iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** III **Month of publication:** March 2022

DOI: https://doi.org/10.22214/ijraset.2022.40721

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue III Mar 2022- Available at www.ijraset.com

Speech to Speech Translation with Response Suggestion (English-Hindi)

Kushal HU[1], Kethu Yashaswi[2], Patil Chanchal Vinod[3], Meghana G[4]
[1, 2, 3, 4]*Vidyavardhaka College Of Engineering*

Abstract: Language barriers are becoming more common in primary care as a result of the surge in international migration. People who do not speak the same language will be able to converse in real-time using natural spoken language thanks to this technology. It can assist in overcoming linguistic hurdles as well as having more effective communications. We can also reply rapidly as a result of it. Medical researchers must translate their findings into several languages, with international tribunals and regional blocs such as the EU requiring particularly stringent translations.

I. INTRODUCTION

India is one of the countries with a wide range of cultural and linguistic diversity. More than 90% of the 599 people who took part in this nationwide cross-sectional survey suffer language challenges at least once a year, with 30.0 percent experiencing them once a week. Using family and friends for translations is the most common method of bridging a language barrier (60.1 percent say it happens in more than half of their encounters), followed by "using gestures" (32.0 percent) and simply tolerating the lack of communication (22.9 percent). Around 15% of the world's population suffers from a language-based learning problem. As a result, there is a communication breakdown, resulting in a significant gap in discourse. To get around this, we'll need a translator in the middle who will act as a link and translate it into the appropriate language.

It is of great assistance to migrants and their families. Illiterates and many people with only basic reading/writing skills will benefit greatly from it. Individuals' learning processes will be enhanced, and their communication flexibility will be improved. Emails and chats have become a big part of our life in recent years, thanks to the introduction of smart devices. Users of laptops and cell phones receive hundreds of messages every day, not only at work but also in their personal lives.

The majority of these responses are domain-specific or require context, but a significant percentage is general and consists of the standard set of words like "Okay," "Wow," or "Good morning," among many others. As a result, it's only natural that our devices have some additional capability to assist us in replying to received messages in such day-to-day circumstances with the addition of quick responses to speech; this saves time and enhances the dialogue in their respective languages.

II. PROBLEM STATEMENT

Language obstacles are projected to become more prevalent in primary care as international migration increases. More than 90% of the 599 people who took part in this national cross-sectional study suffer language challenges at least once a year, and 30.0 percent confront them every week. Using family and friends for translations is the most common method of bridging a language barrier (60.1 percent say it happens in more than half of their encounters), followed by "using gestures" (32.0 percent) and simply tolerating the lack of communication (22.9 percent).

III. LITERATURE SURVEY

Felix Stahlberg et al.[1] They used Modern NMT architectures in this paper, which can be traced back to their origins in word and sentence Embedding. They went over the Encoder and Decoder Families. They employed the Encoder and Decoder Network Concept with Fixed Length Sentence Embedding, but they used Attentional Encoder and Decoder Networks because this provided poor translations for extended sentences. RNN search, GNMT, ConvS2S, and the Transformers are among the most commonly utilised designs discussed in this work. It also delves into advanced NMT topics like explainability and NMT-SMT hybrid systems.

Rico Sennrich et al.[2] They introduced an innovative and effective technique in this study for making the NMT model capable of open-vocabulary translation by encoding rare and unknown words as sequences of sub-word units. To improve the encoding of any rare words in the vocabulary with appropriate sub word tokens while avoiding the insertion of any unknown tokens, simple character n-gram models and segmentation based on the Byte pair encoding compression technique [BPE] were utilised. Thanks to this tokenization process, every word may finally be able to overcome the fear of being forgotten.

©IJRASET: All Rights are Reserved

Mikel Artetxe et al.[3] In this research, they propose a method method for training an NMT system utilising monolingual corpora in an unsupervised manner. They employed an unsupervised embedding mappings approach as well as a slightly modified attentional encoder-decoder model that can be trained on monolingual corpora utilising a combination of demonising and Back Translation on monolingual corpora. This methodology achieved a score of 10.21 BLEU points in the WMT 2014 French to English and German to English translation. This study also explains how to incorporate character-level data into the model, which improves the methodology for dealing with some of the model's limitations.

Anjuli KannanF et al.[4] The purpose of this paper is to propose and demonstrate a new method for semantic clustering of user-generated data that only requires a modest amount of tagged data. It employs LSTM, Deep Learning, Clustering, and Semantics ideas. Language Detection, Tokenization, Sentence Segmentation, Normalization, Quotation Removal, and Salutation/Close Removal were all used to pre-process the data in this paper. Perplexity, Mean Reciprocal Rank, and Accuracy is three standard metrics used to evaluate the LSTM model.

Rico Sennrich et al.[5] Neural machine translation is the art of performing several languages to match with the translation models. On the destination side, the translation data plays an important role in this model. They are providing the source data and were successfully reached to some extent, but obtain the results for all tasks in this survey. The results are back-translated from the target data into the source data. This research will improve fluency and effectiveness in the target language using neural machine translation models.

Sahinur Rahman Laskar et al.[6] Neural machine translation achieves the piece of work in machine translation because of the use of advanced deep learning techniques.

NMT model still suffers from the low quality of source data, to Overcome this challenge multi-modal came into existence. Multi-modal techniques came into existence for textual or visual features to improve the quality of the source data. Multi-modal NMT carries out for higher BLEU, RIBES scores than text NMT models for translations.

Ajay Anand Verma et al.[7] In this paper, the encoder and the decoder suffers the problems of poor performance with long sentences with the out-of-vocabulary To solve the out-of-vocabulary problem introduces NMT along with subword segmentation called subword NMT. The long sentence translation error is identifying the attention-based NMT model using LSTM units. The out-of-vocabulary words translation problem is solved by the sub-word segmentation by the byte-pair encoding(BPE).

Saikiran Gogineni et al.[8] When the length of the source language exceeds the restrictions, the performance of the encoder and decoder models has no bearing on the translation model. The 4 layer bi-directional LSTM approaches are ideal for sentences that are longer than the minimal length, resulting in higher BLEU scores for the network. They're employed in encoders and decoders, attention mechanisms, recurrent neural networks, and BLEU scoring techniques, among other things. Rule-based translation strategies are used in corpus-based translation and hybrid-oriented machine translation in this paper.

Roee Aharoni et al.[9] This paper portrays the work on NMT models which can translate up to 102 languages to/from English with a certain amount of trained directions. They use the TED talks dataset which consists of 58 languages paired with English along with other in-house datasets. Zero-shot accuracy is used as a degree of measurement for model generalization. Hence, they realize that by adding more languages, the model will be forced to create a generalized representation to utilize its capacity, thereby improving the zero-shot performance.

M. Anand Kumar et al.[10] This paper explains the machine translation shared task for language pairs of English to Tamil, Malayalam, Hindi and Punjabi. The entire shared task is the first to be evaluated by humans. Adequacy scores, Fluency scores, Overall ratings of the shared task MTIL-2017 were a part of the analysis results. The shared task discussed in this paper led to discoveries of new methods for machine translation for Indian languages. The study will be further evaluated for translation to other Indian languages from English apart from the 4 languages mentioned and vice versa.

Hendra Setiawan et al.[11] A Robust variational NMT model that performs much better than the best non-latent NMT models has been discussed in this paper. They analyse how different word dropout rates affect the performance of the model. Their experiment also included finding the optimal Latent Dimension and Normalizing flow configuration for certain conditions. Further improvements include applying the NMT model to many other inter-language conversions.

Georgiana Dinu et al.[12] The paper describes a way for adding new terminologies to neural network translation during runtime that can meet production requirements. For integrating terminology limitations, they do not employ constrained decoding algorithms. All studies were done in a zero-shot environment and used a black-box technique in which a generic neural Machine Translator is directly trained to understand how to use an external terminology that is provided at run-time. All experiments were done in a zero-shot setting.

Abigail See et al.[13] This paper explains how pruning schemes like class-blind, class uniform and class distribution help compress NMT models better. The WMT'14 English-German translation task has been worked upon for experimentation. After pruning and retraining the model, they returned with up to 62.5 percent reduction in storage size. These authors are the first ones to be applying Compression techniques to NMT models. They also tested the generalizability of the observations obtained by testing the pruning methods over smaller datasets.

Budhaditya Deb et al.[14] To help diversity response ideas, this research employed a generative latent variable model using Conditional Variational Auto-Encoder (M-CVAE). The main goal of this study is to increase the production click rate by diversifying the responses. The model has two encoders that project messages and responses into a common feature representation. Responses are chosen from a collection of choices during the inference. For diversification, the model employs lexical segmentation and optimization of marginal significance.

Mozhi Zhang et al.[15] This paper explains how they create a dataset for reply suggestions in 10 languages. They downloaded comments from reedit to create a data set. The top 10 languages are chosen in which they consist of at least 100k examples. There are 2 types of models to generate a reply one is the retrieval model which selects the reply from the fixed set of responses and the other is a generative model which generates replies from scratch. This feature is available in many applications including Gmail, LinkedIn, Microsoft Teams, Uber, etc.

IV. METHODOLOGY

In this System, the voice of the speaker will be recognized by the built-in Python Speech Recognizer Model and will be converted into text format. There were will be the following Sections in the System.

1) Neural Machine Translation Model
2) Automatic Reply Model for the Text
3) The Resulted Text will be spoken out by the Python Module.

A. System Description

Figure 1.1 depicts the proposed system's methodology or system flow diagram. The user will first pronounce his sentence, which will then be translated by the algorithm. The speech module in Python will identify the voice and transform it into the embeddings text format. With the help of Word Embeddings, the information from the text will be extracted and passed on to the Encoder-Decoder Model with Attenuation Mechanism, where the text will be translated into Hindi. The same text will be transmitted to the Automatic Speech Reply Model, which will generate a reply to the text, which will then be forwarded to the translation model for translation. Finally, translated input text and translated reply text will be obtained, and these messages will be supplied to the Python voice module to speak aloud depending on the user's needs.

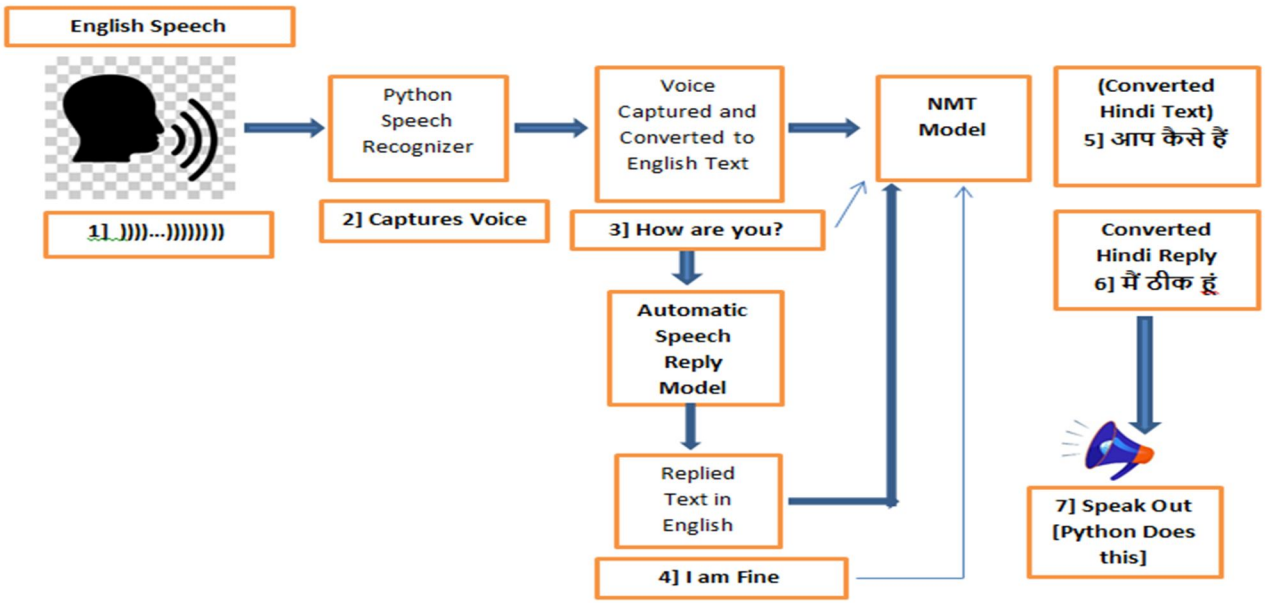


Figure 1.1 System Flow Diagram

V. CONCLUSION

The Papers explains about various models and techniques in Translation which are taken into consideration for the development of speech to speech translation using Neural Machine Translation. They have explained various pre-processing techniques that can be executed for the sentences before sending as input to the model. The various pre-processing techniques include Tokenization, Sentence Segmentation, Normalization, word embedding's and BPE. They also have explained about various models which can be effectively used to translate sentences. The model is evaluated based on the Perplexity, Mean Reciprocal Rank, Precision, Adequacy scores, Fluency scores, Overall ratings and Bleu Scores. It explains different models and techniques to develop the data for the automatic reply suggestions and diversify the response. The dataset can be either a retrieval model from a fixed set of replies or a generative model where the reply suggestions are generated from the scratch.

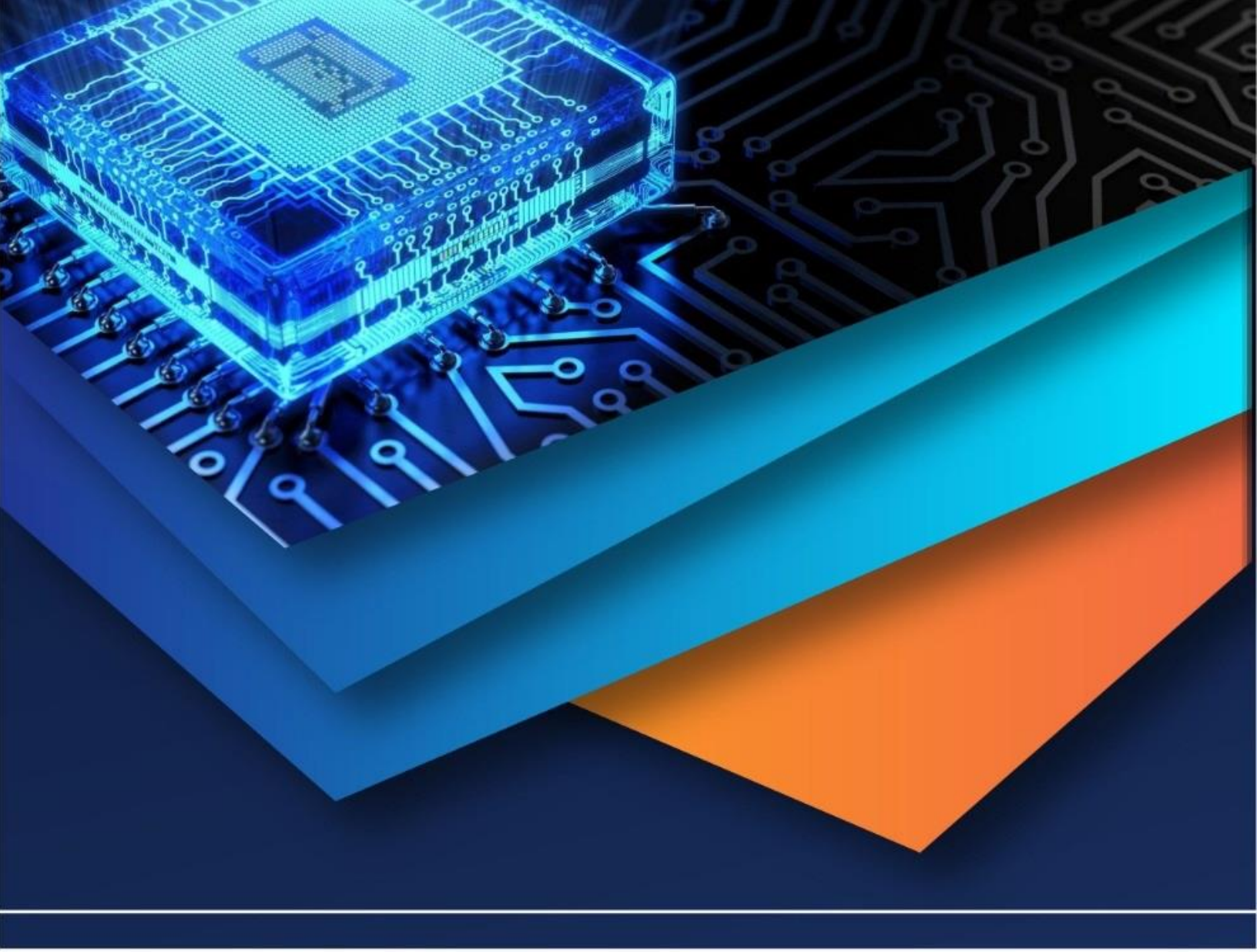
VI. FUTURE SCOPE

The proposed system can be extended for all languages, with more data the accuracy can be improved. More Sophisticated Models can be developed with the help of transformers and Transfer Learning Models. The replythe Suggestion Model can be improved with the help of creating similar sentences in one cluster for each sentence. The reply the suggestion model can be improved with the help of more replies for each text.

REFERENCES

[1] Felix Stahlbergu University of Cambridge, Engineering Department, Trumpington StreetCambridge CB2 1PZ, United Kingdom. Neural Machine Translation.
[2] Rico Sennrich and Barry Haddow and Alexandra Birch School of Informatics, University of Edinburgh. Neural Machine Translation of Rare Words with Subword Units.
[3] Mikel Artetxe, Gorka Labaka & Eneko Agirre IXA NLP Group University of the Basque Country (UPV/EHU), Kyunghyun Cho New York University CIFAR Azrieli Global Scholar. Unsupervised Neural Machine Translation.
[4] Anjuli KannanF, Karol KurachF, Sujith RaviF, Greg Corrado, László Lukács, Marina Ganea, Tobias KaufmannF, Andrew Tomkins, Balint Miklos, Peter Young, Vivek Ramavajjala. Smart Reply: Automated Response Suggestion for Email.
[5] Rico Sennrich and Barry Haddow and Alexandra Birch School of Informatics, University of Edinburgh. Improving Neural Machine Translation Models with Monolingual Data.
[6] Sahinur Rahman Laskar, Abdullah Faiz Ur Rahman Khilji, Partha Pakray, Sivaji Bandyopadhyay Department of Computer Science and Engineering, National Institute of Technology Silchar, Assam, India. Multimodal Neural Machine Translation for English to Hindi.
[7] Ajay Anand Verma, Pushpak Bhattacharyya CFILT, Indian Institute of Technology Bombay, India. Neural Machine Translation.
[8] Saikiran Gogineni, G. Suryanarayana, Sravan Kumar Surendran, CMR College OF Engineering & Technology, Hyderabad. An Effective Neural Machine Translation for : English to Hindi Language.
[9] Roee Aharoni - Bar Ilan University Ramat-Gan Israel Melvin Johnson and Orhan Firat - Google AI Mountain View California. Massively Multilingual Neural Machine Translation.
[10] M. Anand Kumar, B. Premjith, Shivkaran Singh, S. Rajendran and K. P. Soman From the Journal of Intelligent Systems. An Overview of the Shared Task on Machine Translation in Indian Languages (MTIL) – 2017.
[11] Hendra Setiawan, Matthias Sperber, Udhay Nallasamy ,Matthias Paulik. Variational Neural Machine Translation with Normalizing Flows.
[12] Georgiana Dinu, Prashant Mathur, Marcello Federico, Yaser Al-Onaizan - Amazon AI. Training Neural Machine Translation To Apply Terminology Constraints.
[13] Abigail See, Minh-Thang Luong, Christopher D. Manning - Computer Science Department, Stanford University, Stanford, CA 94305. Compression of Neural Machine Translation Models via Pruning.
[14] Budhaditya Deb Peter Bailey Milad Shokouhi Microsoft Search, Assistance and Intelligence. Diversifying Reply Suggestions using a Matching-Conditional Variational Autoencoder.
[15] Mozhi Zhang Wei Wang Budhaditya Deb Guoqing Zheng Milad Shokouhi Ahmed Hassan Awadallah. A Dataset and Baselines for Multilingual Reply Suggestion.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)