Machine Translation of English to Ahirani Language: A Review

Uday Chandrakant Patkar[1], Dr. Suhas Haribhau Patil [2], Dr. Prasadu Peddi [3]

[1]*Research Scholar, JJTU*
[2]*Professor, Bharati Vidyapeeth Deemed University College of Engineering, Katraj, Pune*
[3]*Professor, Shri Jagdish Prasad Jhabarmal Tibrewala University*

---***---

Abstract - *Internet users around the world are growing rapidly. According to Internet live statistics, today there are more than 3 billion Internet users in the world, and the number of non-English speakers there is quite large. People with different language backgrounds could not interact with each other. This translation concept will help people communicate comfortably. Also, it will help fill the communication gap between two linguistically different grounds. It will help people in villages who have taken English language education. While most of the Indian population is not familiar with English, most of the information available on the web and in electronic information is English. So, to reach out to ordinary people in various sections and throughout the field, an automatic language translator is important. The main purpose of the Machine translation (MT) is to break the language barrier, like a multilingual nation.MT gives you a way to translate source languages into target languages. This paper study and analyzes the language translation system, types and conduct review on English to Ahirani machine translation language.*

Key Words: **Language, Machine translator (MT), Communication.**

1. INTRODUCTION

Language is the ability to express one's thoughts through a series of signs, including graphical gestures, acoustics, and even music. It is a unique property a man who is the only living being to use such a structured system. Speech is one of its main components. It is by far the oldest means of communication between humans and even the most widely used. So, people studied it extensively and often tried to make a machine for processing it acoustically. More information about access to the digital world, access rights few people read and understand a particular language. Language technology is a natural interface so that digital content can reach the masses and facilitate the exchange of information between different people who speak different languages; these technologies play an important role in multilingual societies like India with a dialect/native language of about 1652[1].

Machine translation (MT) is the Computational Linguistics field, which explores the use of software text or word translated from one natural language into another natural language. Machine translation performs a simple translation of words in one natural language for another language, but it may not lead to a good translation of the text, i.e. Recognition of whole phrases and their equivalent meanings must be present in the target language. Machine translation refers to the use of a computer to translate some or the entire translation task between human languages. The development of any bilingual machine translation system with the use of electronic resources and tools is a challenging task.

Multilingual systems are bidirectional, but most bilingual systems are unidirectional. Machine translation techniques are usually divided into direct, transfer and Interlingual. Methodologies differ in the analysis of SL and the extent to which language-independent representation is achieved between the source and target languages. Barriers to quality machine translation may be related to ambiguity in natural languages. Ambiguities are divided into two types: structural ambiguity and lexical ambiguity. India is a linguistically rich region. It has 22 constitutional languages and is written in 10 different scripts. Hindi is the official language of the Union. Many of the states have their regional languages, which are either Hindi or other constitutional languages. Also, English is very widely used in the media, commerce and science and technology, and about 5% of the world's population speaks English as their native language. In this situation, there is a large market for translations between English and various Indian languages [2]. The various Indian languages are translated such as Marathi, Hindi, Tamil, Telugu, and Bengali [3] [4]. Spoken languages like Ahirani are not explored until now. Ahirani language is one of the most common languages spoken in Khandesh. Khandesh region mainly constitutes Dhulia, Jalgaon and Nandurbar districts. Table 1 shows the Ahirani words.

Table -1: Ahirani Words change language into English

KARE	LAGIN	UKHADLA	VAKKHAR	CHULA	NHARI
UBYA	JAPIJAY	BHAITAN	PHAPUDA	KHUDA	GHATIYA

2. APPROACHES OF MACHINE TRANSLATION

Generally, machine translation is divided into seven categories: Rule-based, Statistical, Hybrid, Example-based, Knowledge-based, Principle-based, and online interactive methods. The first three approaches to machine translation are widely used. Figure 1 shows the classification of MT in Natural language Processing (NLP) [2].

2.1 Rule-Based Machine Translation (RBMT):

The rule-based machine translation (RBMT) system was developed by human experts in the early 1970 by mapping the structure from the source language to the target language, and also the first RBMT system over time. A rule-based system can be very costly and time-consuming to implement and maintain. As the rules are added and updated from time to time, these systems may have ambiguity or translation degradation over time [2] [4].

2.2 Statistical Machine Translation:

The statistical models consist of words and phrases learned automatically from bilingual parallel sentences, creating a bilingual "database" of translations. The attraction of statistical systems comes from the level of automation in building new systems using its machine learning capabilities, which results in fast turnaround times and low-cost processing power required to build and operate these statistical models.

2.3 Hybrid Machine Translation:

The hybrid-based approach utilizes both statistical and rule-based translation methodologies that have been proven to have better efficiency in the field of MT systems currently; several government and private-based MT sectors are using this hybrid-based approach and use both source-based and rule-based translation methodologies. The hybrid approach can be used in many different ways. In some cases, the translation is performed at the first stage, using a rule-based approach, and then using statistics to adjust or correct the output. In other ways, the rules are used to preprocess the input data, as well as post-

processing the statistical output of a statistical-based translation system. This technique is better than the previous two techniques and has more power, flexibility, and control over translation.

Figure-1: Machine Translation Classification [2]
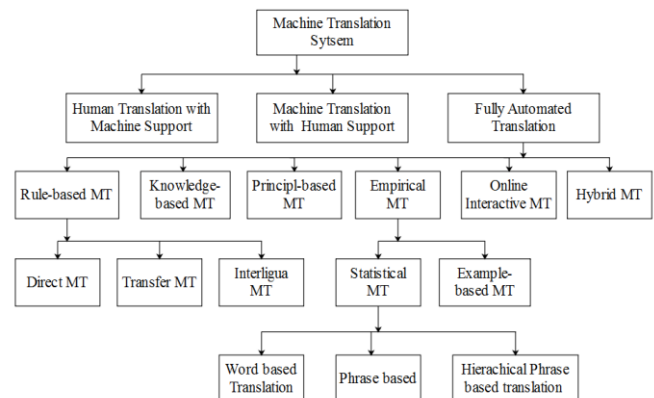
3. LITERATURE SURVEY

Kale Sunil Digambar, Rajesh S Prasad [3], a strategy was proposed for the identification of the original author of the document written in Marathi. Here they used a set of fine lexical and stylistic features for the analysis of the text, and have developed two different models, a Sequential Minimal optimization with Rule-Based Decision Tree Approach .Partha Mukherjee, Soumen Santra,Subhajit Bhowmick, Ananya Paul [4], the author developed a text-to-speech synthesis synthesizer that analyzes and processes text using natural language processing (NLP) and converts text into synthesized speech representation using digital signal processing (DSP) technology. Here is a simple application that reads out to the user can convert the entered text into a synthesized voice and then save it as an mp3 file. Hussain Rangoonwala, Vishal Kaushik, P Mohith and Dhana Lakshmi Samiappan [5], the author propose a method to develop a complete system that can convert text to speech, convert text files to audio, convert text to audio from various languages, convert images to text, and convert images to audio using MATLAB as a programming tool. The various methods used include preprocessing, Unicode conversion, segmentation, concatenation, rhyme, and so on, which are then combined in the application for easy access and ease of use.

Nagmani Wanjaria, Prof. G. M. Dhopavkarb, Nutan B. Zungre [6], this paper proposes a rule-based system to correctly identify the boundaries of sentences written in Marathi. The task of identifying the end of a sentence in Marathi is complicated by the fact that there is no indication of the

beginning of a sentence, for example in English. The system uses certain rules to correctly determine the end of a sentence. S. Venkateswarlu1, D. B. K. Kamesh, J. K. R. Sastry and Radhika Rani [7], introduce an innovative, efficient, and cost-effective technology that allows users to hear the contents of text images instead of reading the contents of text images. It combines the concept of Optical Character Recognition (OCR) and text to speech (TTS) with the Raspberry Pi. This system is an effective vocal interface with a computer that interacts with visually impaired people. Jisha Gopinath, Aravind S, Pooja Chandran, Saranya S [8], The purpose of this paper is to recognize text characters and convert them into audio signals. Manjare Chandraprabha Anil and S. D. Shirbahadurkar [9], the sensitivity of the analysis in this paper relates to speech pitch patterns in words. The pitch feature is derived from punctuation and emotional speech samples. Question marks and exclamation marks in the text are now studied to find prosodic information. First, the pitch features obtained from emotional speech samples are compared with those obtained from neutral speech. Speakers usually tend to emphasize specific syllables within the prosodic word. Itunuoluwa Isewon, Jelili Oyelade and Olufunke Oladipupo [10], they are implementing a useful text-to-speech synthesis file in the form of a simple application that translates the entered text into a synthesized voice and reads it aloud to the user, which can be saved as an mp3. Chaw Su Thu, Theingi Zin [11], this paper is focused on optical character recognition, speech synthesis technology and to develop an effective convenient image to speech conversion system using MATLAB. In this work, the OCR system is implemented by recognizing the capital of the English character A to Z and the number 0 to 9. Each character is recognized simultaneously. The recognized character is saved as text in the notepad file. Rahul Raghavendra Joshi, Maya Shelke, Joya Prakash, Swapnil Bhadhane [12], This paper describes a software interface which accepts Marathi sentences as input, carries out the accurate English translation, identifies the parts of speech, carries out the transliteration to the accurate English, and translates them into English.

K. R. Aida Zade, C. Ardil and A.M. Sharifova [13], this paper describes the main principle of the speech synthesis system. Describe the related problems in developing speech synthesis systems. This paper also presents an approach to be used in the speech synthesis system of the Azerbaijani language and its application. Manoj Kumar Chinnakotla, Sagar Ranadive, Om P. Damani, and Pushpak Bhattacharyya [14], this article present Hindi for English and Marathi for English CLIR systems developed as part of this participation in the special bilingual task CLEF 2007. They use a query-

based translation approach using bilingual dictionaries. Query words not found in the dictionary are transliterated using a simple rule-based transliteration approach. The resulting transliteration is then compared to the unique hull words to return the "k" words most similar to the transliterated word. Satoshi Nakamura, Konstantin Markov and Hirofumi Yamamoto [15], this article describes the ATR multilingual speech-to-speech translation (S2ST) system, which is mainly focused on translation between English and Asian languages (Japanese and Chinese). There are three main modules of this S2ST system: continuous speech recognition with a large vocabulary, machine text-to-text translation (T2T) and text-to-speech synthesis. They are all multilingual and designed using the latest technology developed at ATR.

4. CONCLUSION

In this paper, the translation of Indian languages presented and showed different approaches to machine translation. Machine translation is a large-scale translation system from the source language to the target language. India is a nation where great diversity is observed in culture with diversity in spoken language. People of different cultures and language bases can't easily communicate where the translation system makes it easy to fill that gap. This paper also summarizes the Ahirani language is one of the most common languages spoken in Khandesh.

REFERENCES

[1] Poonam's. shetake, S.a. patil, P. m jadhav, "Review of text to speech conversion methods", IJIEAEE 2014

[2] Dr. Siddhartha Ghosh, SujataThamke and Kalyani U.R.S, "Translation of telugu-marathi and vice-versa using rule-based machine translation", Computer Science Published in ArXiv 2014DOI:10.5121/csit.2014.4501

[3] Kale Sunil Digambar, Rajesh S Prasad, "Author Identification using Sequential Minimal Optimization with Rule Based Decision Tree on Indian Literature in Marathi", International Conference on Computational Intelligence and Data Science ICCIDS, 2018

[4] Partha Mukherjee, Soumen Santra, Subhajit Bhowmick, Ananya Paul, and ArpanDeyasi, "Development of GUI for Text-to-Speech Recognition using Natural Language Processing", TENCON IEEE -2018.

[5] Hussain Rangoonwala, Vishal Kaushik, P Mohith and Dhana Lakshmi Samiappan; "Text to Speech Conversion Module; International Journal of Pure and Applied Mathematics, Volume 115 No. 6, 2017.

[6] Nagmani Wanjaria, Prof. G. M. Dhopavkarb, Nutan B. Zungre, "Sentence Boundary Detection for Marathi Language", International Conference on Information Security & Privacy (ICISP2015), 11-12, Nagpur, INDIA, Elesevir-2016.

[7] S. Venkateswarlu1, D. B. K. Kamesh, J. K. R. Sastry and Radhika Rani, "Text to speech conversion", Indian Journal of Science and Technology, Vol 9(38), -2016.

[8] Jisha Gopinath, Aravind S, Pooja Chandran, Saranya S, "Text to Speech Conversion System using OCR"; International Journal of Emerging Technology and Advanced Engineering (IJETAE), Volume 5, Issue 1, January 2015.

[9] Manjare Chandraprabha Anil and S. D. Shirbahadurkar, "Speech Modification for prosody Conversion in Expressive Marathi Text to Speech Synthesis", International Conference on Signal Processing and Integrated Networks (SPIN), 2014

[10] Itunuoluwa Isewon, Jelili Oyelade and Olufunke Oladipupo, "Design and Implementation of Text to Speech Conversion for Visually Impaired People", International Journal of Applied Information Systems (IJAIS) Volume 7–No. 2, 2014

[11] Chaw Su Thu, Theingi Zin, "Implementation of Text to Speech Conversion", International Journal of Engineering Research & Technology (IJERT) Vol. 3 Issue 3, 2014

[12] Rahul Raghavendra Joshi, Maya Shelke, Joya Prakash, Swapnil Bhadhane, "Developing the Translation Interface for Marathi Text", International Journal of Advanced Engineering Research and Studies-2013.

[13] K. R. Aida Zade, C. Ardil and A.M. Sharifova, "The Main Principle of Text to Speech Synthesis System"; International Journal of Computer and Information Engineering Vol:7, No:3, 2013.

[14] Manoj Kumar Chinnakotla, Sagar Ranadive, Om P. Damani, and Pushpak Bhattacharyya, "Hindi to English and Marathi to English Cross Language Information Retrieval Evaluation"; Springer-Verlag Berlin Heidelberg 2008.

[15] Satoshi Nakamura, Konstantin Markov and Hirofumi Yamamoto, "The ATR Multilingual Speech to Speech Translation System", IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 2, 2006.