International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 1 (2018) pp. 202-209

Survey: Machine Translation for Indian Language

Shachi Mall
Guest Faculty, Department of Computer Science and Engineering
Madan Mohan Malaviya University of Technology, Gorakhpur, India.
Orcid Id: 0000-0002-4443-4885

Umesh Chandra Jaiswal
Professor, Department of Computer Science and Engineering
Madan Mohan Malaviya University of Technology, Gorakhpur, India.

Abstract

This paper presents Literature survey on Machine Translation and disuses the important tools and various approaches used for translating from one language into another language. There are various Machine Translations system has been developed for Indian languages such as Google Translator, BabelFish Translator etc. but they are failing to provide a good quality of translation. There are various challenges faced in Machine Translation such as Parts of speech tagging, morphological analyser, parsing, Word sense disambiguation and Translation. These are important tool of Natural Language processing.

Keywords: Machine Translation, Word sense disambiguation, Natural Language processing..

INTRODUCTION

The aim of machine translation [1] is a process which translates from one language called source language to another language called target language. Users can use this service for translating one language to another. Machine translation is from the broad area of Artificial Intelligence Natural language processing is based on different corpora(vocabulary), this corpora are used for the processing of NLP to generate and develop a standard model which can be used for many purposes such as speech recognition technique, etc. To develop a machine Translation process, there are four major goals:

1) Morphological analyser [2] (MA) analyses word level information and generates all possible roots. It takes <Root, Suffix> pair as input and outputs a set of features along with the set of roots.
2) Parts of speech tagging [3] is the process of assigning a part-of-speech like noun, verb, pronoun, preposition, adverb, adjective or other lexical class marker to each word in a sentence.
3) Chunk [4] is the task of identifying and labelling different types of phrases such as Noun phrase (NP), Verb phrase (VG), Adjectival phrase (JJP) etc. in a sentence.
4) Parsing [5] parse sentence by combination of parts of speech tagging and chunking.
5) Word sense disambiguation [6] (WSD) is the problem of automatically deciding the correct meaning of an ambiguous word based on the surrounding context in which it appears
6) Translation performance based on WordNet [7] for translating from Hindi to English words

MOTIVATION

The main motivation of Machine Translation is to build computational models of natural language for its analysis and generation. In particular this work is interdisciplinary field called computational linguistics driven from researches in Artificial Intelligence .There are three primary motivations for this type of research:

1) First, the technological motivation is to build intelligent computer systems.
2) Second, the linguistic and cognitive science motivation is to gain a better understanding of how humans communicate by using natural language. This system will help the people to understand English language.

Classification of Machine Translation

Machine Translation approaches as depicted in Figure1 can be classified into following categories, namely:

1) Direct Translation [8] approach directly translates the token with the bilingual dictionary.
2) Rule-Based [9] approach uses hand written rules built for Machine Translation. It is classified in two types First Transfer based approach [10] follows three basic steps such as analyses the sentence, Transfer appropriate grammar to each word and generate the target language. Second Interlingua based [11] follows the three basic steps of Transfer based approach but it not use intermediate structure.
3) Corpus-Based approach are of two types Statistical based [12] Example based [13] uses two approach Steaming process Stemming is process to extract original word (root) and longest suffixes from input

word. Morph analyser return correct grammatical information by conducting morphological rules and Adaptive method used to find the root word if not then adopt the similar word which match to the root word.

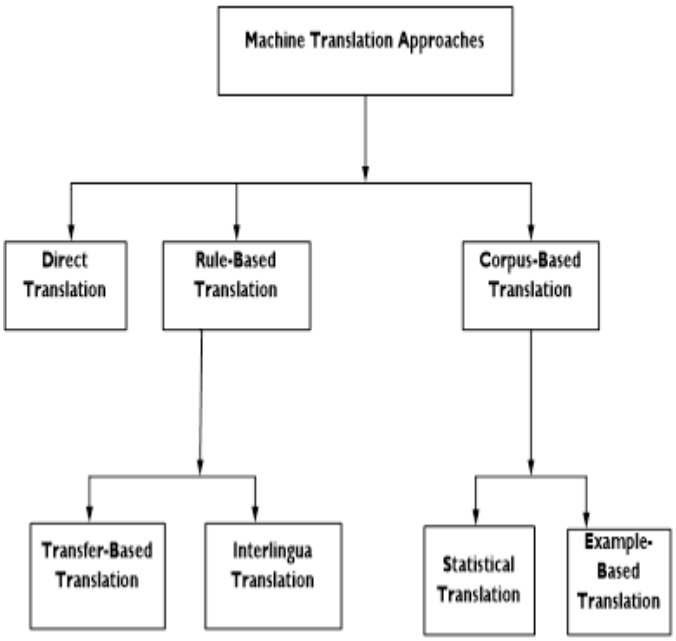

Figure 1: Machine Translation with different approaches

RELATED WORK

Researches in the field of Machine Translation in India are working on various projects of Machine Translation system which is sponsored by Department of Electronics (DoE), state governments etc. since 1990[9] another good effort done by Government of India, has initiated 'Technology Development for Indian Languages' (TDIL) project in the year 1990.TDIL covers number of research done in the field of machine translation in different languages.

Indian languages Machine Translation

Different Indian Researchers are working to improve machine translation system. This work divided into different steps which are as follows:

1) Token Generation Module at word level Tokenize Sentence and Word [14] in a given sentence: - Hindi Sentence is break at word level delimited by punctuation symbol purnviram or question mark "?" and words are delimited by whitespace between two words.
2) Morphological analyses Module: -The minimal parts of words that deliver aspects of meaning to them are called morphemes.
3) Lexemes: - Forms word is expressed in linguistic form in the given context. The concept is to set the alternatives form which can express it.
4) Parts of Speech Tagging [15]: - It tags the token or word with their related Parts of Speech like Noun, Pronoun, verb etc. as shown in Table 1
5) Chunk: - Chunk determines the beginning of phrase and inside the phrase in the sentence for example NP –Noun Phrase and VP- Verb Phrase etc.
6) Parsing: - Parsing means break the Hindi sentence to analyse the syntactic structure of the sentence.
7) Word sense disambiguation: - It is peradventure the most decisive task in the field of machine translation, either supervised and unsupervised approach is used for disambiguation.
8) Ambiguous lexemes: - Hindi word which has two separate lexemes with distinct and unrelated meaning for example **सोना: सोना**[1] can be gold, and **सोना**[2] can be sleep also.

Table 1: . Parts of speech tagset with their Abbreviation

S.No.	Symbol	Parts of speech
1	NN	Noun
2	NNS	Noun Plural
3	NST	Noun denoting spatial and temporal expressions
4	NNP	Proper Nouns
5	PRP	Pronoun
6	DEM	Demonstratives
7	VM	Verb Main
8	VAUX	Verb Auxiliary
9	JJ	Adjective
10	RB	Adverb
11	PSP	Postposition
12	RP	Particle
13	CC	Conjuncts
14	WQ	Question Words
15	QF	Quantifiers
16	QC	Cardinals
17	QO	Ordinals

Indian language is morphologically rich language. This is necessary to defined morphological structure and well defined grammar. Morphology is root of all challenges that are arise in parts of speech tagging as well as this aspect of language is also proved a boon to resolve all problems arise in parts of speech tagging. For good quality of machine translation researchers are concentrating on morphological structure of language to develop better Morphology based parts of speech tagger, word sense disambiguation and Translation of source language to target language. In aspect of implementation of morphological based parts of speech tagger for Hindi language, there are different algorithm based on morphological structure of Hindi language and morphology based database. For implementation of Hindi parts of speech tagger, some intermediate tools like Morph Analysers and Stemmer Analysis and word sense disambiguation tool, are used. Stemming is process to extract original word (root) and longest suffixes from input word. Morph analyser return correct grammatical information by conducting morphological rules. For strong morphological analysis there are various grammatical rules and morphological structure based on rules for determining correct tags from complex sentences. When same word has more than one sense, 'Word Sense Disambiguation' is used to find correct sense and handling the unknown word are still a problem.

i. Authors [16] proposed AnglaBharti adopted Rule based and Example Based methodology. Result-90% acceptable translation in case of simple, compound and complex sentences up to a length of 20 words.

ii. Authors [17] used morphological analyser. The accuracy of the system reaches 69%. The drawback of the system was construction of a bilingual dictionary and work include in the development/adaptation of English parser.

iii. Authors [18] proposed a system Anusaaraka (English-Hindi) based on Paninian grammar formalism and shallow parser approach. Drawback-word sense disambiguation is not resolved

iv. Authors [19] developed a system for Hindi to English machine translation using Context free Grammar parsing technique. Drawback-Case (karaka) and gender is not resolved. Mapping system that connects words and tags of the source language (English) with the corresponding tags for target language (currently considering Hindi).

v. Authors [20]. Approach- Dependency parsing. Result-76.5%. Drawback-person, number gender is not resolved.

vi. Authors [21] used Statistical phrase-based approach for word alignments. They present a model that decouples the steps of lexical selection and lexical reordering with the aim of minimizing the role of word-alignment in machine translation. Drawback-The bag-of-words model per- formed very well in predicting lexical items but was not as good as Moses at ordering them

vii. Authors [22] used Hybrid approach for word alignment for English-Hindi. Result-AER obtained using 270 training sentences 57.06%.

viii. Authors [23] used finite rules like Moses and Stanford Phrasal. BLEU (Bilingual Evaluation Understudy) is an algorithm. Result-Moses 37.4% and Phrasal 29.1%. Drawback-Data was set before training, the English -Hindi corpus (of Indian names) using Phrase based statistical machine translation.

ix. Authors [24] developed Word sense disambiguate algorithm in which they combine supervised and unsupervised method. The accuracy of the work is evaluated for 30 words and produces 80% result.

x. Authors [25] use WordNet tools for several applications.

xi. Authors [26] discuss the development of Hindi WordNet and the co-occurrence vector generated from Hindi Corpus. This approach is used for collocation information, co-occurrence information this information is used to assign different senses for ambiguous word. The accuracy results obtained for 60 lexical semantic ambiguous words, precision obtained is 88.92%.

xii. Authors [27] proposed an extended version of Hindi WordNet. The method is derived mathematically for fuzzy relations and the composition of the fuzzy relations for the extended version. They shows the concept of composition of fuzzy relations can be used to infer a relation between two words that otherwise are not directly related in Hindi WordNet.

xiii. Authors [28] developed a unique approach for polysemy based on the clue words

Foreign languages Machine Translation

i. Authors [29] presents a word senses disambiguation method create the graph for building dictionary. They use semantic proximity to measure words within dictionary. Result obtained 50% for coarse polysemy and 40% for Fine polysemy.

ii. Authors [30] presents a new linear time algorithm for lexical chaining that adopts the assumption of one sense per discourse. The result was compared by two lexical chaining algorithms and their algorithm new linear time algorithm for lexical chaining shows good result.

iii. Authors [31] presents graph-based unsupervised word sezses disambiguation system the problem based on Maximum A Posteriori (MAP) Inference Query on a Markov Random Field (MRF) used WordNet tools.

iv. Author [32] developed a probabilistic posterior inference algorithm for disambiguation. Latent Dirichlet allocation and WORDNET-W ALKs used for component models for other language tasks.

The proposed model for Machine Translation consists of following module which is as follows:

1) The issues identified from the literature review

related to morphology. Hindi is morphological rich language and the word exist in many form. Morphological analyses faces problem in productivity and creativity in languages, word that are not licensed than it will remain unparsed. This is known as unknown word. In this module we aim to identify feature extraction by adding suffix or prefix with the word, this way we can identify new word.

2) Parts of speech tagging plays an important role in Machine Translation there are various approach and method are used for grammatical tagging First is Rule Based is hand-written are used to resolve the ambiguous tag and Second is Hybrid based [34]: this approach is combination of both method Rule Based, Hidden Markov Model and Statistical based. Third is Empirical based this is divided in two types Example and Stochastic based grammatical taggers. Stochastic is based on Hidden Markov Model which uses probability method [33] they combine two methods. First method tagged each token for the given sentence by relative frequency counts. Second method using Hidden Markov Model tag each token in the sequence which maximizes the product of word likelihood. Below is the use of Probability to label tag with their related parts of speech. let T is sequence of word T = T1, T2, Tn

3) And R is different parts of speech, the sequence of tag R = R1, R2..Rn

Let be the select of parts of speech

Where P = product of maximizes the word likelihood.

There are many work carried out by the researcher by using above approach for parts of speech tagging Table 1 [35] shows the results of Parts of speech tagging for various languages. The issues identified from the literature review related to Part of speech tagging. Each token displays Part of speech tagging ambiguity and sometimes same word with different meaning in different context in same sentence. So it is critical problem in Part of speech tagging for Hindi. Such a word has multiple entries in dictionary database

Table 2: Parts of speech tagging for various languages

Chunk symbol	Precision (%)	Recall (%)	F-Score (%)
B-CCP	79.15	67.21	72.97
B-JJP	50.00	10.00	16.67
B-NP	78.17	90.27	83.79
B-RBP	44.83	27.08	33.77
B-VG	76.50	79.76	78.09
I-CCP	42.86	37.50	40.00
I-JJP	100.00	16.67	28.57
I-NP	82.45	71.19	76.41
I-RBP	38.46	27.78	32.26

In this module we aim to improve the performance of Parts of speech tagging for Hindi language with the existing approaches by using Rule Based and Hidden Markov model. We use different part of speech by using Trigrams'n'Tags (TNT) this uses second order Hidden Markov Model and Viterbi algorithm this can be explained with the help of flow chart take a text file in Shakti Standard Format (SSF) this format take input and output specification with part of speech tagging, Chunk and morphological analyser with the information of feature such as root, gender category, person, number, case etc. convert this SSF format in TNT format, this create a sub directory of the text file apply Viterbi algorithm for parts of speech tagging. For unknown word we apply suffix smoothing. This method will improve the result of Part of speech tagging.

In this module we aim to improve the performance of Parts of speech tagging for Hindi language with the existing approaches by using Rule Based and Hidden

Markov model. We use different part of speech by using Trigrams'n'Tags (TNT) this uses second order Hidden Markov Model and Viterbi algorithm this can be explained with the help of flow chart take a text file in Shakti Standard Format (SSF) this format take input and output specification with part of speech tagging, Chunk and morphological analyser with the information of feature such as root, gender category, person, number, case etc. convert this SSF format in TNT format, this create a sub directory of the text file apply Viterbi algorithm for parts of speech tagging. For unknown word we apply suffix smoothing. This method will improve the result of Part of speech tagging.

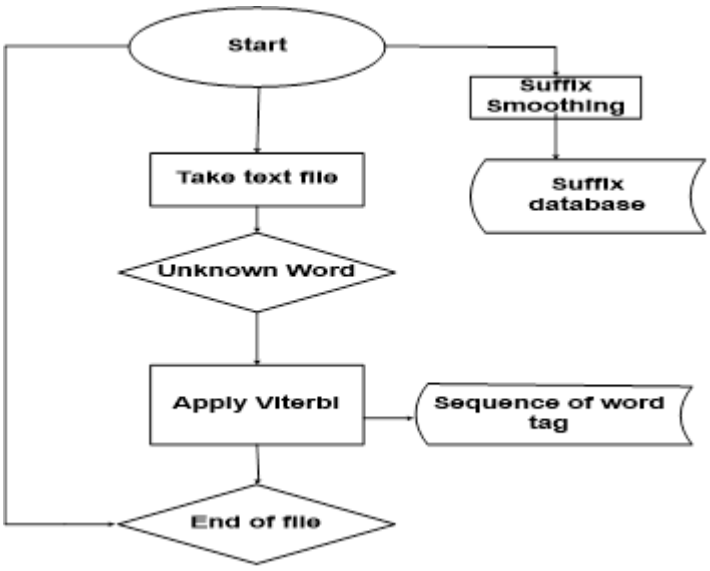

Figure 2: Flow chart of Part of speech tagging

1) The issues identified from the literature review related to Chunk. The role of chunk is to identify different phrases such as B-CCP, B-JJP, B-NP, B-RBP, B-VG, I-CCP, I-JJP, I-NP, I-RBP, I-VG etc. in a given sentence. Where B represents beginning phrase and I represent inside phrase. Various

approaches are used such as maximum entropy models [36]. There are many work carried out by the researcher by using above approach for Chunk. Table 2 shows accuracy result of Precision, Recall and F-Score for Chunk [36] and Table 3 shows Chunk result for various languages. This method will improve the result of Chunk.

Table 2: Accuracy result of Precision, Recall and F-Score for Chunk

Language	No. of Test token	Correct Tag	Accuracy
Bengali	5225	3897	74.58%
Hindi	4924	3858	78.35%
Telugu	5193	3909	75.27%

Table 3: Chunk result for various languages

Language	Result (%)
Telugu	78.15
Bengali	81.74
Hindi	79.97

In this module we aim to improve the performance of Chunk for Hindi language with the existing approaches by using maximum entropy models. Our approach to label chunk the sentence is divided into number of Chunk. We count the chunk with maximum probability of part of speech tags.

1) Parsing there are various approach used such as machine learning approach in which data-driven dependency parsing [37]. This approach shows the combination of two clause inter- and intra-clausal relations. In this module we aim to combine different part of speech and Chunk by using second order Hidden Markov Model, Viterbi algorithm and maximum entropy models. To train the data we use Conditional Random Field (CRF). For simulation we create confusion matrix and compare our result of part of speech and chunk for Hindi sentence with Gold Standard. Gold Standard is a training data which is collection of correct part of speech and chunk for Hindi sentence. Figure 3 shows the overall flow chart of the Parsing. This method will improve the result of Parsing.
2) Construction of electronic dictionary we can use Hindi and English WordNet which understood the data structure and directly obtains result. This can be done by lookup operation.
3) The issues identified from the literature review related to Hindi language. Hindi contains multi-sense words for example Hindi word सोना have contains multi-sense 'सोना1' meaning in English is gold and another सोना2 meaning in English sleep. Some drawback related to Word sense disambiguation are as follows:
   i. Same word are multiple tag in given Parts of speech, polysemy words has different senses in specified domains.
   ii. To identify the correct meanings in multi-sense words in Hindi languages.

   In this module we aim to resolve Word sense disambiguation that will automatically decide the correct meaning of an ambiguous word based on the surrounding context in which they appears. A methodology is based on Morphological analysis, Part of speech tagging, Parsing and Word sense disambiguation. We used machine learning techniques such as supervised, unsupervised, overlap based method and domain specific sense with the information of WordNet tool. This approach and method will resolve the problem of Word sense disambiguation Figure 4 shows the overall flow chart of the Word sense disambiguation.
4) There are various translating website are available such as Google Translator [38] and Babelfish Translator [39] which are failed to resolve polysemy word in Hindi to English Translation. Table 4 shows the output of Goggle Translator and Babelfish Translator. In this table we shows that both translators are failing to resolve Word sense disambiguation for Hindi to English language.

CONCLUSION

In this paper we have illustrated about Machine Translation system there are various approach and method was discussed in this paper to develop Machine Translation system but still we find some drawback such as Morphological analyser, Part of speech tagging, Chunk, Parsing and Word sense disambiguation. In future we will try to develop Hindi to English language Machine Translation system in which we will improve the performance of Morphological analyser, Part of speech tagging, Chunk, Parsing and Word sense disambiguation. In the field of Machine Translation Word sense disambiguation plays important role to produce correct translation. We follow supervised unsupervised and domain specific sense disambiguation. We use Hindi English WordNet tool which is based on dictionary knowledge like gloss overlaps. This will improve the performance of Word sense disambiguation and improve the performance of Hindi to English language machine translation.

S.No.	Input Hindi Sentence	Output Google English Translation	Output Babelfish English Translation	Correct Translation
1	उसे **कर** देना पडेगा।	He will have to give.	I have to give him.	He will have to pay the tax.
2	इस काम को पूरा कर दो।	Let's complete the task.	Complete this work.	Do Complete this work.
3	मुझे मेरे **कर** पर विश्वास है।	I have my trust.	I do believe me.	I have faith on my hand.
4	खून का रंग **लाल** होता है।	Red is the color of blood.	The color of blood red.	The colour of the blood is red.
5	मेरे **लाल** की वाहवाहीहो रही है ।	My red is Wahwahiho.	.Woohoo my red.	My child is getting the applause.
6	क्या आप **सोना** चाहते हैं?	do you want to sleep?	Do you want to sleep?	Do you want gold or Do you want to sleep?
7	यह पा**त्र** मुझे दे दो ।	Give me the character.	It give me eligible	Give me this utensil or role/character to act/to play.
8	तुम इस नौकरी के **पात्र** हो ।	You become eligible for the job.	You deserve this job.	You are eligible for this job.
9	मैं तुम्हारे **प्रति** आस्था खोता जा रहा हूँ ।	I'm losing faith towards you	I'm going to lose faith in you	I am continuing to lose belief on you.
10	**प्रति** दिन का किराया क्या होगा?	Rent per day, what will happen?	Rent per day?	What will be the rent per day?

REFERENCES

[1] Xiao, T., Wong, D. F., & Zhu, J. (2016). A Loss-Augmented Approach to Training Syntactic Machine Translation Systems. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(11), 2069-2083.

[2] Hettige, B., & Karunananda, A. S. (2006, December). A Morphological analyzer to enable English to Sinhala Machine Translation. In 2006 International Conference on Information and Automation (pp. 21-26). IEEE.

[3] Bellegarda, J. R. (2010). Part-of-Speech tagging by latent analogy. IEEE Journal of Selected Topics in Signal Processing, 4(6), 985-993.

[4] Tran, T., & Nguyen, D. T. (2016, November). Method of Mapping Vietnamese Chunked Sentences to Definite Shallow Structures. In Advanced Computing and Applications

[5] (ACOMP), 2016 International Conference on (pp. 74-80), IEEE.

[6] Li, Z., Zhang, M., Che, W., Liu, T., & Chen, W. (2014). Joint Optimization for Chinese POS Tagging and Dependency Parsing. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(1), 274-286.

[7] Lopez, F. R., Arevalo, I. L., Pinto, D., & Sosa, V. J. S. (2015). Context Expansion for Domain-Specific Word Sense Disambiguation. IEEE Latin America Transactions, 13(3), 784-789.

[8] Hwang, M., Choi, C., & Kim, P. K. (2011). Automatic enrichment of semantic relation network and its application to word sense disambiguation. IEEE Transactions on Knowledge and Data Engineering, 23(6), 845-858.

[9] Rajpirathap, S., Sheeyam, S., Umasuthan, K., & Chelvarajah, A. (2015, September). Real-time direct translation system for Sinhala and Tamil languages. In Computer Science and Information Systems (FedCSIS), 2015 Federated Conference on (pp. 1437-1443). IEEE.

[10] Chen, Y., & Wang, J. Z. (2003). Support vector learning for fuzzy rule-based classification systems. IEEE Transactions on Fuzzy Systems, 11(6), 716-728.

[11] Winiwarter, W. (2007, October). Machine translation using corpus-based acquisition of transfer rules. In Digital Information Management, 2007. ICDIM'07. 2nd International Conference on (Vol. 1, pp. 345-350). IEEE.

[12] Xiong, W., & Jin, Y. (2011, November). A new Chinese-English machine translation method based on rule for claims sentence of Chinese patent. In Natural Language Processing andKnowledge Engineering (NLP-KE), 2011 7th International Conference on (pp. 378-381). IEEE.

[13] Zhang, M., Liu, Y., Luan, H., & Sun, M. (2016). Listwise Ranking Functions for Statistical Machine Translation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 24(8), 1464-1472.

[14] Alawneh, M. F., & Sembok, T. M. (2011, September). Rule-based and example-based machine translation from English to Arabic. In Bio-Inspired Computing: Theories and Applications (BIC-TA), 2011 Sixth International Conference on (pp. 343-347). IEEE.

[15] Mall, S., & Jaiswal, U. C. (2015, October). Innovative algorithms for Parts of Speech Tagging in hindi-english machine translation language. In Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on (pp. 709-714). IEEE.

[16] Pammi, S. C., & Prahallad, K. (2007, January). POS tagging and chunking using decision forests. In IJCAI Workshop on Shallow Parsing for South Asian Languages (pp. 33-36).

[17] Sinha, R. M. K., & Thakur, A. (2005). Machine translation of bi-lingual hindi-english (hinglish) text. 10th Machine Translation summit (MT Summit X), Phuket, Thailand, 149-156.

[18] Hettige, B., & Karunananda, A. S. (2006, August). A Parser for Sinhala Language-First Step Towards English to Sinhala Machine Translation. In First International Conference on Industrial and Information Systems (pp. 583-587). IEEE.

[19] Chaudhury, S., Rao, A., & Sharma, D. M. (2010, August). Anusaaraka: An expert system based machine translation system. In Natural Language Processing and Knowledge Engineering (NLP-KE), 2010 International Conference on (pp. 1-6). IEEE.

[20] Sugandhi, R. S., Shekhar, R., Agarwal, T., Bedi, R. K., & Wadhai, V. M. (2011, December). Issues in Parsing for Machine Aided Translation from English to Hindi. In Information and Communication Technologies (WICT), 2011 World Congress on (pp. 754-759). IEEE.

[21] Ambati, B. R., Husain, S., Jain, S., Sharma, D. M., & Sangal, R. (2010, June). Two methods to incorporate local morphosyntactic features in Hindi dependency parsing. In Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages (pp. 22-30). Association for Computational Linguistics.

[22] Venkatapathy, S., & Bangalore, S. (2009). Discriminative machine translation using global lexical selection. ACM Transactions on Asian Language Information Processing (TALIP), 8(2), 8.

[23] Srivastava, J., & Sanyal, S. (2012, November). A hybrid approach for word alignment in english-hindi parallel corpora with scarce resources. In Asian Language Processing (IALP), 2012 International Conference on (pp. 185-188). IEEE.

[24] Halder, M., & Tyagi, A. D. (2013, December). English-Hindi Transliteration by Applying Finite Rules to Data before Training Using Statistical Machine Translation. In IT Convergence and Security (ICITCS), 2013 International Conference on (pp. 1-4). IEEE.

[25] Saktel, P., & Shrawankar, U. (2012, March). Context based Meaning Extraction for HCI using WSD algorithm: A review. In Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on (pp. 208-212). IEEE.

[26] Noor, N. M. M., Ali, N. H., & Ibrahim, N. S. (2010, June). A new framework to extract WordNet lexicographer files for semi-formal notation: a preliminary study. In 2010 International Symposium on Information Technology (Vol. 2, pp. 1027-1031). IEEE.

[27] Agarwal, M., & Bajpai, J. (2014, August). Correlation based Word Sense Disambiguation. In Contemporary Computing (IC3), 2014 Seventh International Conference on (pp. 382-386). IEEE.

[28] Jain, A., & Lobiyal, D. K. (2016). Fuzzy Hindi WordNet and Word Sense Disambiguation Using Fuzzy Graph Connectivity Measures. ACM Transactions on Asian and Low-Resource Language Information Processing, 15(2), 8.

[29] Dhungana, U. R., Shakya, S., Baral, K., & Sharma, B. (2015, February). Word Sense Disambiguation using WSD specific WordNet of polysemy words. In Semantic Computing (ICSC), 2015 IEEE International Conference

on (pp. 148-152). IEEE.

[30] Gaume, B., Hathout, N., & Muller, P. (2004, August). Word sense disambiguation using a dictionary for sense similarity measure. In Proceedings of the 20th international conference on Computational Linguistics (p. 1194). Association for Computational Linguistics.

[31] Galley, M., & McKeown, K. (2003, August). Improving word sense disambiguation in lexical chaining. In IJCAI (Vol. 3, pp. 1486-1488).

[32] Chaplot, D. S., Bhattacharyya, P., & Paranjape, A. (2015, January). Unsupervised Word Sense Disambiguation Using Markov Random Field and Dependency Parser. In AAAI (pp. 2217-2223).

[33] Boyd-Graber, J. L., Blei, D. M., & Zhu, X. (2007, June). A Topic Model for Word Sense Disambiguation. In EMNLP-CoNLL (pp. 1024-1033).

[34] Merialdo, B. (1994). Tagging English text with a probabilistic model. Computational linguistics, 20(2), 155-171.

[35] Sunitha, C. (2015, August). A hybrid Parts Of Speech tagger for Malayalam language. In Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on (pp. 1502-1507). IEEE.

[36] Sastry, G. R., Chaudhuri, S., & Reddy, P. N. (2007). An HMM based Part-Of-Speech tagger and statistical chunker for 3 Indian languages. Shallow Parsing for South Asian Languages, 13.

[37] Rao, D., & Yarowsky, D. (2007). Part of speech tagging and shallow parsing of Indian languages. Shallow Parsing for South Asian Languages, 17.

[38] Husain, S., Gadde, P., Nivre, J., & Sangal, R. (2011). Clausal parsing helps data-driven dependency parsing: Experiments with Hindi. In IJCNLP (pp. 1279-1287).

[39] https://translate.google.co.in/

[40] https://www.babelfish.com/