Hindi to English Machine Translation

Sheetal Mahadik, Priya Gour, Paras Tank, Shubham Yadav, Abhishek Vishwakarma
Department of Electronics Engineering
Shree L.R Tiwari College of Engineering

Abstract:- Machine Translation (MT) is a procedure in Natural Language Processing (NLP), where the automatic systems are used to translate the text from one language to another language without changing the meaning of source language. In this work, we provide our efforts in developing a rule-based translation system on the Analyze-Transfer-Generate paradigm which employs morphological and syntactic analysis of source language. We utilized shallow parser for Hindi language along with dependency parse labels for syntactic analysis of Hindi language, developed modules for transfer of Hindi to English and generation of English language. Due to wide difference in word order of the two languages (Hindi following SOV and English SVO word order), a lot of re-ordering rules need to be crafted to capture the irregularity of the language pair. As a result of drawbacks of the aforementioned approach, we shifted to statistical methods for developing a system. A wide variety of machine translation approaches have been developed in past years. As each model has its pros and cons, we propose an approach where we try to capture the advantages of each system, thereby developing a better MT system. We then incorporate semantic information in phrase-based machine translation using monolingual corpus where the system learns semantically meaningful representation.

I. INTRODUCTION

Machine Translation is a sub-field of Machine Learning, where automatic translation of a sentence or a document or number of documents is done from one language to another language. In the past decades multiple approaches for solving complex MT problems have been suggested with each approach having its pros and cons. The system combination approach combines the positives from the corresponding translation of each system thereby performing better than each of the individual systems.
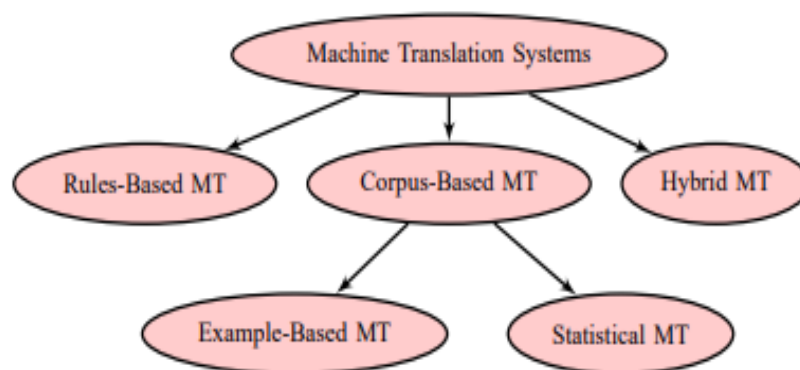

Fig 1:- Machine Translation Systems

A. Rule Based Approach

A rule based machine translation contains various sets of rules called grammar rules, lexicon and software programs.

B. Corpus Based Approach

Corpus is a large-scale database with tremendous collection of linguistic (language) information in real use, which is provided for retrieval of data by computers for research purpose.

C. Hybrid Approach

Hybrid solutions are the combination of advantages of individual approaches to achieve an overall best translation.

➤ Language Barrier

The majority of the Indians, especially the remote villagers, doesn't have any knowledge about English language; therefore implementing an efficient language translator is needed. Machine translation systems, that translate text from one language to another, will help the knowledgeable society of Indians by reducing any language barrier. The major use can be seen in areas of news articles translations and the translation of documents in judicial systems. The hearings and document works in regional law courts are performed in local language, which creates an issue when the case moves to high court where English is the official language. The time taken by the human translators to translate each document in local language to English postpones the justice and weakens the judicial system.

> *Language Ambiguities*

Ambiguity in language is often an obstacle to be ignored or a problem to be solved for people to understand each other. Ambiguities found in languages are:

- *Lexical Ambiguities*

When one single word is understood in two possible senses or ways than it is lexical ambiguities. Examples like:
✓ Read the "book".
✓ "Book" the flight.

Here in first statement word "book" is acting like "*noun*" whereas in second statement same word acts like "*verb*".

- *Homonym*

Different words are pronounced or spelled in a same manner but have different meanings. Examples like:
✓ to, too, & two.
✓ shit & sheet
✓ by & bye

- *Metonym*

A word used in place of another word or expression to convey the same meaning. Examples like:
✓ "Hand" used in place of "Help"
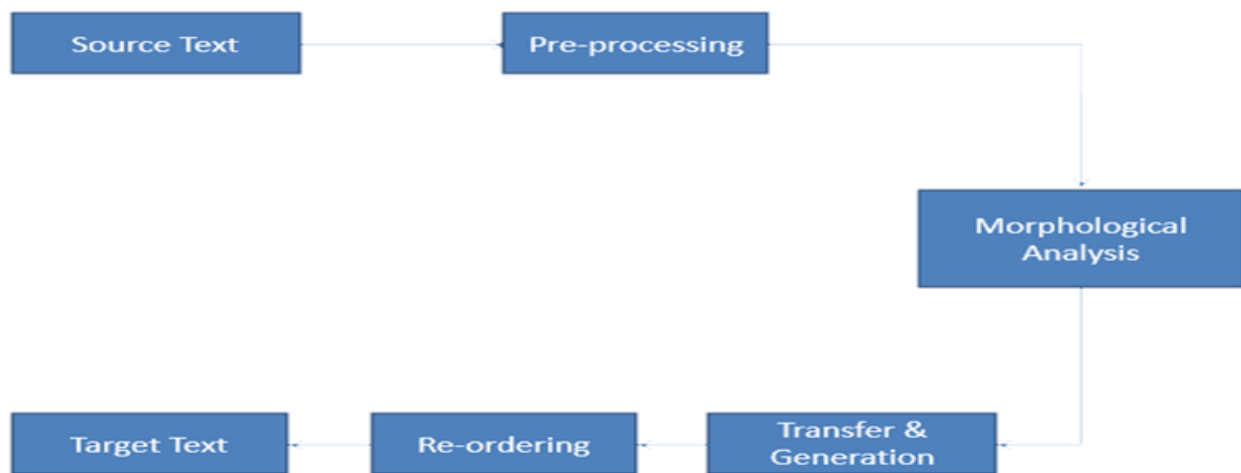✓ "Tongue" used in place of "Language"

II.   BLOCK DIAGRAM

Fig 2:- Block Diagram
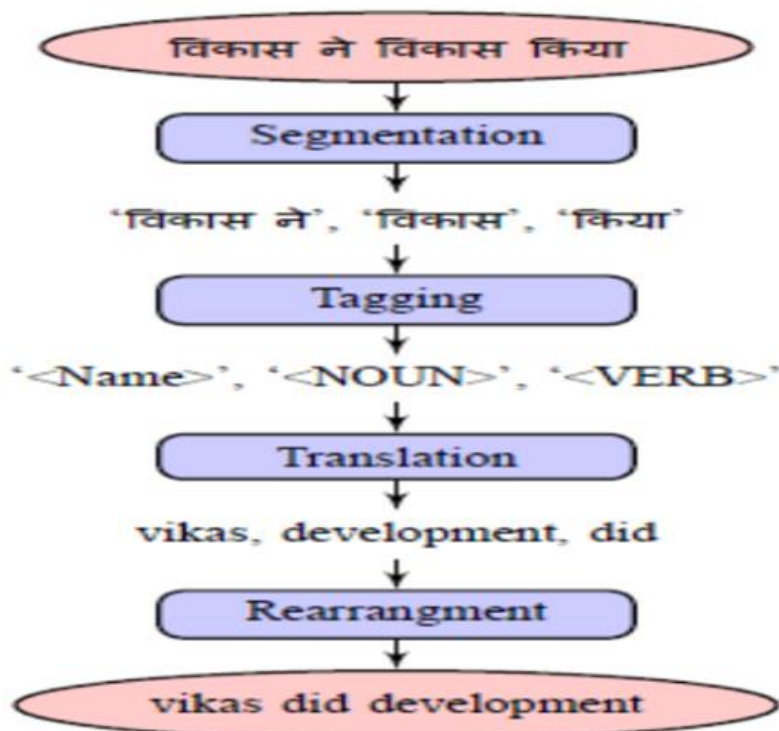
III.   FLOWCHART

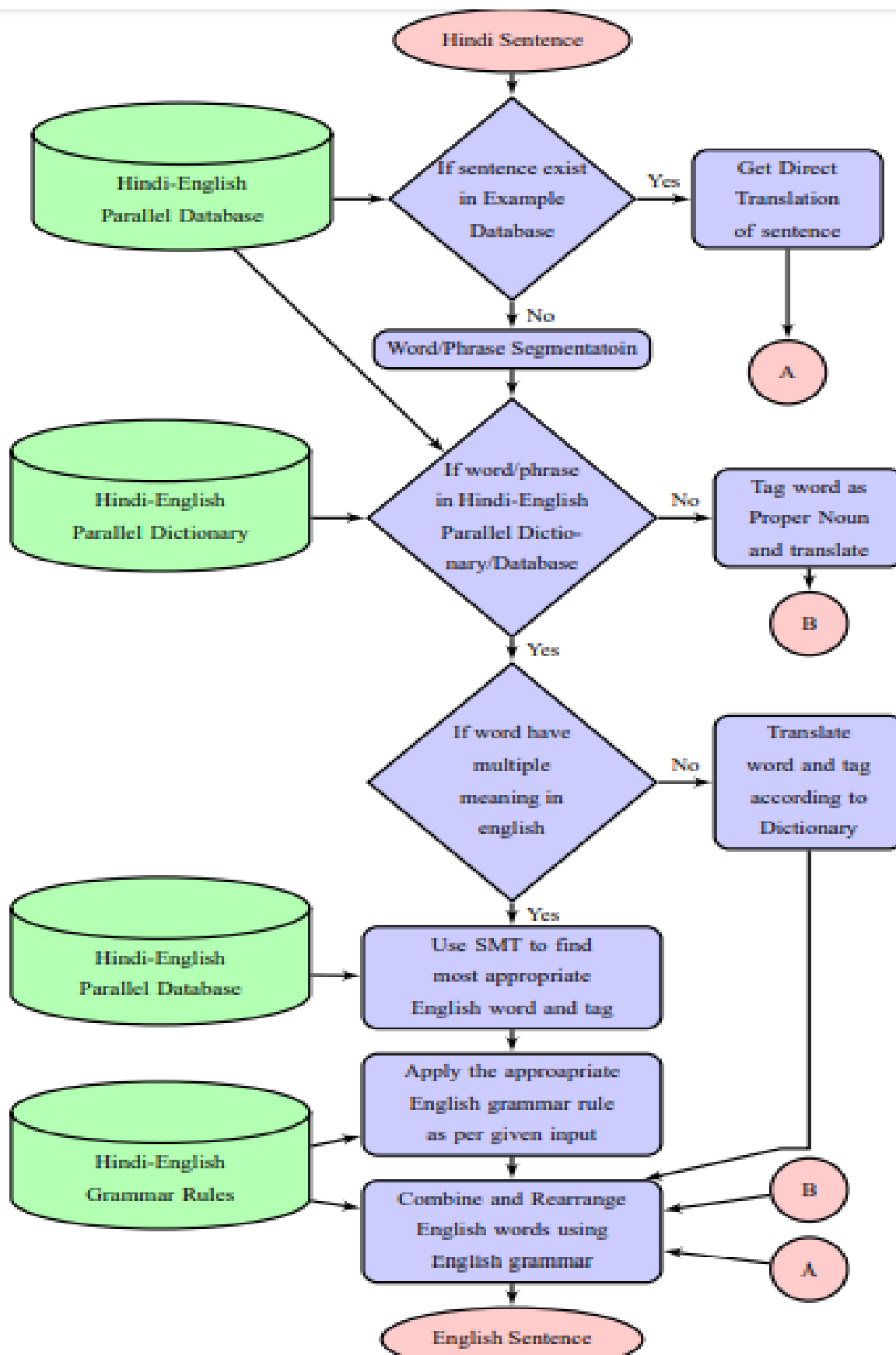Fig 3:- Flowchart

www.ijisrt.com

IV. ALGORITHM

Fig 4:- Algorithm

V. SOFTWARE REQUIREMENT

A. Language

Python is a high-level interpreted programming language for programming in which all the libraries used for machine translation are available. Created by "*Guido van Rossum*" and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace. It provides constructs that enable clear programming on both small and large scales. It is a muli-paradigm programming language.

B. Libraries

➤ **NLTK** *(Natural Language Tool Kit).*
   NLTK features:
- Classification.
- Tokenization.
- Stemming.
- Tagging.
- Parsing.

➤ **SpaCy**
   SpaCy features:
- Named Entity Recognition.
- Text Classification.
- Part-of-speech Tagging.
- Efficient Binary Serialization.

➤ **Tensor Flow**
   Tensor Flow features:
- Provides stable Python and C API's.
- Compatibility with C++, Go, Java,etc.

➤ **Keras**
   Keras features:
- Contains numbers of neural network building blocks.
- Provide host of tools to work with images and text data easily.

C. Tools

The software used for the coding purpose is integrated development environment (IDE) known as IntelliJ IDE. The software supports number of language such as JAVA, Python, Groovy, Scala, Perl, Rust etc.

VI. CONCLUSION

In this paper we have listed the techniques of translating one language into another one by applying various rules of both languages used. Also we have stated problems faced during translation. Further we will be solving the problems which are been faced in the same and will try get as many solutions as possible. After achieving the target, more functions will be added to it.

REFERENCES

[1]. M. Chandhana Surabhi, "NATURAL LANGUAGE PROCESSING FUTURE", *IEEE Optical Imaging Sensor and Security (ICOSS) 2-3 JULY 2013, Coimbatore, India.*

[2]. Jayashree Nair, "An Efficient Machine Translation System Using Hybrid Mechanism", *IEEE Advances in Computing Communications and Informatics (ICACCI) 21-24 SEPTEMBER 2016, Jaipur, India.*

[3]. M.A. Perez-Quinones, O.I. Padilla-Falto, K. McDevitt, "Automatic Language Translation For User Interface" ,*Richard Tapia Celebration of Diversity in Computing Conference,19-20 October 2005,USA.*

[4]. Sheena Angra, Sachin Ahuja, "Machine Learning and Its Application", *International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), 23-25 March 2017, Chirala, India.*

[5]. ieeexplore.ieee.org/document/7078789/

[6]. en.wikipedia.org/wiki/Machine_translation

[7]. www.sdltrados.com/solutions/machine-translation/

[8]. githubengineering.com/

[9]. www.leadingindia.ai/

[10]. en.wikipedia.org/wiki/Natural_language_processing

[11]. www.nlp.com/what-is-nlp/