spelling. For
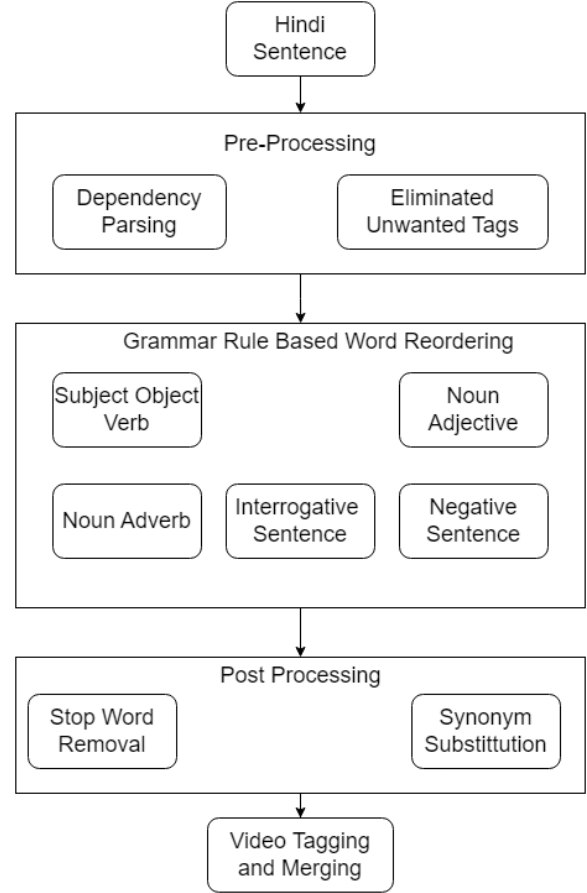


Figure 1: Architecture of Hindi to Indian Sign Language Translation System

the finger-spelling of personal pronouns, we recorded 26 videos corresponding to each English character.

### 3.3 Phases in Video Generation

We have segmented the process into five phases: Pre-Processing, Grammar Transformation, Stopword Removal, Synonym Substitution, and Video Translation. We utilize the Stanza Hindi pipeline (Qi et al., 2020) for tokenization, dependency parsing, lemmatization and POS tagging. The overall system architecture is shown in Figure 1.

### 3.4 Pre-Processing

We first extract parts-of-speech (POS) tags from the Stanza pipeline. We then remove words with the following POS tags as they have no meaning in the ISL Dictionary: VAUX (Auxiliary Verb), CC (Coordinating Conjunction), SYM (Symbol), PSP (Postposition).

For example, in the sentence वह स्कूल जा रहा है। (vaha skūla jā rahā hai, He is going to school)

the auxiliary verb रहा है (rahā hai) is removed during preprocessing, leaving the sentence as वह स्कूल जा (vaha skūla jā) (He school go).

## 3.5 Word Reordering

Unlike Hindi and other Indian languages, which are relatively free-order languages, ISL is a rigid-order language. This means the sequence of words in a sentence is fixed and must be adhered to for clear communication. Further, ISL is not just about gestures and pantomimes; it has its own grammar and rules. In this phase, we change the Hindi sentences to follow the grammar and rules of ISL. We have implemented six rules for the same:

### 3.5.1 Subject Object Verb

In Hindi, the relative order of subject, object and verb can be different for the same sentence. For example, consider the sentence "Ram sees Sita" in Hindi. All the following word orders are correct:
- Word Order 1: राम ने सीता को देखा। (rāma nē sītā kō dēkhā) (Subject-Object-Verb)
- Word Order 2: सीता को राम ने देखा। (sītā kō rāma nē dēkhā) (Object-Subject-Verb)
- Word Order 3: देखा राम ने सीता को। (dēkhā rāma nē sītā kō) (Verb-Subject-Object)

Since ISL follows the subject-object-verb (SOV) pattern, we fixed the order to maintain this rigid structure. Thus, the sentence is arranged in Word Order 1 according to the above example.

### 3.5.2 Noun and Adjective Case

Since Hindi is a free-order language, adjectives can be used predicatively (following), and attributively (preceding) as shown below for the sentence "This boy is clever":
- Adjective Order 1: यह लड़का होशियार है। (yaha laṛakā hōśiyāra hai) (Predicatively)
- Adjective Order 2: यह होशियार लड़का है। (yaha hōśiyāra laṛakā hai) (Attributively)

For our purpose, we follow predicative order, i.e., Adjective Order 1.

### 3.5.3 Noun and Adverb Case

In Hindi, adverbs can appear either before or after the verb, depending on the context: (the sentence is "Ram goes slowly")
- Adverb Order 1: राम धीरे चलता है। (rāma dhīrē calatā hai) (Adverb before verb)
- Adverb Order 2: राम चलता है धीरे। (rāma calatā hai dhīrē) (Adverb after verb)

We follow the Adverb Order 2 for our translation system.

### 3.5.4 Negative Sentences

In Hindi, there are mainly three words used to convey negative meaning to a sentence; these are नहीं, मत, ना (nahīṁ, mata, nā). The strategy used for handling these is to position it at the end of sequence as shown below: (the sentence is "I have not seen her so far")

Sentence: मैंने उसे अभी तक नहीं देखा। (maiṁnē usē abhī taka nahīṁ dēkhā)

ISL-Format: मैंने उसे देखा नहीं। (maiṁnē usē dēkhā nahīṁ)

The words अभी तक (abhī taka, so far) are removed as they are stopwords (Sec 3.6).

### 3.5.5 Interrogative Sentences

In Hindi, interrogative words like क्या, कौन, कब, कहाँ, क्यों, कैसे, कितना, किसका, किसे, आदि (kyā, kauna, kaba, kahāṁ, kyōṁ, kaisē, kitanā, kisakā, kisē, ādi) can be used anywhere in a sentence, be it at the beginning, in the middle, or at the end. To provide a rigid order, they are handled the same way as negative sentences in Sec 3.5.4, as explained below: (the sentence is "Have you done your homework?")

Sentence: क्या तुमने अपना होमवर्क किया? (kyā tumanē apanā hōmavarka kiyā)

ISL-Format: तुमने अपना होमवर्क किया क्या? (tumanē apanā hōmavarka kiyā kyā)

Also if there is both a negative word and an interrogative word, then the negative word is placed before the interrogative word as shown below: ("Have you not placed your bag?")

Sentence: कहाँ तुमने अपना बैग नहीं रखा? (kahāṁ tumanē apanā baiga nahīṁ rakhā)

ISL-Format: तुमने अपना बैग रखा नहीं कहाँ? (tumanē apanā baiga rakhā nahīṁ kahāṁ)

### 3.5.6 Multiple Verbs, Adjectives, and Adverbs

Multiple verbs in a sentence are placed in the order they are given. For multiple adjectives, they are placed as per the Stanza (Qi et al., 2020) dependency relation so that relative ordering of noun and its corresponding adjective is fixed. The same strategy is followed for multiple adverbs.

3

## 3.6 Stop Word Removal

In ISL, it is important to simplify sentences by removing unnecessary words that do not add meaning. Through iterative examination of the sentences, we have compiled a definitive list of such *stopwords*. This process is conducted manually by assessing the utility of each word. After rigorous testing and optimization, we have identified a list of 60 stopwords for Hindi.

## 3.7 Synonym Substitution

In this phase, we take each word and check whether it is already in the ISL dictionary. If the word is in dictionary then there is no need for substitution; otherwise, we replace it with the the closest word. Synonym substitution is performed by replacing words with the closest matches in the ISL dictionary, using a combined module of Hindi and English lemmatized forms. For translation from Hindi to English, we used IndicTrans module (Ramesh et al., 2022). Lemmatization for Hindi is done using Stanza (Qi et al., 2020) and for English using Spacy (`https://spacy.io/models/en`). For synonymn substitution task, we made different categories for words which are shown in Figure 2 in Appendix:

- **Category 1:** We analyzed the Hindi language and came out with 25 verbs and their lemmatized forms. If the verb comes out to be in this category, we map it to the corresponding video tag.
- **Category 2:** If the lemma of a Hindi word is in ISL Dictionary, we map it to corresponding video tag.
- **Category 3:** If the Hindi word is not in ISL Dictionary, then we find its Hindi synonym using pyiwn (Panjwani et al., 2018). We map the synonym if found in the ISL Dictionary.
- **Category 4:** We translate the Hindi word to English word using IndicTrans model (Ramesh et al., 2022). The lemma of the English word is mapped to the ISL Dictionary if found.
- **Category 5:** We find the synonyms of English using WordNet (Miller, 1992) and check the synonym with the ISL Dictionary.

| Sentence Category | Accuracy | # Words finger-spelled |
|---|---|---|
| Adverb, Adjective | 76.92 | 8 |
| Subject, Verb | 100.00 | 3 |
| Subject, Object, Verb | 100.00 | 3 |
| Verb, Adverb | 93.33 | 5 |
| Verb, Adjective | 90.00 | 4 |
| Subject, Object, Adjective | 85.71 | 5 |
| Subject, Object, Adverb | 100.00 | 9 |
| Interrogative Sentences | 100.00 | 12 |
| Negative Sentences | 94.73 | 6 |

Table 1: Accuracy and Finger-spelling requirements (in each category, 20 sentences were tested).

## 3.8 Video Tagging and Merging

After synonym substitution, words are mapped to their video tags. All videos corresponding to these words are merged together and shown collectively. Exceptions include proper nouns and words for which no corresponding ISL tag can be found; these are finger-spelled character-by-character.

## 4 Results and Discussion

For testing purposes, we created 9 categories based on sentence grammar. Within each category, we included 20 sentences. We tested all these sentences and observed varying accuracies across different categories (Table 1). Additionally, we calculated the number of sentences that required finger-spelling.

Additionally, we test on 300 random sentences sourced from Kaggle. Our evaluation indicates that the code performs effectively on 259 sentences. For the rest 31 sentences, the intended meaning is altered due to incorrect substitution of synonym word or non-optimal performance of the used tools (Stanza).

Results of synonym substitutions are shown in Table 2 in Appendix.

## 5 Conclusions and Future Work

Our paper presents a pioneering effort to address the communication challenges faced by the deaf community in India by developing a system that converts Hindi sentences into Indian Sign Language (ISL) videos. In this initial phase, we have focused on translating simple Hindi sentences.

In future, we aim to handle complex sentences, other languages and animations.

4

## 6 Limitations

We only cover simple sentences in Hindi. In future, we plan to cover more complex sentences.

## 7 Ethics Statement

We do not see any ethical concerns in this work. The videos generated follow rules and outputs from standard tools.

## References

Purushottam Kar, M. Veera Prathap Reddy, Amitabha Mukerjee, and Achla M. Raina. 2006. Limited domain formulaic translation from hindi strings to indian sign language.

George A. Miller. 1992. WordNet: A lexical database for English. In *Speech and Natural Language: Proceedings of a Workshop Held at Harriman, New York, February 23-26, 1992.*

Ritesh Panjwani, Diptesh Kanojia, and Pushpak Bhattacharyya. 2018. pyiwn: A python based api to access indian language wordnets. In *Global WordNet Conference.*

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 101–108, Online. Association for Computational Linguistics.

Gowtham Ramesh, Sumanth Doddapaneni, Aravinth Bheemaraj, Mayank Jobanputra, Raghavan AK, Ajitesh Sharma, Sujit Sahoo, Harshita Diddee, Mahalakshmi J, Divyanshu Kakwani, Navneet Kumar, Aswin Pradeep, Srihari Nagaraj, Kumar Deepak, Vivek Raghavan, Anoop Kunchukuttan, Pratyush Kumar, and Mitesh Khapra. 2022. Samanantar : The largest publicly available parallel corpora collection for 11 indic languages. *Transactions of the Association for Computational Linguistics*, 10:145–162.

Paras Vij and Parteek Kumar. 2016. Mapping hindi text to indian sign language with extension using wordnet. In *Proceedings of the International Conference on Advances in Information Communication Technology & Computing*, AICTC '16, New York, NY, USA. Association for Computing Machinery.

Ulrike Zeshan, Madan N. Vasishta, and M. J. Sethna. 2005. Implementation of indian sign language in educational settings.
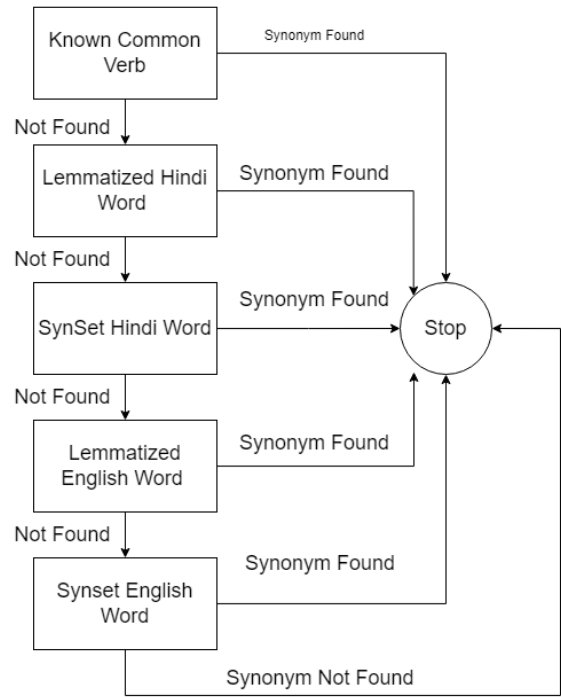
Figure 2: Synonym Substitution Algorithm

| Sentence | Category | After Grammar Rules | Video Played |
|---|---|---|---|
| बड़ी बहन धीरे-धीरे चल रही है। (baṛī bahana dhīrē-dhīrē cala rahī hai) | Adverb, Adjective | बहन बड़ी चल धीरे-धीरे (bahana baṛī cala dhīrē-dhīrē) | Sister Big # # |
| तुम सोते हो। (tuma sōtē hō) | Subject, Verb | तुम सोते (tuma sōtē) | You Sleep |
| वह खिलौना छोड़ता है। (vaha khilaunā chōṛatā hai) | Subject, Object, Verb | वह खिलौना छोड़ता (vaha khilaunā chōṛatā) | He Toy Leave |
| तुम जल्दी सीख लो। (tuma jaldī sīkha lō) | Verb And Adverb | तुम सीख जल्दी (tuma sīkha jaldī) | You Learn Fast |
| गाड़ी तेज चलती है। (gāṛī tēja calatī hai) | Verb And Adjective | गाड़ी तेज चलती (gāṛī tēja calatī) | Vehicle Fast Run |
| राधा सुन्दर फूल लेती है। (rādhā sundara phūla lētī hai) | Subject, Object, Adjective | राधा फूल सुन्दर लेती (rādhā phūla sundara lētī) | @ Flower Beautiful Take |
| उसने कुछ खुशी से किया। (usanē kucha khuśī sē kiyā) | Subject, Object, Adverb | उसने कुछ खुशी किया (usanē kucha khuśī kiyā) | He Some # Do |
| क्या यहाँ कोई बच्चे खेल रहे हैं? (kyā yahām kōī baccē khēla rahē haiṁ) | Interrogative Sentences | कोई बच्चे यहाँ खेल क्या (kōī baccē yahām khēla kyā) | Any Child Here Game What |
| वह उस परीक्षा में सफल नहीं हुआ। (vaha usa parīkṣā mēṁ saphala nahīṁ huā) | Negative Sentences | वह उस परीक्षा सफल हुआ नहीं (vaha usa parīkṣā saphala huā nahīṁ) | He That Exam Success Happen No |

Table 2: Synonymn substitution results (# are finger-spelled and @ are proper nouns which should be finger-spelled)