

# Meaningful Translation and Transliteration for Marathi Language

Mahendra Prabhakar Shinde

Lecturer, DVK MIT World Peace University, Pune, Maharashtra, India

## ABSTRACT

### Article Info

Volume 7, Issue 5

Page Number: 111-114

### Publication Issue :

September-October-2021

### Article History

Accepted : 07 Oct 2021

Published : 17 Oct 2021

Meaningful translation and transliteration is NP problem in case of languages like Marathi language as there are so many word disambiguation and multiple use and meaning of single word in different context is available. That is why identifying correct informational need and translating text into meaningful information is a tedious and error prone task. Google translate works on machine neuron network and WorldNet is an online reference system works on psycholinguistic theory of human memory. Both approaches are promising tools for language translation. Complete translation of Marathi text to English or English to Marathi also having problem of more complicated meaningless or tedious translation. Proposed algorithm is taking into consideration meaningful translation or transliteration as per user's informational need. This novel approach consider machine neuron network for meaningful formation of translated sentence and morphological structure for correct translation of word based on ontological analysis of word.

**Keywords:** Natural Language processing, machine learning, Indic languages, information analysis.

## I. INTRODUCTION

Many colleges from Maharashtra now a days offering syllabus in Marathi language. Also many manuscripts from Marathi is being translated into different languages. Machine translation is a buzz word now a days. People are using WorldNet or Google translate to translate information into regional languages. While using these systems they encounter so many problems mostly because of Marathi language and its forms, dialects, , meaning of same word is different in different context. For example keyword “लावणे” . Translation for keywords “ To close the door” is दरवाजा लावणे and translation for keywords दिवा लावणे is to lighten the lamp. In both sentence “लावणे” represents complete different meaning. Also due to cosmopolitan culture

people are more familiar with original English keyword for example “भ्रमणध्वनी” is more complicated than mobile.

Google's translation feature is depends on neural machine translation (NMT) [4]s system and it is more promising way of translation as it offer zero-shot translation . Prior to this Google were using phrase based translation the key algorithm behind this service. Since then, rapid advances in machine intelligence have improved our speech recognition and image recognition capabilities, but improving machine translation remains a challenging goal.

Marathi wordnet is based on idea of English WordNet. It is more than a conventional Marathi dictionary. It gives different relations between synsets or synonym

sets which signifies exclusive concepts. This is another important tool for Marathi translation. It is an online lexical reference system and mainly on psycholinguistic theory of human lexical memories.

Other approaches such as rule based corpus based knowledge based are also available . In this paper I am focusing of GNMT ,IITB's wordnet[3] , and rule based translation proposed by Mamulkar et all [9] and usable transliteration which is not covered in any of the approach.

## II. LITERATURE SURVEY

### A. Machine Translation Approaches

Machine Translation method is built on dictionary records, which means that the translation is done word to word, typically without much connection of meaning between them. Dictionary search may be done with or without structural analysis or lemmatization. This approach to Machine Translation is probably the least refined but it is ideally suitable for the translation of long lists of sentences on the sub-sentential (not a full sentence) level. Examples catalogs of products and inventories data. It can also be used to advance manual translation, if the well trained person who is fluent in both languages and therefore capable of amending syntax and grammar

**Neural Machine Translation (NMT)** is an end-to-end erudition approach for mechanical translation, which is having potential to overwhelm many of the weaknesses of conventional phrase-based translation systems. Unfortunately, NMT systems are known to be computationally costly both in training and in translation implication. In the case of very large data sets and large models. Several authors have also charged that NMT systems lack robustness, mainly when input sentences comprise infrequent words. These issues have stuck NMT's use in practical deployments and services, where both correctness and speed are important. For example most of the colleges are now a days offering engineering courses in native languages. In that case accurate translate and transliteration is essential.

### B. ANUBHARTI

Prof R. M. K. Sinha,[10][11] et al from IITK developed a Machine Translation system by Generalized Example-Base and Raw Example-Base (REB) Machine Translation system approach for hybridization. The combination of sample-based method and traditional grammar rule based approach is used in Anubharati-II. The example based approach imitates human learning method for extracting knowledge from past experiences or recognition . as above mention lavane word meaning can easily understood by Marathi native persons . and to be used in future. The source language is Hindi. The entered Hindi text is converted into a standard form to handle the word order. The Hindi sentences converted into standard form are matched with a top level standard form of example-base. If no match is found then a shallow chunker is used to portion the input sentence or combination of string into small units and then they are matched with a hierarchical example-base. The small chunks obtained by shallow chunker are translated and positioned by matching with sentence level example base.

### C. IndoWordnet

In countries like India machine translation and cross lingual search are common searches. These problems need large Data corpus - like wordnets and lexicons- that of high quality and coverage. Wordnets are lexical structures poised of synsets and semantic relations. Synsets are sets of synonyms. Wordnets are linked by semantic relations like hypernymy is-a kind of relationship , meronymy (part-of ), troponymy (manner-of). IndoWordnet is a related erection of wordnets of major Indian languages from Indo-Aryan, Dravidian and Sino-Tibetan families. For example synset वनस्पती is hypernymy of झाड़ and आंबा is hyponymy

## III. PROPOSED SYSTEM

Proposed system aims at meaningful translation and transliteration of Marathi language sentences . which comprises of neural machine translation and synsets for meaningful classification of words .

चक्र + पाणी means wheel + water and english word for this is whirlpool but चक्र + पाणि or चक्रपाणि means one who is holding wheel in his hand , shrikrishna . in this पाणि means hand not water. Neural machine translation translates चक्रपाणि into whirlpool which depicts wrong translation.

Whenever I am translating universal serial bus into Marathi NMT gives me result as सार्वत्रिक सिरीयल बस which is also wrong. here we need transliteration and should be transliterate into युनिव्हर्सल सीरिअल बस as it in Marathi font.

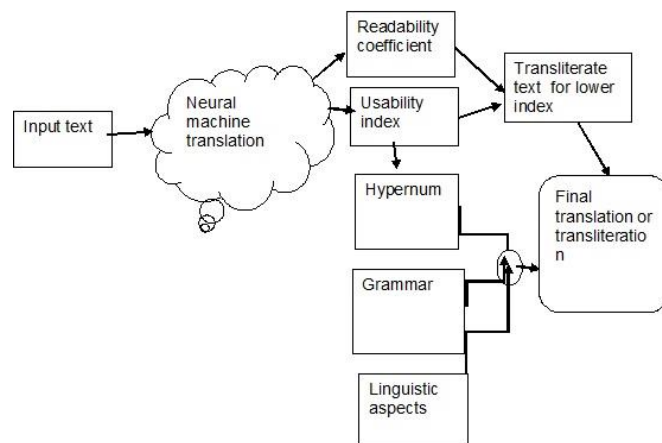
Wordnet and anubandh provides translation into Marathi with grammatical correct sentence with Marathi dictionary backup. but whenever I am going to translate मी मुलाखत देतोय , देतोय मी मुलाखत , मुलाखत देतोय मी all three sentence are grammatically correct . it gives me differed sentences with approach they are using.

Our proposed algorithm uses Flesch reading ease to calculate readability of translated word

$206.835 - (1.015 * \text{number of words} / \text{number of sentences}) - (84.6 \text{ number of syllables the number of words})$  this will be taken into consideration for each translated large word.

After that we are going to take following parameter for word and sentence formation Sense. Sense indicates that usability of translated keyword if it doesn't make any sense the keep it as it is.

- Repeated word
- Style of writing
- Discourse genre through synets
- Discourse structure of sentences.
- Usability index based on naïve byes theorem



#### IV. CONCLUSION

Translation of text into another language is a NP problem. There are numerous approaches available for translation. Neural machine translation is a promising solution available for languages like Marathi where multiple meaning of text available and its meaning only can be detected after reading entire text.

In this paper I am aiming to solve problem meaningful translation and if translation is not useful then transliteration only can make system usable by considering usability index on considering know probability of popular words , readability coefficient , aspect value and grammar for syntactic .

#### V. REFERENCES

- [1]. G.V.Gajre, G.Kharate, H. Kulkarni (2014), "Transmuter: An approach to Rule-based English to Marathi Machine Translation" , International Journal of Computer Applications (0975 – 8887) Volume 98 – No.21, July 2014
- [2]. R. Mahesh K. Sinha, A Journey from Indian Scripts Processing to Indian Language Processing, IEEE Annals of the History of Computing, Jan-March 2009. pp. 2-25.
- [3]. <https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html>
- [4]. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B.,

- Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X. Tensorflow: A system for large-scale machine learning. Tech. rep., Google Brain, 2016. arXiv preprint.
- [5]. Brown, P., Cocke, J., Pietra, S. D., Pietra, V. D., Jelinek, F., Mercer, R., and Roossin, P. A statistical approach to language translation. In Proceedings of the 12th Conference on Computational Linguistics - Volume 1 (Stroudsburg, PA, USA, 1988), COLING '88, Association for Computational Linguistics, pp. 71-76
- [6]. Chung, J., Cho, K., and Bengio, Y. A character-level decoder without explicit segmentation for neural machine translation. arXiv preprint arXiv:1603.06147 (2016)
- [7]. Gers, F. A., Schmidhuber, J., and Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural computation* 12, 10 (2000), 2451-2471. 18 Gülçehre, Ç., Ahn, S., Nallapati, R., Zhou, B., and Bengio, Y. Pointing the unknown words. CoRR abs/1603.08148 (2016). 19 Gupta, S., Agrawal, A., Gopalakrishnan, K., and Narayanan, P. Deep learning with limited numerical precision. CoRR abs/1502.02551 (2015).
- [8]. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi yonghui,schuster,zhifengc,qvl,mnorouzi@google.com
- [9]. IndoWordnet, Pushpak Bhattacharyya Bhattacharyya P., Fellbaum C. and Vossen P. (eds.) (2010), Principles, Construction and Application of Multilingual Wordnets, Proceedings of the 5th Global Wordnet Conference, Mumbai, Narosa Publishing House, India.
- [10]. Supervised, Semi-Supervised and Unsupervised WSD Approaches: An Overview Lokesh Nandanwar, Kalyani Mamulkar International Journal of Science and Research (IJSR)Open Access | Fully Refereed | Peer Reviewed ISSN: 2319-7064
- [11]. R. Mahesh K. Sinha, A Robust POS Tagger Using Multiple Taggers and Lexical Knowledge, ICAI'10 - The 2010 International Conference on Artificial Intelligence, Las Vegas, Nevada, USA, July 12-15, 2010
- [12]. R. M. K. Sinha and Anil Thakur, Syntax and Semantics of 'kaa' in Hindi, Proceedings of International Symposium on Machine Translation, NLP and Translation Support System (iSTRANS- 2004), November 17-19, 2004, Tata Mc Graw Hill, New Delhi, pp: 226-229.

**Cite this article as :**

Mahendra Prabhakar Shinde , "Meaningful Translation and Transliteration for Marathi Language ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 7 Issue 5, pp. 111-114, September-October 2021. Available at doi : <https://doi.org/10.32628/CSEIT217534> Journal URL : <https://ijsrcseit.com/CSEIT217534>