

A REVIEW OF MACHINE TRANSLATION METHODS

BINGUMALLA SUDHEESH , BUDDAREDDYGARI NITHEESHKUMAR REDDY , AKULA LAKSHMI VENKATA PHANEENDRA

Abstract:

An effort to access information in other languages refers to the growth of a language conversion system with a wide range of properties and methods. The advantage of machine translation is frequently utilized by librarians and an information to fulfil the requirements of users. Machine translation methods differ, and each has its own unique strengths and weaknesses. There is no translation tool will provide an exact copy of an original language, but it will provide a summary of content which can be used to understand with a sort of information is present in the source text. After producing translation results with a transcription system, sometimes it is necessary After being edited by a human translator, a text that was built heritage by an MT engine reaches its final form. In this article we are going to discuss the different methods used in language conversion, including dictionary-based, rule-based, corpus-based, and hybrid methods. Despite the fact that the Hybrid approach is better than all other ways since it incorporates the qualities of several translation methods, the study concludes that no optimal translation systems exist.

Keywords:

machine translation, corpus-based, rule based, hybrid machine translation, computational linguistics

1.Introduction

The concept of translation software is now being created to address language variety problems. People will still not be able to understand and grasp all of the languages of the world. There must be approximately 7000 languages in the world, highlighting the importance of speech(tongue) relocation developments and methodologies. Language translation researchers are looking into to the possibilities of information getting transmitted from one speech to the other. Federal agencies and research organizations are supporting projects to improve computer textual translation systems, that could be helpful for multinational corporate communications and assisting professionals better to serve their customers. MT is a subfield of language modelling that examines the use of software applications to convert words either phrase from one original text to another target culture. Machine translation system has recently gained translation quality. The very first generation of automated translation software was dictionary-based machine translation, that relies entirely on electronic dictionaries. It only changes different words, not full conversations. The following step would have been to design Rule Based Machine Translation (RBMT) systems, Corpus Based Systems, and Hybrid Machine Translation systems. In terms of grammatical and semantic rules, RBMT generates translates from multilingual textual data based on structural, language producing, and contextual information about the source and target languages. Corpus-based systems, on the other hand, generate translates from multilingual textual data in real - time. The hybrid procedure is a technical methodology for producing a good overall translation software by combining the benefits of various methods.

2.WHERE ARE MACHINE TRANSLATIONS USED?

Translation systems that allow people to communicate with each the others from various places, allowing them to realize the benefits of information and communications technologies [13]. Machine conversion is widely applicable in numerous of actions, and a lot of conversion systems, especially, corporate ones are supporting the use of methods [6]. Main benefits of machine conversion will conduct research via analysing internet services and materials. Moreover, conversion software technology will benefit on business, official plans, customer support, manuals and

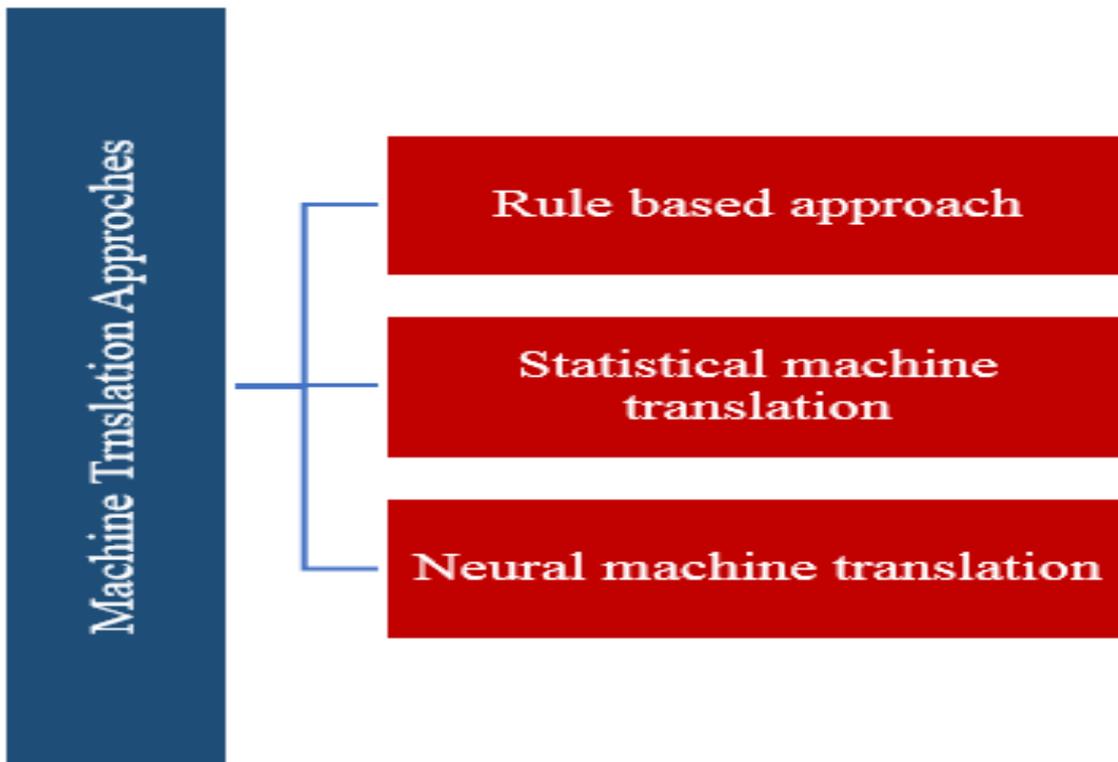
documents translation, application translation, email translation on service requests, customer support, personal communication such as ticket bookings, financial services abroad, and so on.

3. ATTEMPTS TO MEASURE FOR CHOOSING MACHINE TOOLS FOR TRANSLATION

As stated in [2] the speed and accuracy are the are two key aspects to consider when analysing the performance of MT technologies. Furthermore, as stated in [3] the factors for evaluations involve linguistic quality and simplicity of connection with available technology. Linguistic performance means that post-editing of translation output requires less time, & flexibility of connectivity allows for better communication with translation control systems via API. It is important to find the time spent and trace the edit and reviewing distance in order to calculate the language quality. Edit distance is the amount of editing necessary to bring a machine translation output up to publishing requirements. The rewrite of a vendor's translation along with in linguists is known as review distance. [4] WER is used to evaluate the security level because it is more understandable than BLEU, which is used in [3].

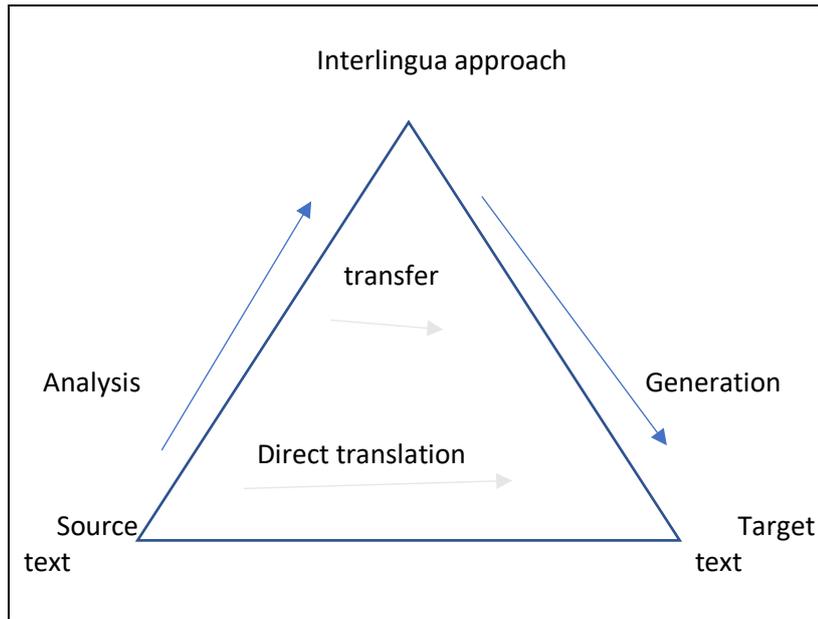
4.MACHINE TRANSLATION APPROACHES

For natural languages such as English, Sanskrit, Hindi, Frenchie, Spanish and others, many machine translation systems have already been created.



5. RULE BASED MACHINE TRANSLATION (RBMT)

Knowledge Based Machine Translation (KBMT) is a technique for retrieving rules from lexicons and grammatical rules based on semantic data about the source and the target. RBMT produces goal samples based on every language's structural, grammatical, and lexical symmetries. It translates input to targeted grammatical structure and is expandable and sustainable, as described in [1]. RBMT systems are divided into three categories.



i) Using the direct method (Dictionary Based Machine Translation)

Machine Translation Direct One of most basic type of machine translation is this. Direct translation software breaks the input sentence into words, checks these to a given dictionary, and then adjusts the output based on grammar rules using a simple rule structure. An example of a system that takes a direct approach is Anusaarka. It is designed by IITHyderabad.

ii) RBMT Systems Transfer

In Transfer based systems, the essential methods are grammatical and syntax analytics. With the use of grammatical rules and multilingual dictionary, input text is converted into something less specific language form, as well as the same level of understanding is produced

iii) RBMT Systems that are multilingual

This approach is supposed to achieve global grammatical uniformity. The original language is converted into middleman expression which is not depending on any other languages in this technique. This additional form of

Language	Edit Distance	Review Distance	Volume (words)
French	46.33%	9.1%	38900
Italian	49.05%	16.94%	40149
Spanish	33.67%	6.30%	56269
Simplified Chinese	54.43%	2.69%	80367

expression is often used to produce new language [13]. The key feature of this approach is that it has a single representation for various languages, make multinational translation software significantly easier.

The rule-based machine translation system's edit and review distances are listed below.

The Advantages of RBMT:

There are no multilingual texts required. This facilitates the generation of speech processing for language with no common text and no digitised data at all.

There is no ceiling of good quality. Even if the trigger situation is highly unusual, every problem may be addressed with a targeted rule. In statistics systems, on the other side, rare types are systematically completely erased.

Independent of the area. Examples are typically written in a top-level domain manner, this implies that large number of policies would "just work" in any area, only with a few specific cases per scope needing rule development.

The Disadvantages of RBMT:

As mentioned in [9], RBMT has various drawbacks, the first most of which was the shortage of quality dictionary. The building of new dictionary is a extremely enormous investment. Another drawback is that certain grammatical data must be physically provided. In contrast, handling rules interaction and uncertainty in a big system is extremely hard. RBMT allows for such generation of different rules and the addition of new rules, but that these updates are really expensive.

6. CORPUS BASED MACHINE TRANSLATION

Corpus Based Machine Translation is among the most popular ways of translation software just because it provides a good accuracy during in the translator. Following the development of a content method that is used in various computer-aided translating tools [8] huge amounts of translation are presented. The corpus-based model was trained using multilingual datasets which have been POS-tagged and analysed, and also phrase and set of words, and develops translating rules between both the input context and output [14]. A various type on Corpus Based Machine Translation models is listed below.

i) Statistical Machine Translation (SMT)

This method uses statistical models to create converted results with the use of multilingual texts. Scientific method developed the concept of Statistical Machine Translation. Language learners should not have to do any customization effort with this method because the software learn translated methods via statistical analysis of multilingual texts [11]. This technique is much less costly as Rule-based technologies. Another advantage of a SMT approach would be that it promotes good utilization of resources, giving in much more organic transactions. An example of the an SMT technology is n-gram oriented SMT. Various kinds of statistical models included statistical word-based translation models, statistical phrase-based translation models, and statistical syntax-based translation models [1].

The limitations of SMT are:

-> With limited funds, building a corpus is expensive.

-> Need not perform well and there are major differences in grammatical structure among languages.

-> unable to predict the output

ii) Example Based Machine Translation

Its technique is also known as Memory based translation because it gets a collection of statements from of the original text and provides goal to translations in the native tongue. Whenever different types of statements are converted via samples, as well as in the past converted statement is reproduced, a repeat of same conversation is

likely to be accurate one more time [7]. The key benefit of this technology would be that it performs efficiently with such little volume of information and that may produce results faster rapidly if the translation engine is learned. As seen in [1], the example-based technique is basically worked to convert two completely different tongues, such as Spanish and franchise. One of the key drawbacks of the Example-based technology would be that deep speech research cannot be applied. EBMT tool PanEBMT [10] is an example.

Statistical Translation Software is a significant method that incorporates corpus-based translation software, and the table under shows the edit and review distance of this technique.

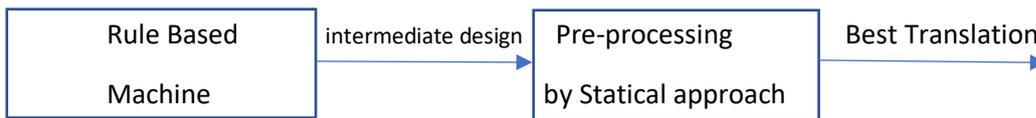
Language	Edit Distance	Review Distance	Volume (words)
Danish	36.18%	0.32%	89747
Norwegian	37.19%	N/A	105674
Swedish	43.56%	2.41%	115544

6. HYBRID MACHINE TRANSLATION (HMT)

RBMT and Statistical Language Processing are combined in HMT. It begins using RBMT and afterwards modifies the principles via statistical tools. In order to effectively guide the analytical machine, procedure is utilized to reconducting input. Hybrid approaches were different in a number of ways [13], that are

A) Rules post-processed by Statistics

For the initial translation, a rule-based model is utilized. The translated result of a rule-based technology is altered using a statistical model.



B) Statistics guided by rules

Rules are implemented to preconditioning inputs in this approach, that offers extra advice to the statistical method. Standards are often used to standardize the statistical result after it has been post conducted. At the time of conversion, this approach provides greater flexibility, strength, and management.

Hybrid Machine Translation Engine [11] is an example of DFKI-LT. The following table illustrates the edit and review distances for HMT.

Language	Edit distance	Review distance	Volume (words)
German	77.88%	13.77%	17269
Russian	69.11%	5.85%	33764

As these different languages have such rich vocabulary, the edit distance of the information in the above table is really quite high. Hybridization has happened recently for French and Italian, but still no results were presently available [3]. For both oral and written information, hybrid machine translation outperforms approaches [12]. Telugu

to English MT [13] use a hybrid technique, that shows many features, namely (a) fast business applications, (b) scalable machine translation, (c) high precision, and (d) a simple and easy framework. Furthermore, because Straight MT requires less humans post-editing, it is in high demand both interpreters and consumers. It generates copies that are affordable, rapid, and high-quality. It also has several restrictions, such as part-of-speech matching issues, unnecessary commas, and wrong spelling, as shown in [3].

7. Neural Machine Translation (NMT)

NMT differs other phrase-based statistical techniques in that it uses individual parts that are independently designed. [15] Neural machine translation (NMT) is not a significant departure from what's been achieved in statistical machine translation in the before (SMT). The use of vector representations ("autoencoder," "continuous spatial forms") for phrases and emotions is the fundamental departure. The designs have fewer components over phrase-based approaches. There's only the single series model used to predict 1 word, with no separate language, translations, or rearranging methods. Its pattern forecasting, on the other side, is based on the full source sentence and the full previously produced target sequence. DL and reinforcement learning are used in NMT models.

Initially, word sequence analysis was typically done with such a recurrent neural network (RNN). The neural network utilizes a multimodal rnn, defined as an encoding, that decode an input word for an another RNN, known as a decoder, that forecasts word in the mother tongue. [16] Large inputs are difficult for rnn to incorporate into a single feature vector. A attention mechanism [17] may solve for this by enabling the decoding to concentrate on various parts of the source during producing every letter of the result. Also, there are Covering Approaches which tackle issues with focus techniques like neglecting previous showed similar, that results to excessive and under-translation. [18]

8. DISCUSSION OF RESEARCH FINDINGS

Machine translation is a way of converting a given text to a mother tongue based on textual principles. Its ability to understand human language is critical to a successful for translation software. As previously noted, different approaches for machine translation are possible. The kind of machine translation method utilized is largely defined by the original and required language combinations. Once modification is done repeatedly, RBMT performed best and gives better results. However, it's much less efficient as compared to Corpus-based and Hybrid techniques. Whenever the required language does not have a much vocabulary, Corpus Based MT, especially Statistical MT, is a good option. Whenever the original and required languages are much more difficult, Hybrid MT is preferred because it combines the benefits of different methodologies.

9. CONCLUSION

Machine translation is indeed an advanced method for which information is translated from one human language to another via software program. In transcription, the interpreter must understand the content of the original text and develop the goal language sentence structure. This method necessitates a thorough understanding of the sender and receiver language syntax, word order, and concepts. Language conversion is being used in a wide range of applications today, including service quality, text transcription, messaging, program development, online transcription, and etc. A most common methodologies for translation software includes dictionary-based, rule-based, corpus-based, and hybrid techniques. As previously said, each of them has its own set of upsides and downsides. Any 2 transcription systems could yield similar copies of same content in the very same pair of words. For high-quality conversion, postediting is also needed.

REFERENCES

- [1] S. Tripathi, J. K. Sarjgek, "Approaches to machine translation", *Annals of Library and Information Studies*, Vol. 57, Dec 2010, pp. 388-393
- [2] L.R. Nair, D.S. Peter, P.R. Renjith, "Design and Development of a Malayalam to English Translator-A Transfer Based Approach", *IJCL*, Vol. 3, Issue. 1, 2012, pp.
- [3] C. Dove, O. Loskutova, and R. Fuente, "What's Your Pick: RbMT, SMT or Hybrid?", 2012, available at: <http://amta2012.amtaweb.org/AMTA2012Files/papers/Doveetal.pdf>
- [4] F.J. Och, "Minimum Error Rate Training in Statistical Machine Translation", *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, July 2003, pp. 160-167
- [5] M. Nagao, "A Framework Of A Mechanical Translation Between Japanese And English By Analogy Principle". In A. Elithorn and R. Banerji. *Artificial and Human Intelligence*. Elsevier Science Publishers, 1984
- [6] M. N. Al-Kabi, T. M. Hailat, E.M Al-Shawakfa, I. M Alsmadi, "Evaluating English to Arabic Machine Translation Using BLEU", *IJACSA*, Vol. 4, 2013, pp. 66-73
- [7] E. Sumita, H. Iida, "Experiments and Prospects of Example-based Machine Translation", available at: <http://acl.ldc.upenn.edu/P/P91/P91-1024.pdf>
- [8] M. Guidère, "Toward Corpus-Based Machine Translation for Standard Arabic", *Translation Journal*, vol. 6, no. 1, January 2002,
- [9] A.-L. Lagarda, V. Alabau, Casacuberta, R. Silva, E. Díaz-de-Liaño, "Statistical Post-Editing of a Rule-Based Machine Translation System". *Proceedings of NAACL HLT 2009*, pages 217–220,
- [10] R.D. Brown, "Example Based Machine Translation in Pangloss System", available at: <http://www.scism.lsbu.ac.uk/inmandw/ir/example-basedmachine-translation.pdf>
- [11] P. Kohen, "Statistical Machine Translation", Cambridge University Press, New York, 2010, pp-53
- [12] S. Nirenburg, H. Somers, Y. Wilks, "Readings in Machine Translation", Asco Typesetters, Hong Kong, pp.157, 233
- [13] T. V. Prasad, G.M. Muthukumar, "Telugu to English Translation using Direct Machine Translation Approach, *International Journal of Science and Engineering Investigations*, Vol. 2, Issue. 2, 2013, pp. 25-35
- [14] D. Dinh, N. L. Ngan, D. X. Quang, V. . Nam, "A Hybrid Approach to Word Order Transfer in the English-to-Vietnamese Machine Translation", available at: <http://www.amtaweb.org/summit/MTSummit/FinalPapers/58-Dien-final.pdf>
- https://en.wikipedia.org/wiki/Neural_machine_translation
- [15] Wołk, Krzysztof; Marasek, Krzysztof (2015). "Neural-based Machine Translation for Medical Text Domain. Based on European Medicines Agency Leaflet Texts". *Procedia Computer Science*. **64** (64): 9. [arXiv:1509.08644](https://arxiv.org/abs/1509.08644). [Bibcode:2015arXiv150908644W](https://doi.org/10.1016/j.procs.2015.08.456). [doi:10.1016/j.procs.2015.08.456](https://doi.org/10.1016/j.procs.2015.08.456). [S2CID 15218663](https://doi.org/10.1016/j.procs.2015.08.456).
- [16] Dzmitry Bahdanau; Cho Kyunghyun; Yoshua Bengio (2014). "Neural Machine Translation by Jointly Learning to Align and Translate". [arXiv:1409.0473](https://arxiv.org/abs/1409.0473) [cs.CL].
- [17] Bahdanau, Dzmitry; Cho, Kyunghyun; Bengio, Yoshua (2014-09-01). "Neural Machine Translation by Jointly Learning to Align and Translate". [arXiv:1409.0473](https://arxiv.org/abs/1409.0473) [cs.CL].