English text to Indian Sign Language Machine Translation: A Rule Based Method

Gouri Sankar Mishra, Parma Nand, Pooja

Abstract: *The hearing challenged community all over world face difficulties to communicate with others. Machine translation has been one of the prominent technologies to facilitate a two way communication to the deaf and hard of hearing community all over the world. We have explored and formulated the fundamental rules of Indian Sign Language and implemented as a translation mechanism of English Text to Indian sign Language glosses. The structure of the source text is identified and transferred to the target language according to the formulated rules and sub rules. The intermediate phases of the transfer process is also mentioned in this research work.*

Index Terms: *Indian Sign Language (ISL), Machine Translation (MT), Rule Based Machine Translation (RBMT), Syntactic transfer.*

I. INTRODUCTION

Sign Languages are used by the deaf and hard of hearing community all over the world. The sign language used in India is Indian Sign Language (ISL) [3]. A study by M. Vashisth et. al. [25] specifies that the ISL used in different parts of India is almost same in its structure [19] with a differentiation in signs. It is a social need to facilitate the hearing impaired people of Indian sub-continent with a tool which can translate the English text to ISL. ISL has all the properties of a natural language and is also considered as a natural language like British Sign Language (BSL), American Sign Language (ASL)[9], Australian Sign Language (AUSLAN).

The idea of automatic machine translation in the area of Sign Language translation has been evolving very fast in the last two decades as the technology needs to be used for the hearing impaired people all over the world. Most of the researchers experimented with the rule based machine translation methodology [3][4][5] to translate a spoken or, written language to Sign Language as this method is relied upon the dictionary and the grammars of the source and target languages.

The generic translation technology used is Rule Based Machine Translation [15]. Today also this technique is very useful in certain scenarios and also the technology is amalgamated with other technologies like Phrase based statistical machine translation [13] & Word based statistical machine translation [14], to discover new hybridized techniques. As the name suggests, the translation process

requires the knowledge of the language pairs, i.e. the source and target languages.

II. RULE BASED MACHINE TRANSLATION

In this methodology the rules of the language pairs should be identified preferably through a linguistic experts. The process of RBMT [17]can be better understood by a triangle proposed by Bernard Vauquois and is popularly known as Vauquois triangle which is shown in the figure-1.

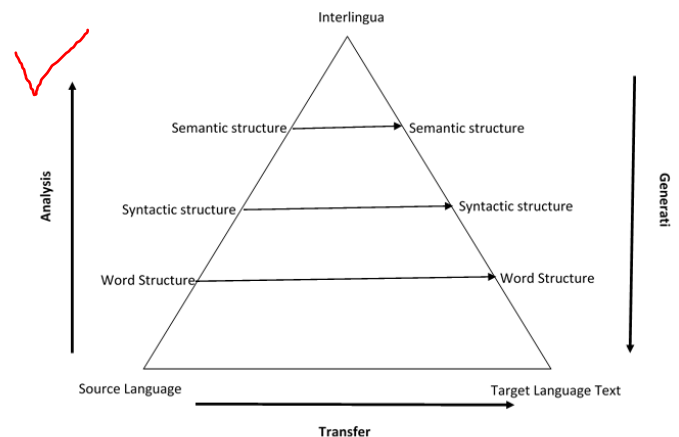

Figure-1 : Vauquois triangle expressing approaches to machine translation

The translation process[6][7[8] is operated in different levels of this triangle. The left hand side is the ascending side and the right hand side is the descending side. The left corner shows the source language text and the right corner the target language text. While ascending in the left hand side, the source language text may go through various analysis phases and while descending in the right hand side the respective generation phases occur. The parallel horizontal arrows shows the transfer of respective analysis phase of source language to generation phase of target language. The words of the source language is directly transferred to a target language words by using a dictionary is called direct transfer. This happens in the morphological analysis and generation phase. The next level of translation considers the syntax of source and target language. A syntactic parse tree is framed and is transferred to a corresponding parse tree of a target language. This process is also called syntactic transfer. The semantic transfer follows the syntactic transfer. Here the meaning of the source language is transferred to the target language. And at last the highest level of Vaquois triangle is interlingua [9]-[10]where the sentences are represented in a language independent structure.

From here the target languages can be generated. However, in each phase there exists a number of the ambiguity which became the real challenge of machine translation.

III. ENGLISH - SIGN LANGUAGE RULE BASED TRANSLATION SYSTEMS-STATE OF ART

The ViSiCAST project [9]-[10] is used for translating English text to British Sign Language by implementing CMU Link parser. The given English text is parsed and then Prolog declarative clause grammar rules are used to convert the output into a Discourse representation structure (DRS)[6].In generation phase, head driven phrase structure rules are used to produce symbolic signs. Though both ASL and BSL share many fundamental linguistic properties, they are not mutually intelligible.

The ZARDOZ system [12] is developed to make a framework to translate BSL [14], Irish Sign Language and Japanese Sign Language on the basis of ASL [15]. The translation system uses a set of hand coded schemata which is framed as a language independent structure [16] i.e. Interlingua. The methodology for analysis and generation is based on knowledge representation in Artificial Intelligence. In analysis phase, the source English sentence is decomposed by implementing sophisticated idiomatic concept before syntactic parsing so that the commonsense and other reasoning components in the system could later easily operate on the semantic information. In any case if an input text is not fit in the above schemata, the authors proposed a direct translation is applied on such texts.

The ASL Workbench [17] is designed to be used as a tool by the professional translators. This system is not a fully automatic system, rather it requires human intervention during ambiguity in lexicon selection. The ASL Workbench is a rule based translation system in which the analysis of the source English text is converted into a functional structure by using lexical-functional grammar (LFG)[18][19]. The transfer rules are hand-crafted for converting an English functional structure into an ASL structure. By implementing the LFG rules, the ASL output is produced. The system uses a transfer-specific lexicon to map English words/phrases to respective ASL signs/phrases.

This system is developed in the University of Pennsylvania. TEAM[10] [23] is an English to ASL translation system which works in two different phases. In the first phase the input English text is translated by using Synchronous Tree Adjoining Grammar rules to ASL gloss notations along with embedded parameters. The embedded parameters are the extra information about the non-manual signs. In the second phase, the glosses along with the embedded parameters are expressed by an animated software through animated virtual human as signing avtar.

In the translation of Indian Sign Language two prominent researches have been reported. The English to Indian Sign Language translation through a rule based system is reported by T. Dasgupta et. al. [6]. They proposed an analysis of preprocessed sentence by a dependency parser to identify the syntactic and functional information. The information is used

to form an intermediate language independent recursive case frame structure. The intermediate structures are assigned a numeric value and is the input to the generation phase. The transfer grammar rules are applied for the structural transfer of source to target structure.

P. Kar et. al. [5] reported Hindi to English translation system and named as INGIT. This translation system is designed to translate the conversation in the Indian Railway reservation counter where mostly the local people talk in Hindi with the officials in the reservation counter. In this system a semantically mediated formulaic framework for Hindi to ISL is adopted. To handle formulaic inputs a construction grammar is adopted for construction of lexicon with single constituent or, larger phrases with direct semantic mapping at each level. The experimentations are validated manually by the native ISL signers.

IV. PROPOSED METHODOLOGY

This methodology is used for translating a given text in English to Indians Sign Language glosses. The syntax of both the languages are different. In the proposed methodology, the input sentence is tokenized through a module called Tokenizer. The tokens are then passed through a module called as POS Tagger to find out the POS tags of the tokens.
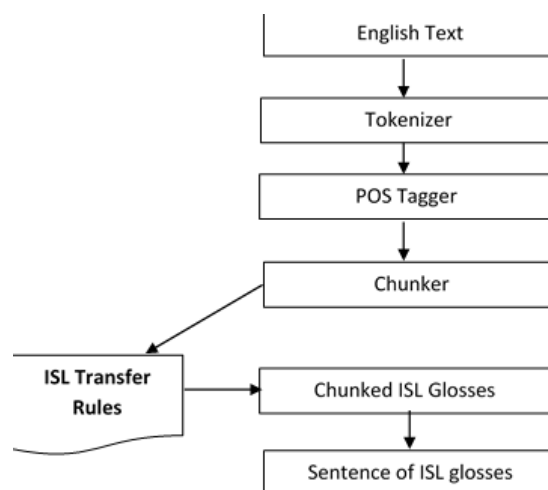

Figure-2: Architecture of Proposed methodology

The tagged words are then sent to the next module called as chunker. Here the chunks are found out and the chunks are applied with the transfer rules of ISL. The transfer rules[22][24][25] are defined according to the syntax of ISL. The transfer rules and the experimental results are explained in detail in the following section. The chunked ISL glosses are generated by applying the transfer rules. From the chunked ISL glosses the target sentence is generated containing ISL glosses.

V. TRANSFER RULES AND EXPERIMENTAL RESULTS

The followings are the syntactic rules along with the examples of ISL derived from U. Zeeshan book [28]. These rules are implemented in the tool NLTK. Along with the rules, the examples and the results from the experiments are also given for easy reference.

Rule 1: placing the participants first and the predicate last.

Sub rule 1.a. Participants in a sentence should be placed at the beginning of each sentence. The structure followed in ISL is SOV i.e. Subject-Object-Verb. In English, the structure of sentences is SVO, i.e. Subject-Verb-Object.

Example:
    English   : ' I am deaf '
    ISL   : ' मैं   बहरा '
    ISL gloss : ' I   DEAF '

The source sentence in English follows the SVO structure and the translated ISL follows SOV. As per the rule the participant is in the first place then comes the predicate by removing 'am' as it is not a valid token in ISL. Now the sentence with three tokens is tokenized as:

Tokenize :   ['I', 'am', 'deaf.']
POS tag :[('I', 'PRP'), ('am', 'VBP'), ('deaf.', 'JJ')]
After Chunking: (Tree('S',[Tree('SOURCE', [('I', 'PRP'), ('am', 'VBP'), ('deaf.','JJ')])])])

Transferred: ('Transferred Tree', Tree('S', [Tree('TARGET',

                  [('I', 'PRP'), ('DEAF.', 'JJ')])])])

The parsed tree is transferred by implementing the grammar of ISL to find the above parse tree:

Sub rule 1.b. If there is no object in a sentence then in both the languages the verb is followed by the subject.
Example:
English   : ' I   grew up '
ISL   : ' मैं बड़ा होना '
ISL gloss :   ' I   GROW UP '

The simple past is converted to simple present. To use tense in an ISL sentence, we have another rule to be discussed in this section. The experimental results are as follows:

Tokenize   : ['I', 'grew', 'up']
POS tag   : [('I', 'PRP'), ('grew', 'VBD'), ('up', 'RP')]
After Chunking :(Tree('S', [Tree('SOURCE', [('I', 'PRP')]), ('grew', 'VBD'), ('up', 'RP')])])
Transferred   : ('Transferred Tree', Tree('S', [Tree('SOURCE', [('I', 'PRP'), ('grow', 'VBP')]), ('up', 'RP')]))

Sub rule 1.c. It is not necessary to include each word of the sentence. The words may be dropped in the target sentence as per the requirements of translation irrespective of the availability of ISL tokens for the dropped words.
Example:
English   : 'There are only few deaf people'
ISL   : ' बहरा   कम'
ISL gloss :   ' DEAF LITTLE'

Here 'There', 'are', 'only' and 'people' words are omitted in the corresponding translated text. The word 'people' is a valid token in ISL. These words are not required to communicate the content of the source sentence in ISL. This shows the property of ISL to communicate the meaning of the sentence but not the words. The experimental results are as follows:

Tokenize   : ['There', 'are', 'only', 'few', 'deaf', 'people'])
POS tag   : [('There', 'EX'), ('are', 'VBP'), ('only', 'RB'), ('few', 'JJ'), ('deaf', 'NN'), ('people', 'NNS')]
After Chunking : Tree('S', [('There', 'EX'), Tree('SOURCE', [('are', 'VBP')]), ('only', 'RB'), Tree('SOURCE', [('few', 'JJ')]), ('deaf', 'NN'), ('people', 'NNS')])
Transferred   : ('Transferred Tree', Tree('S', [('DEAF', 'NNP'), ('LITTLE', 'NNP')]))

Sub rule 1.d. The first alphabet of the noun is denoted for the noun (Place, Person name, etc.) which don't have a sign in ISL,
Example:
English : ' The   club   (will)   become bigger. '
ISL   : '   क्लब   तरक्क़ी '
ISL gloss: '   C   develop '

In the above example, the word 'CLUB' doesn't have a sign in ISL. So the first alphabet 'C' is spelled out to represent 'CLUB'.

The experimental results are as follows:

Tokenize   : ['The', 'club', 'become', 'bigger.']
POS tag   : [('The', 'DT'), ('club', 'NN'), ('become', 'NN'), ('bigger.', 'NN')]
After Chunking: (Tree('S',[('The', 'DT'),('club', 'NN'),('become', 'NN'), ('bigger.', 'NN')])
Transferred   : ('Transferred Tree', Tree('S', [('Club', 'NNP'), ('develop.', 'NN')]))

Rule 2: 'Movement signs' in combination with an agent and a location mostly prefer the order agent-location-predicate:

Sub rule 2.a.: The plural form is generally represented in singular.
Example:
English   : ' I   went   to   clubs '
ISL   : ' मैं   क्लब   जाना '
ISL gloss   : ' I   C   GO '

By following the Rule-1, the subject and object are placed first, then the verb. The simple past is also translated to simple present. In this example, 'clubs' become 'CLUB' but 'MANY' is not used here as the entire sentence it doesn't give stress on 'many clubs'. So, 'MANY' is used when it's necessary to use and the stress is on the plurality. If its required to specify the plural form then the ISL token used for plural form is ' बहुत ' or, ' बहुत सारे '. We have used the ISL gloss 'MANY' for ' बहुत ' sign. We found the following results:

Tokenize   : ['I', 'went', 'to', 'clubs.']
POS tag   : [('I', 'PRP'), ('went', 'VBD'), ('to', 'TO'), ('clubs', 'VB')]
After Chunking : Tree('S', [Tree('SOURCE', [('I', 'PRP')]), ('went', 'VBD'), ('to', 'TO'), ('clubs.', 'VB')])
Transferred : Tree('S', [Tree ('SOURCE',[('I','PRP'), ('club' , 'VBP')]), ('go.', 'NN')])

Sub rule 2.b.: For several participants in a sentence, not all of them are required to be expressed by individual sign.
Example:
English   : 'Then in 2015 Bill Gates and Merinda Gates came to New Delhi. '
ISL   : ' बाद दो-शुन्य-एक-पांच बिल गेट्स दो दिल्ली उड़ना '
ISL gloss : ' THEN TWO-ZERO-ONE-FIVE BILL GATES TWO Delhi FLY '

Experimental results:

Tokenize   : ['Then', 'in', '2015', 'Bill', 'Gates', 'and', 'Merinda', 'Gates', 'came', 'to', 'New', 'Delhi.']
POS tag   : [('Then', 'RB'), ('in', 'IN'), ('2015', 'CD'), ('Bill', 'NNP'), ('Gates', 'NNP'), ('and', 'CC'), ('Merinda', 'NNP'), ('Gates', 'NNP'), ('came', 'VBD'), ('to', 'TO'), ('New', 'NNP'), ('Delhi.', 'NNP')]
After Chunking : Tree('S', [('Then', 'RB'), ('in', 'IN'), ('2015', 'CD'), ('Bill', NNP'), ('Gates', 'NNP'), Tree('SOURCE', [('and', 'CC')]), 'Merinda', 'NNP'), ('Gates', 'NNP'), ('came', 'VBD'), ('to', 'TO'), ('New', 'NNP'), ('Delhi.', 'NNP')])
Transferred   : Tree('S', [('THEN', 'NNP'), ('TWO-ZERO-ONE-FIVE', 'NNP'), ('BILL', 'NNP'), ('GATES', 'NNP'), ('TWO', 'NNP'), ('Delhi', 'NNP'), ('FLY.', 'NNP')])

In the above example, out of two participants only one i.e. 'BILL GATES' is mentioned in the target language and another participant Merinda Gates is clubbed with Bill Gates by including ISL gloss 'TWO' which represents along with Bill Gates another participant is present. If there are more than two participants, then all of them can be counted and along with the first participant the number can be clubbed together.

Sub rule 2. C.: For number translation, each digit of a number is signed sequentially from left to right.

As mentioned in the above example, 2015 is expressed as 'TWO-ZERO-ONE-FIVE'. There are 10 signs in ISL from 0 to 9. We have included the digits in the ISL glosses.

Rule 3: Experiencer — predicate — stimulus OR stimulus — experiencer — predicate OR experiencer — stimulus — predicate:

There is no definite pattern of word order for experincer, predicate and stimulus. This can be accepted as a "flexibility Condition'. In the following three sub rules each example demonstrates possible construction with "देखना" / "to see".

Sub rule 3.a.: The more inflected the verb is, the freer the word order will be. The word order is: experiencer — predicate — stimulus:

Example:
English   : ' He looks right and left into the side mirror '
ISL   : ' वह देखना बगल-का-शीशा दाईं देखना बगल-का-शीशा बाईं '
ISL gloss : ' He see side_mirror-right see side_mirror left. '

Tokenize   : ['He', 'looks', 'right', 'and', 'left', 'into', 'the', 'side', 'mirror.']
POS tag   : [('He', 'PRP'), ('looks', 'VBZ'), ('right', 'JJ'), ('and', 'CC'), ('left', 'VBD'), ('into', 'IN'), ('the', 'DT'), ('side', 'NN'), ('mirror.', 'NN')]
After Chunking : Tree('S', [Tree('SOURCE', [('He', 'PRP')]), ('looks', 'VBZ'), Tree('SOURCE', [('right', 'JJ'), ('and', 'CC')]), ('left', 'VBD'), ('into', 'IN'), ('the', 'DT'), ('side', 'NN'), ('mirror.', 'NN')])
Transferred   : Tree('S', [Tree('SOURCE', [('He', 'PRP')]), ('see', 'VB'), Tree('SOURCE', [('side_mirror-right', 'JJ')]), ('see', 'NNS'), Tree('SOURCE', [('side_mirror', 'VBP')]), ('left.', 'VB')])

Sub rule 3.b.: The predicate requires an animate experiencer, a second inanimate participant will automatically be interpreted as the stimulus. The word order is: experiencer — predicate — stimulus:

Example:
English   : ' I only watch TV series '
ISL   : ' टीवी नाटक   मैं देखना बस '
ISL gloss : 'TV Drama I   see   only'

Tokenize   : ['I', 'only', 'watch', 'TV', 'series']
POS tag   : [('I', 'PRP'), ('only', 'RB'), ('watch', 'VBP'), ('TV', 'NN'), ('series', 'NN')]
After Chunking : Tree('S', [Tree('SOURCE', [('I', 'PRP')]), ('only', 'RB'), Tree('SOURCE', [('watch', 'VBP')]), ('TV', 'NN'), ('series', 'NN')])

Transferred   : Tree('S', [('TV', 'NN'), ('Drama', 'NNP'), Tree('SOURCE', [('I', 'PRP'), ('see', 'VBP'), ('only.', 'JJ')])])

The only systematic exception from the sentence final rule is that the sign ' बस ', a discourse particle which is used to end an utterance, occurs after some of the function signs. The meaning of ' बस ' may be paraphrased as 'That's all./What else can I say?', which implies an interrogative component of meaning. It also occurs as a filler of hesitation pauses and at the end of incomplete utterances.

Sub rule 3.c.: Elicit sentences whose predicates are not spatially modifiable and in which all participants are represented by individual signs. The word order is: experiencer — stimulus — predicate:

Example:
English   : 'In the morning I have a look at the news paper'
ISL   : ' सुबह   मैं   अखबार   देखना बस '
ISL gloss : 'morning I   newspaper see   end'

Tokenize : ['In', 'the', 'morning', 'I', 'have', 'a', 'look', 'at', 'the', 'news', 'paper']
POS tag : [('In', 'IN'), ('the', 'DT'), ('morning', 'NN'), ('I', 'PRP'), ('have', 'VBP'), ('a', 'DT'), ('look', 'NN'), ('at', 'IN'), ('the', 'DT'), ('news', 'NN'), ('paper', 'NN')]
After Chunking : Tree('S', [('In', 'IN'), ('the', 'DT'), ('morning', 'NN'), Tree('SOURCE', [('I', 'PRP'), ('have', 'VBP')]), ('a', 'DT'), ('look', 'NN'), ('at', 'IN'), ('the', 'DT'), ('news', 'NN'), ('paper', 'NN')])
Transferred : Tree('S', [('morning', 'NN'), Tree('SOURCE', [('I', 'PRP')]), ('newspaper', 'NN'), Tree('SOURCE', [('see', 'VBP')]), ('end', 'NN')])

Rule 4: A rather frequent principle of ordering involves the most important sign which is semantically most prominent to be placed at the beginning of the sentence. Then the whole sentence follows which again contains the same sign.

Sub rule 4.a.: A title or headline is presented first and then the whole subject matter follows. This brings the topic-comment structures of spoken languages, with the particularity that the sign which is extracted from the sentence appears again in the complete sentence.

Example:
English : ' I got a letter from the club. '
ISL : ' चिट्ठी | क्लब चिट्ठी है '
ISL gloss : ' Letter C   letter is '

Tokenize : ['I', 'got', 'a', 'letter', 'from', 'the', 'club']
POS tag : [('I', 'PRP'), ('got', 'VBD'), ('a', 'DT'), ('letter', 'NN'), ('from', 'IN'), ('the', 'DT'), ('club', 'NN')]
After Chunking : Tree('S', [Tree('SOURCE', [('I', 'PRP')]), ('got', 'VBD'), ('a', 'DT'), ('letter', 'NN'), ('from', 'IN'), ('the', 'DT'), ('club', 'NN')])
Transferred : Tree('S', [('Letter', 'NNP'), ('club', 'NN'), ('letter', 'NN'), ('is', 'VBZ')])

Sub rule 4.b.: If the word in a sentence is important, i.e. the entire sentence explains about the word, then the word is kept first, and then in the sentence it is also repeated at its normal position.

Example:
English : 'We deaf people all play (cricket), (so) we all have bat (cricket) with us. '
ISL : ' bat | bahra pura khel bat hai '
ISL gloss : ' Bat | deaf all play bat is '

Tokenize : ['We', 'deaf', 'people', 'all', 'play,', 'we', 'all', 'have', 'bat', 'with', 'us']
POS tag : [('We', 'PRP'), ('deaf', 'VBP'), ('people', 'NNS'), ('all', 'DT'), ('play,', 'VBP'), ('we', 'PRP'), ('all', 'DT'), ('have', 'VBP'), ('bat', 'VBN'), ('with', 'IN'), ('us', 'PRP')]

After Chunking : Tree('S', [Tree('SOURCE', [('We', 'PRP'), ('deaf', 'VBP')]), ('people', 'NNS'), ('all', 'DT'),Tree('SOURCE', [('play,', 'VBP')]), Tree('SOURCE', [('we', 'PRP')]), ('all', 'DT'), Tree('SOURCE', [('have', 'VBP')]), ('bat', 'VBN'), ('with', 'IN'), Tree('SOURCE', [('us', 'PRP')])])

Transferred : Tree('S', [('Bat', 'NNP'), ('deaf', 'NN'), ('all', 'DT'), Tree('SOURCE', [('play', 'VBP')]), ('bat', 'NN'), ('is', 'VBZ')])

Rule 5: Using temporal expressions in ISL. In ISL there is no temporal inflection, but the tense is indicated by the use of time signs such as 'पहले' (before), 'बाद' (after, then), 'आगे' (further, later, in future, etc.).

Sub rule 5.a.: The time signs are articulated along a timeline. Usually a temporal frame is fixed at the beginning of an utterance and the signs are interpreted within this frame.

Example:
English : 'In 1991 I passed the tenth class '
ISL : ' पहले स्कूल दस सफल एक नौ नौ एक'
ISL gloss : ' Past school ten success one nine nine one '

Tokenize : ['In', '1991', 'I', 'passed', 'the', 'tenth', 'class']

POS tag : [('In', 'IN'), ('1991', 'CD'), ('I', 'PRP'), ('passed', 'VBD'), ('the', 'DT'), ('tenth', 'NN'), ('class', 'NN')]

After Chunking : Tree('S', [('In', 'IN'), ('1991', 'CD'), Tree('SOURCE', [('I', 'PRP')]), ('passed', 'VBD'), ('the', 'DT'), ('tenth', 'NN'), ('class', 'NN')])

Transferred : Tree('S', [Tree('SOURCE', [('Past', 'JJ')]), ('school', 'NN'), ('ten', 'NN'), ('success', 'NN'), ('one', 'CD'), ('nine', 'CD'), ('nine', 'CD'), ('one', 'CD')])

Sub rule 5.b.: 'पहले' (before), 'बाद' (after, then) indicate an event occurring before and after the present time or another temporal point of reference. If the temporal distance between the point of reference and the event is indicated, these signs occur in combination with expressions that stand for a certain period of time.

Example:
English : 'Then I was a teacher at a school for two years '
ISL : ' बाद स्कूल दो साल सिखाना'
ISL gloss : ' then school two year teach '

Tokenize : ['Then', 'I', 'was', 'a', 'teacher', 'at', 'a', 'school', 'for', 'two', 'years']
POS tag : [('Then', 'RB'), ('I', 'PRP'), ('was', 'VBD'), ('a', 'DT'), ('teacher', 'NN'), ('at', 'IN'), ('a', 'DT'), ('school', 'NN'), ('for', 'IN'), ('two', 'CD'), ('years', 'NNS')]

After Chunking : Tree('S', [('Then', 'RB'), Tree('SOURCE', [('I', 'PRP')]), ('was', 'VBD'), ('a', 'DT'), ('teacher', 'NN'), ('at', 'IN'), ('a', 'DT'), ('school', 'NN'), ('for', 'IN'), ('two', 'CD'), ('years', 'NNS')])

Transferred : Tree('S', [('then', 'RB'), ('school', 'NN'), ('two', 'CD'), ('year', 'NN'), ('teach', 'NN')])

Sub rule 5.c.: If a point in time such as a certain date is mentioned, time signs expressing the relative temporal distance from the point of reference often precede it and the whole complex stands at the beginning of the sentence.

Example:
English : 'Then three years latter in 1996 I became a teacher at Pie Academy '
ISL : 'तीन साल बाद एक नौ नौ छह शिक्षक स्कूल पी स्कूल वाह '
ISL gloss: ' three year then one nine nine six teacher school P school delight '

Tokenize : ['Then', 'three', 'years', 'later', 'in', '1996', 'I', 'became', 'a', 'teacher', 'at', 'Pie', 'Academy']

POS tag   : [('Then', 'RB'), ('three', 'CD'), ('years', 'NNS'), ('later', 'RBR'), ('in', 'IN'), ('1996', 'CD'), ('I', 'PRP'), ('became', 'VBD'), ('a', 'DT'), ('teacher', 'NN'), ('at', 'IN'), ('Pie', 'NNP'), ('Academy', 'NNP')]

After Chunking : Tree('S', [('Then', 'RB'), ('three', 'CD'), ('years', 'NNS'), ('later', 'RBR'), ('in', 'IN'), ('1996', 'CD'), Tree('SOURCE', [('I', 'PRP')]), ('became', 'VBD'), ('a', 'DT'), ('teacher', 'NN'), ('at', 'IN'), ('Pie', 'NNP'), ('Academy', 'NNP')])

Transferred : Tree('S', [('three', 'CD'), ('year', 'NN'), ('then', 'RB'), ('one', 'CD'), ('nine', 'CD'), ('nine', 'CD'), ('six', 'CD'), Tree('SOURCE', [('teacher', 'JJ')]), ('school', 'NN'), ('P', 'NNP'), ('school', 'NN'), ('delight', 'NN')])

Rule 6: Using function signs i.e. ' जाओ ' (go), ' करो ' (do), ' मत करो '(don't do), ' नहीं ' (no), ' ना ना ' (no-no), ' क्या ' (what), ' हो गया ' (done), ' है ' (is), . These sign glosses share functional as well as formal characteristics.

Sub rule 6.a.: The function signs appear at the end of the sentence and so follow the same syntactic rule.

Example:
English   : 'You have to speak.'
ISL   : 'बोलना   करो'
ISL gloss : 'speak do'

Tokenize : ['You', 'have', 'to', 'speak']
POS tag   : [('You', 'PRP'), ('have', 'VBP'), ('to', 'TO'), ('speak', 'VB')]

After Chunking : Tree('S', [Tree('SOURCE', [('You', 'PRP'), ('have', 'VBP')]), ('to', 'TO'), ('speak', 'VB')])

Transferred   : Tree('S', [('speak', 'NN'), ('do', 'VB')])

Sub rule 6.b.: The imperative ' जाओ ' (go) is used in impolite orders, in particular orders to servants and other inferiors, and implies that the addressee will have to move physically ('go and do').

Example:
English   : 'Go and wash the clothes!'
ISL   : 'कपड़ा कपड़ा   धोना   जाओ'
ISL gloss   : 'cloth   cloth wash go'

Tokenize   : ['Go', 'and', 'wash', 'the', 'clothes']
POS tag   : [('Go', 'NNP'), ('and', 'CC'), ('wash', 'VBD'), ('the', 'DT'), ('clothes', 'NNS')]

After Chunking : Tree('S', [('Go', 'NNP'), Tree('SOURCE', [('and', 'CC')]), ('wash', 'VBD'), ('the', 'DT'), ('clothes', 'NNS')])

Transferred   : Tree('S', [('cloth', 'DT'), ('cloth', 'NN'), ('wash', 'NN'), Tree('SOURCE', [('go', 'VBP')])])

Sub rule 6.c.: The 'करो' (do) is neutral as to politeness, does not imply physical movement and might be translated as 'have to', 'should', 'is necessary' etc.

Example:
English   : 'Don't beat me!'
ISL   : ' मारना नहीं करो'
ISL gloss : 'beat not do'

Tokenize   : ['Don\x92t', 'beat', 'me']
POS tag   : [('Don\x92t', 'NNP'), ('beat', 'VBD'), ('me', 'PRP')]

After Chunking : Tree('S', [('Don\x92t', 'NNP'), ('beat', 'VBD'), Tree('SOURCE', [('me', 'PRP')])])

Transferred   : Tree('S', [('beat', 'NN'), ('not', 'RB'), ('do', 'VB')])

Sub rule 6.d.: ' नहीं ' (no) is used to negate sentences. To express the negation a headshake may be combined with the negative sign.

Example:
English   : 'I haven't had tea yet.'
ISL   : 'मैं चाय नहीं'
ISL gloss : 'I tea no'

Tokenize   : ['I', 'haven\x92t', 'had', 'tea', 'yet']
POS tag   : [('I', 'PRP'), ('haven\x92t', 'VBP'), ('had', 'VBD'), ('tea', 'NN'), ('yet', 'RB')]

After Chunking : Tree('S', [Tree('SOURCE', [('I', 'PRP'), ('haven\x92t', 'VBP')]), ('had', 'VBD'), ('tea', 'NN'), ('yet', 'RB')])

Transferred   : Tree('S', [Tree('SOURCE', [('I', 'PRP'), ('tea', 'VBP')]), ('no', 'DT')])

Sub rule 6.e.: ' नहीं ' (no) is the neutral negation, ' ना-ना ' (no-no) is used contrastively to negate a question or statement that has been either explicitly mentioned before or is implicitly understood as relevant context.

Example: Answering the question 'Would you like a cup of tea?'
English   : 'I don't want any tea.'
ISL   : 'मैं चाय ना-ना'
ISL gloss : ' I tea no-no'

Tokenize : ['I', 'don\x92t', 'want', 'any', 'tea']

POS tag : [('I', 'PRP'), ('don\x92t', 'VBP'), ('want', 'VBP'), ('any', 'DT'), ('tea', 'NN')]

After Chunking : Tree('S', [Tree('SOURCE', [('I', 'PRP'), ('don\x92t', 'VBP'), ('want', 'VBP')]), ('any', 'DT'), ('tea', 'NN')])

Transferred : Tree('S', [Tree('SOURCE', [('I', 'PRP'), ('tea', 'VBP')]), ('no', 'DT'), ('no', 'DT')])

In rule based machine translation, it is preferred to evaluate the experimental results manually by the experts or professionals. Automatic evaluation of the translated output is not error free. We have adopted manual evaluation of the above experiments.

VI. ANALYSIS OF EXPERIMENTAL RESULTS

In machine translation the evaluation of translated texts happens in two ways i.e. manual and automatic. The manual evaluation by the linguistic experts are considered more authentic, however the automatic evaluations are implemented now a days. The following is the comparison of different rule based sign language machine translation systems.
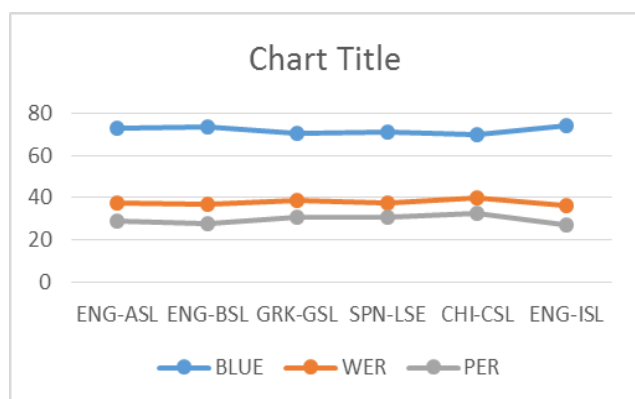

Figure-3: Comparision of Sign Language RBMT systems

We have considered three parameters of evaluation for the different language pairs, i.e. BLUE score, WER, PER. We have considered the translations of ASL, BSL, GSL (Greek Sign Language), LSE (Lengua de Signos i.e. Spanish Sign Language), CSL (Chineese Sign Language) and our proposed system of ISL. From the statistical analysis it is found that the BLUE score of English to Indian Sign Language (ENG-ISL) is 74.32% which is more than the ENG-ASL system and WER 36.56% and PER 27.32% which is less than the ENG-BSL systems.

VII. CONCLUSION

In this research we have experimented with the fundamental rules followed in the Indian Sign Language. These rules are formulated and coded and experimented in NLTK tool which is an open source widely accepted tool used in NLP. New rules can be formulated to make the translation system more effective. In sign language translation the manual signs can be generated from the corresponding glosses, but the non-manual signs require extra information along with the glosses. The research work can be extended to embed more features of non-manual signs with the glosses.

REFERENCES

1. Vivek Kumar Verma,Sumit Srivastava," A Perspective Analysis of Phonological Structure in Indian Sign Language", *Proceedings of First International Conference on Smart System, Innovations and Computing,* 2018, pp.175-180
2. Pawan Kumar, Savita Khatri, "Generating Indian Sign Language Text Using English/Hindi Text" , *International Journal of Recent Research Aspects, Special Issue: Conscientious and Unimpeachable Technologies* 2016, pp. 30-33
3. Taner Arsan and Oguz Ülgen, "Sign language converter", *International Journal of Computer Science & Engineering Survey (IJCSES),* August 2015, Vol.6, No.4,
4. M.Suresh Anand, A.Kumaresan, "An Integrated Two Way ISL (Indian Sign Language) Translation System – A New Approach", *International Journal of Advanced Research in Computer Science,* Jan-Feb 2013, Vol. 4, No. 2, pp 7-12
5. Kar, P., Reddy, M., Mukherjee, A., and Raina A. M. , "INGIT: Limited Domain Formulaic Translation from Hindi Strings to Indian Sign Language".*International Conference on Natural Language Processing (ICON), Hyderabad.* 2007.
6. Gouri Sankar Mishra, Ashok Kumar Sahoo, Kiran Ravaulakollu, "Machine Translation of English Text to Indian Sign Language: A phrase based Approach", *Proceedings Smart Technologies in Computer & Communication,* 2017, Amity University, Rajasthan,
7. Gouri Sankar Mishra, Ashok Kumar Sahoo, Kiran Kumar Ravulakollu, "Word based statistical machine translation from English text to Indian sign language", *ARPN Journal of Engineering and Applied Sciences*, 2017
8. T. Dasgupta, S. Dandpat, A. Basu,"Prototype Machine Translation System From Text-To-Indian Sign Language", *Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages, Asian Federation of Natural Language Processing,*, *Hyderabad, India,* January 2008, p 19–26.
9. I. Marshall & É. Sáfár. " Extraction of semantic representations from syntactic SMU link grammar linkages", *Proceedings of Recent Advances in Natural Language Processing (RANLP)*, 2001, pp. 154-159
10. É. Sáfár & I. Marshall, "The architecture of an English-text-to-Sign-Languages translation system", *Proceedings of Recent Advances in Natural Language Processing (RANLP)*, 2001, pp. 223-228
11. É. Sáfár & I. Marshall, "Sign language translation via DRT and HPSG", *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics*, *Springer Verlag.* 2002. pp. 58-68, [CICLing, Mexico, Lecture Notes in Computer Science 2276]
12. J. A. Bangham, S. J. Cox, R. Elliot, J. R. W. Glauert, I. Marshall, S. Rankov, & M. Wells, "Virtual signing: Capture, animation, storage and transmission – An overview of the ViSiCAST project", *IEEE Seminar on Speech and language processing for disabled and elderly people.* 2000.
13. T. Veale, A. Conway, & B. Collins, "The challenges of cross-modal translation: English to sign language translation in the ZARDOZ system", *Machine Translation*, 1998, vol. 13, pp. 81-106. .
14. d'A.L. Speers, "Representation of American Sign Language for Machine Translation", *PhD Dissertation, Department of Linguistics, Georgetown University,* 2001.
15. S. Liddell & R. Johnson, "American Sign Language: The Phonological Base", *Sign Language Studies*", 1989, vol. 64, pp. 195-277.
16. L. Zhao, K. Kipper, W. Schuler, C. Vogler, N. Badler, & M. Palmer. "A Machine Translation System from English to American Sign Language", *Association for Machine Translation in the Americas.* 2000.
17. Matthew P. Huenerfauth, "A Survey and Critique of American Sign Language Natural Language Generation and Machine Translation Systems", *Technical Report, Computer and Information Sciences University of Pennsylvania,* September 2003, MS-CIS-03-32

18. U. Zeshan, "Sign Language in Indo-Pakistan, A description of a signed language", Book, *John Benjamins publishing company Philadelphia/Amsterdam.* 2000

19. Alison Wray, Stephen Cox, Mike Lincoln and Judy Tryggvason, "A formulaic approach to translation at the post office: reading the signs", *Language and Communication*, 2004, vol. 24, pp. 59-75

20. A. L. Sexton, "Grammaticalization in American Sign Language", *Language Sciences*, 1999, vol. 21, pp.105-141

21. Dilip Deshmukh, "Sign Language and Bilingualism in Deaf Education", *India: Deaf Foundation,Ichalkaranji,* 1996.

22. Ian Marshall, Éva Sáfár, "A Prototype Text to British Sign Language (BSL) Translation System", *The Companion Volume to the Proceedings of 41st Annual Meeting of the Association for Computational Linguistics*, 2003, pp. 113-116

23. Liwei Zhao, Karin Kipper, William Schuler, Christian Vogler, Norm Badler, Martha Palmer "A MachineTranslation System from English to American Sign Language", *Lecture Notes in AI series of Springer-Verlag*, 2000 pp 54-67

24. Madan Vasishta, James Woodward and Susan DeSantis, "An Introduction to Indian Sign Language", *All India Federation of the Deaf (Third Edition)*, 1998.

25. Siegmund Prillwitz, Regina Leven, Heiko Zienert, Thomas Hamke, and Jan Henning, "HamNoSys Version 2.0: Hamburg Notation System for Sign Languages: An Introductory Guide", *International Studies on Sign Language and Communication of the Deaf. Signum Press, Hamburg, Germany,* 1989. volume 5

26. Tony Veale and Alan Conway, "Cross-Modal Comprehension in Zardoz, An English to Sign Language Translation system", *The Fourth International Workshop on Natural Language Generation, Maine, USA,* 1994.

27. Stokoe, W. C., "Sign language structure: an outline of the visual communication systems of the American deaf." *Silver Spring, MD: Linstok Press.* 1960.

28. Zeshan, U., "Indo-Pakistani Sign Language Grammar: A Typological Outline", *Sign Language Studies*, 2003, vol.3, no. 2, pp. 157-212

29. M.Jerin Jose, V.Priyadharshni,M.Suresh Anand, A.Kumaresan,Dr.N.MohanKumar, "Indian Sign Language (ISL) Translation System For Sign Language Learning", *International Journal of Innovative Research & Development,* 2013, vol. 2 issue 5, pp. 358-367

AUTHORS PROFILE

Gouri Sankar Mishra Researcher in the field of Machine translation, Natural Language processing pursuing PhD from Sharda University, Greater NOIDA. Contributing in academics since last 18 years. Author of 10 international publications. Delivered invited talks in conferences and lectures in faculty development programs organized in different Universities.

Prof. Parmanand A visionary, eminent academician, researcher in the field of Algorithms, AI & Machine learning, NLP, wireless networks etc. with PhD in Computer Science & Engineering from IIT Roorkee and Bachelor & Masters degree from in Technology from IIT Delhi, India. Having vast experience of 24 years, he is the author of more than 60 research papers in international publications. Being a constant source of inspiration within the students and researchers, he has delivered talks on International conferences, workshops and symposiums..

Dr. Pooja Analytical and research-oriented data scientist with PhD in Computer Science and Engineering from National Institute of Technology, Jalandhar. Collaborate with industry and research institutions towards learning and co-creation of contents in the fields of ML, NLP, AI, Image Processing. Recently worked with NITTTR, Chandigarh towards co-creation and delivery of course for "Swayam portal" on "Smart grid analytics" implementing "Machine Learning module". Experienced professional with hands on skills on Deep Learning, Machine Learning, Image Processing, Pattern Recognition, Natural Language Processing, Genetic Algorithms and Data Mining.