Paper Presentation on

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

Paper By-
Abhigyan Ghosh
Radhika Mamidi

By-
Yashvir Singh Nathawat (231110059)
Aditya Katare(231110005)
Kartik Jain(231110023)

Paper Published on ACL

Table of Contents

Problem Statement

1. Some Stats
    a. India is home to approximately 63 million people of the deaf
    b. Comparable to total population of France and Italy
2. Need of ISL based system-
    a. Used as Communication Medium between deaf peoples
    b. Used as teaching medium to teach in schools
    c. Severe shortage of ISL interpreters in India
    d. Very Less Work is done compared to ASL
3. Nearly 10000 words in ISL now

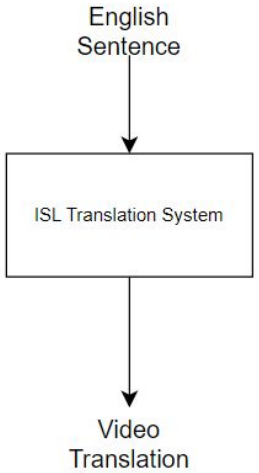


Figure : ISL Translation System

Reference - https://en.wikipedia.org/wiki/Deafness_in_India#:~:text=India%20is%20home%20to%20approximately,of%20hearing%20community%20(DHH).

Overall Architecture of the Proposed System

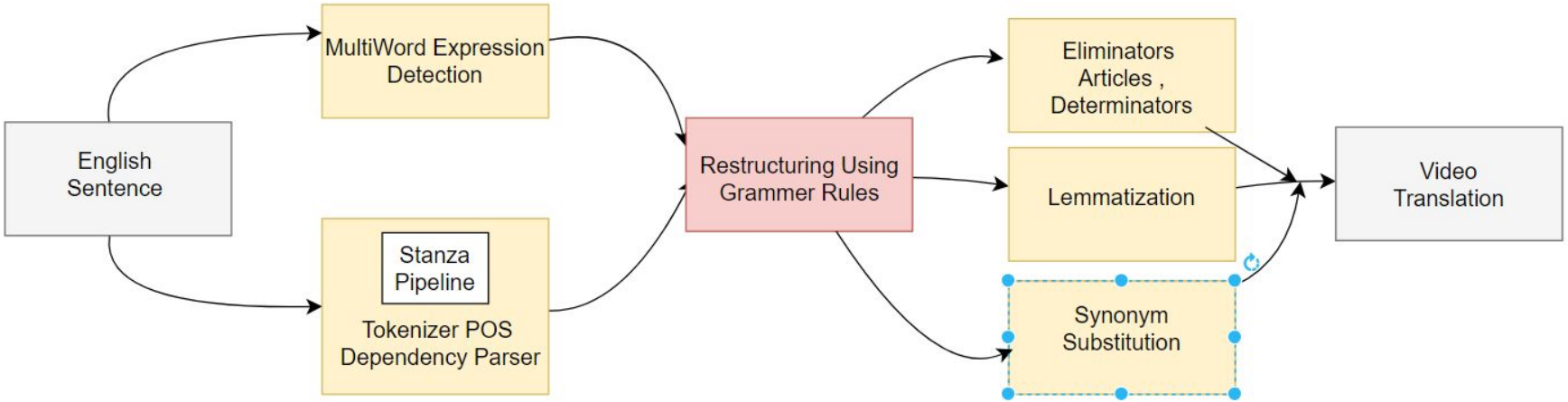
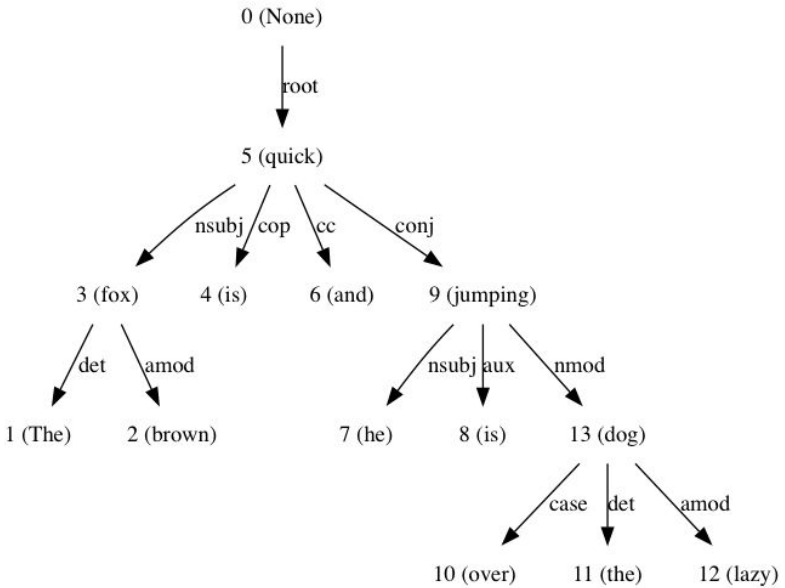


Figure - Architecture of English Text to ISL Video Translation

Stanza PipeLine

1. Tokenization
2. POS Tagging
3. Dependency Parsing



https://stackoverflow.com/questions/13468872/how-to-get-a-dependency-tree-with-stanford-nlp-parser

Grammar Rules

1. Source: $VERB(root) \rightarrow NOUN(obj)$
   Transformation: $VP \rightarrow [NOUN(obj)][VERB(root)]$
   Example: $He\ eats\ mangoes \rightarrow HE\ MANGO\ EAT$

2. Source: $VERB(any) \rightarrow AUX(aux)$
   Transformation: $VP \rightarrow [AUX(aux)][VERB(any)]$
   Example: $He\ was\ eating \rightarrow HE\ WAS\ EATING$
   All auxiliaries to a verb come before the verb

https://aclanthology.org/2022.icon-main.17.pdf

Grammar Rules

3. Source: $VERB(any) \rightarrow ADV(advmod)$
   Transformation: $VP \rightarrow [VERB(any)][ADV(advmod)]$
   Example: $He\ ran\ quickly \rightarrow HE\ RUN\ QUICKLY$
   Explanation: All adverbs to a verb come after the verb

4. Source: $VERB(root) \rightarrow VERB(any)$
   Transformation: $VP \rightarrow [VERB(root)][VERB(any)]$
   This case is written to handle multiple clauses in a sentence. In case of multiple clauses in a sentence, the clauses will be processed in order.

Example - She went to the store, and she bought some groceries

https://aclanthology.org/2022.icon-main.17.pdf

Grammar Rules

5. Source: $NOUN(any) \rightarrow ADJ(amod)$
   Transformation: $NP \rightarrow [NOUN(any)][ADJ(amod)]$

   Example: $He\ has\ a\ blue\ book \rightarrow HE\ BOOK\ BLUE\ HAVE$
   Explanation: Adjectives follow the noun they describe

6. Source: $NOUN(any) \rightarrow NUM(nummod)$
   Transformation: $NP \rightarrow [NOUN(any)][NUM(nummod)]$
   Example: $He\ has\ three\ sons \rightarrow HE\ SON\ THREE\ HAVE$
   Numbers are handled in two ways in ISL:

   (a) Reduplication: Repeat the noun to signal plurality
   (b) Numbers follow the noun denoting the quantity

https://aclanthology.org/2022.icon-main.17.pdf

Grammar Transfer Rules

7. Source: $NOUN(any) \rightarrow ANY(acl : relcl)$
   Transformation: $NP \rightarrow [NOUN(any)][ANY(acl : relcl)]$
   Example: *I saw the book which you bought* → *I BOOK YOU BUY SEE*
   Explanation: A relative clause is similar to an adjective and hence it comes after a noun.

https://aclanthology.org/2022.icon-main.17.pdf

Post Processing

Till this stage we have a rough ordering of the signs we need to show. In post processing, we handle special cases, filter out the unwanted words and reduce the words to their root forms.

1. Interrogative Sentences
2. Negative Sentences
3. Synonym Substitution
4. Stopword Removal and Lemmatization
5. Video Translation

Post Processing - Interrogative Sentences

It is difficult to handle cases of interrogative sentences since we dont know whether or not a wh-word used in a sentence is a question or not, eg : “I know what he is talking about.”

Take the wh-word and pick it and place it at the end of the sentence.

e.g: Did you do your homework is changed to YOU HOMEWORK DO WHAT

Post Processing - Negative Sentences

Negative sentences are sentences that have a not or no in them.

Pick the negative word and place it at the end of the sentence, e.g. He is not a doctor → HE DOCTOR NOT.

If there is negative question sentence such as “Who will not come with me”, sign the negative word before the question word and sentence becomes “I WITH GO NOT WHO. “

Post Processing - Synonym Substitution

As such, till now, there are only 10000 words in the ISL dictionary

Hence, there are many words in English that do not have an equivalent sign

If such a words occur in a sentence, substitute those words that are not in the video dictionary with words close to the original word's meaning and have a corresponding word in the dictionary.

Post Processing - Stopword Removal

After all the above processing is done, we need to clean up the sentences. ISL does not contain words such as articles and certain functional words that do not necessarily convey meaning.

For this, they created their own list of stop-words for English.

Then these stop words are removed.

Eg of stop-words are “the,” “and,” “is” etc.

Post Processing - Lemmatization

Lemmatization reduces words to their base or dictionary form. For example, the lemma of "running" is "run," and the lemma of "better" is "good."

Followed by the removal of the stop-words, Convert all the remaining tokens into their root forms.

Video Translation

After the sentences have been converted into text for sign language, the output tokens are then searched for in the video database.

If the tokens are not available in the database, then they use WordNets to check if there are synonyms for that given word that have the same POS and are present in the dictionary

If any of the words' synonyms are not found in the dictionary, then that sentence cannot be translated into ISL.

MWE Processing

In MWE processing, integration of the jMWE dataset by Kulkarni and Finlayson (2011) with the MWETokenizer from NLTK by Bird and Loper (2004) enhances the system's ability to accurately identify and handle multiword expressions (MWEs) in natural language text. Leveraging annotated data and specialized tokenization techniques improves the parsing and processing of MWEs, thereby contributing to a more robust and context-aware understanding of language.

Example → "Piece of cake" (meaning:very easy to do)

Results

Total Sentences Checked -741

Sentences with synonyms substituted -304

Sentences with MWEs detected -24

ACCURACY-95.833

Demonstration

Sentence = A woman sells newspapers

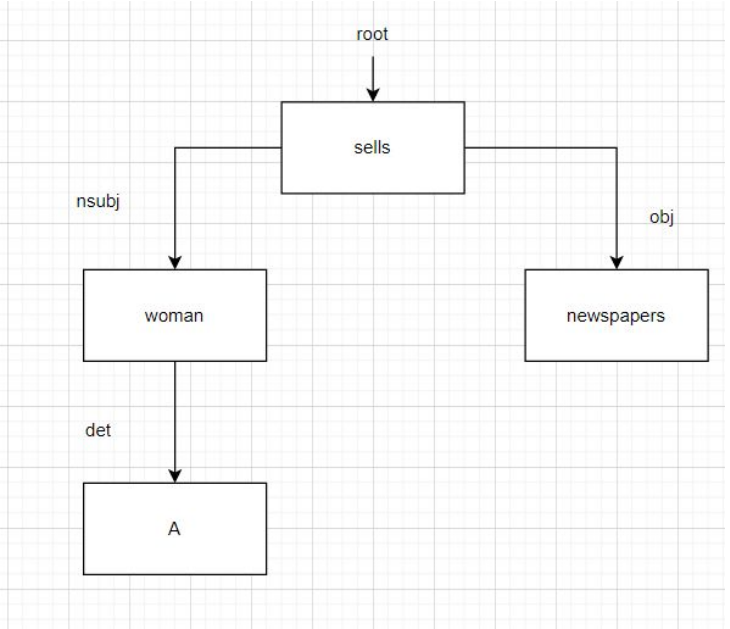
POS TAGGING

WORDS	POS TAG
A	DETERMINANT
Women	NOUN
sells	VERB
newspapers	NOUN

Demonstration

Figure - Dependency Parse Tree for given sentence.

Sentence : A woman sells newspapers



Demonstration

1. **GRAMMAR RULES APPLIED**
   a. RULE 1: As ISL follow subject order verb(SOV) form so verb “sells” will come after the object “newspaper”
   b. RULE 2: A (Determinant) will be removed
   c. ISL-Grammer : **women newspaper sells**
2. **SYNONYM SUBSTITUTION -** newspaper will be substituted by paper
3. **FINAL SENTENCE**: **woman paper sell**

Demonstration

Final Sentence - **woman paper sell**

Limitations of Proposed Solution

1. Not accurately identifying and handling interrogative sentences in sign language processing due to the ambiguity in the dependency roles of wh-words and the variability in their usage across different sentence contexts.
2. Rules are not accurate for the complex sentences which have more then one subject and object
3. Mapping videos to words can present challenges as different videos may play sequentially, leading to potential confusion. Instead, employing animations or pictures can offer a clearer representation, allowing for better comprehension and retention of information.

Conclusion

Study addresses the challenge of semantically complex sentences by employing innovative techniques such as Multi-Word Expression detection and Synonym substitution. Through testing on simple English sentences, demonstrate the efficacy of the approach in capturing hidden meanings within sentence structures. By filling the gaps left by previous methods, this approach paves the way for more robust and nuanced understanding of language semantics.

References

1. https://en.wikipedia.org/wiki/Deafness_in_India#:~:text=India%20is%20home%20to%20approximately,of%20hearing%20community%20(DHH).
2. https://aclanthology.org/2022.icon-main.17.pdf
3. https://stackoverflow.com/questions/13468872/how-to-get-a-dependency-tree-with-stanford-nlp-parser
4. https://www.geeksforgeeks.org/removing-stop-words-nltk-python/

THANK YOU!