Kingdom of Saudi Arabia

Ministry of Education

Taibah University

College of Computer Science & Engineering

CS463 – Natural Language Processing

Semester II, 2020

Machine Translation

Submit By:

Basem Awwad Al-moziny
No. 3706683
(TA14)

❖ Index:

❖ Introduction:

Machine translation, sometimes referred to by the abbreviation MT (not to be confused with computer-aided translation, machine-aided human translation (MAHT) or interactive translation) is a sub-field of computational linguistics that investigates the use of software to translate text or speech from one language to another. On a basic level, MT performs simple substitution of words in one language for words in another, but that alone usually cannot produce a good translation of a text because recognition of whole phrases and their closest counterparts in the target language is needed. Solving this problem with corpus statistical, and neural techniques is a rapidly growing field that is leading to better translations, handling differences in linguistic typology, translation of idioms, and the isolation of anomalies.[1]

Current machine translation software often allows for customization by domain or profession (such as weather reports), improving output by limiting the scope of allowable substitutions. This technique is particularly effective in domains where formal or formulaic language is used. It follows that machine translation of government and legal documents more readily produces usable output than conversation or less standardized text.

Improved output quality can also be achieved by human intervention: for example, some systems are able to translate more accurately if the user has unambiguously identified which words in the text are proper names. With the assistance of these techniques, MT has proven useful as a tool to assist human translators and, in a very limited number of cases, can even produce output that can be used as is (e.g., weather reports).

The progress and potential of machine translation have been debated much through its history. Since the 1950s, a number of scholars have questioned the possibility of achieving fully automatic machine translation of high quality, first and most notably by Yehoshua Bar-Hillel. Some critics claim that there are in-principle obstacles to automating the translation process. [2]

❖ Background on Research on The Topic:

The human translation process may be described as:

1. Decoding the meaning of the source text; and
2. Re-encoding this meaning in the target language.

Behind this ostensibly simple procedure lies a complex cognitive operation. To decode the meaning of the source text in its entirety, the translator must interpret and analyse all the features of the text, a process that requires in-depth knowledge of the grammar, semantics, syntax, idioms, etc., of the source language, as well as the culture of its speakers. The translator needs the same in-depth knowledge to re-encode the meaning in the target language.

Therein lies the challenge in machine translation: how to program a computer that will "understand" a text as a person does, and that will "create" a new text in the target language that sounds as if it has been written by a person.

In its most general application, this is beyond current technology. Though it works much faster, no automated translation program or procedure, with no human participation, can produce output even close to the quality a human translator can produce. What it can do, however, is provide a general, though imperfect, approximation of the original text, getting the "gist" of it (a process called "gisting"). This is sufficient for many purposes, including making best use of the finite and expensive time of a human translator, reserved for those cases in which total accuracy is indispensable.

This problem may be approached in a number of ways, through the evolution of which accuracy has improved. [3]

Machine translation can use a method based on linguistic rules, which means that words will be translated in a linguistic way – the most suitable (orally speaking) words of the target language will replace the ones in the source language.

Generally, rule-based methods parse a text, usually creating an intermediary, symbolic representation, from which the text in the target language is generated. According to the nature of the intermediary representation, an approach is described as interlingual machine translation or transfer-based machine translation. These methods require extensive lexicons with morphological, syntactic, and semantic information, and large sets of rules. [4]

Given enough data, machine translation programs often work well enough for a native speaker of one language to get the approximate meaning of what is written by the other native speaker. The difficulty is getting enough data of the right kind to support the particular method. For example, the large multilingual corpus of data needed for statistical methods to work is not necessary for the grammar-based methods. But then, the grammar methods need a skilled linguist to carefully design the grammar that they use.

❖ Review of Previous Studies and Existing Systems:

While no system provides the holy grail of fully automatic high-quality machine translation of unrestricted text, many fully automated systems produce reasonable output. The quality of machine translation is substantially improved if the domain is restricted and controlled. [5]

Despite their inherent limitations, MT programs are used around the world. Probably the largest institutional user is the European Commission. The MOLTO project, for example, coordinated by the University of Gothenburg, received more than 2.375 million euros project support from the EU to create a reliable translation tool that covers a majority of the EU languages. The further development of MT systems comes at a time when budget cuts in human translation may increase the EU's dependency on reliable MT programs. The European Commission contributed 3.072 million euros (via its ISA programme) for the creation of MT@EC, a statistical machine translation program tailored to the administrative needs of the EU, to replace a previous rule-based machine translation system.[6]

The notable rise of social networking on the web in recent years has created yet another niche for the application of machine translation software – in utilities such as Facebook, or instant messaging clients such as Skype, GoogleTalk, MSN Messenger, etc. – allowing users speaking different languages to communicate with each other. Machine translation applications have also been released for most mobile devices, including mobile telephones, pocket PCs, PDAs, etc. Due to their portability, such instruments have come to be designated as mobile translation tools enabling mobile business networking between partners speaking different languages, or facilitating both foreign language learning and unaccompanied traveling to foreign countries without the need of the intermediation of a human translator.

Despite being labelled as an unworthy competitor to human translation in 1966 by the Automated Language Processing Advisory Committee put together by the United States government, the quality of machine translation has now been improved to such levels that its application in online collaboration and in the medical field are being investigated. The application of this technology in medical settings where human

translators are absent is another topic of research, but difficulties arise due to the importance of accurate translations in medical diagnoses.[7]

❖ Architectural Design of Systems:

Shown in the Figure (1): Bernard Vauquois' pyramid showing comparative depths of intermediary representation, interlingual machine translation at the peak, followed by transfer-based, then direct translation.

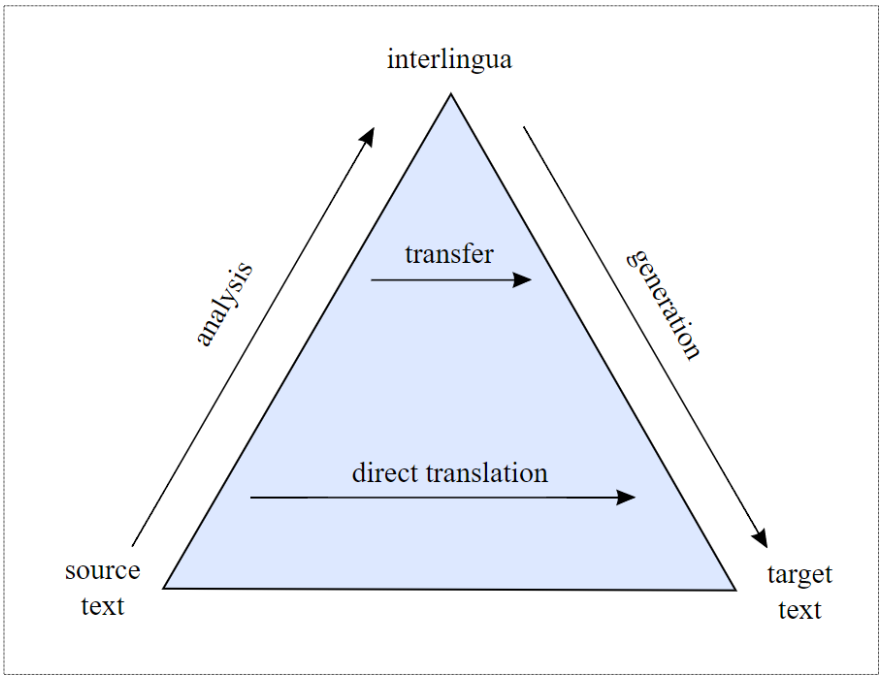

Figure (1)

❖ Possible Suggestions for Improving Existing Systems:

There are various means for evaluating the output quality of machine translation systems. The oldest is the use of human judges[58] to assess a translation's quality. Even though human evaluation is time-consuming, it is still the most reliable method to compare different systems such as rule-based and statistical systems.[59] Automated means of evaluation include BLEU, NIST, METEOR, and LEPOR.

Relying exclusively on unedited machine translation ignores the fact that communication in human language is context-embedded and that it takes a person to comprehend the context of the original text with a reasonable degree of probability. It is certainly true that even purely human-generated translations are prone to error. Therefore, to ensure that a machine-generated translation will be useful to a human being and that publishable-quality translation is achieved, such translations must be reviewed and edited by a human.[61] The late Claude Piron wrote that machine translation, at its best, automates the easier part of a translator's job; the harder and more time-consuming part usually involves doing extensive research to resolve ambiguities in the source text, which the grammatical and lexical exigencies of the target language require to be resolved. Such research is a necessary prelude to the pre-editing necessary in order to provide input for machine-translation software such that the output will not be meaningless.[8]

❖ Summary:

The Machine Translation has been a branch of Natural Language Processing, which comes under the broad area of Artificial Intelligence. Machine Translation system refers to computer software that translates text or voice from one natural language into another with or without human assistance. Worldwide, large number of machine translation systems have been developed using several approaches including human-assisted, rule-based, statistical, example-based, hybrid and agent-based techniques. Among others, Statistical machine translation approach is by far the most widely studied machine translation method in the field of machine translation. The multi-agent approach is a modern approach to handle complexity of the systems in past five years.

❖ References:

1. Albat, Thomas Fritz. "Systems and Methods for Automatically Estimating a Translation Time." US Patent 0185235, 19 July 2012.
2. Madsen, Mathias: "The Limits of Machine Translation" (2010).
3. Lehrberger, J., & Bourbeau, L. (1988). Machine Translation: Linguistic characteristics of MT systems and general methodology of evaluation (Vol. 15). John Benjamins Publishing.
4. Kouremenos, D., Fotinea, S. E., Efthimiou, E., & Ntalianis, K. (2010). A prototype Greek text to Greek Sign Language conversion system. Behaviour & Information Technology, 29(5), 467-481.
5. "Human quality machine translation solution by Ta with you" *(in Spanish). Tauyou.com. 15 April 2009*.
6. "Machine Translation Service". *5* August 2011.
7. Randhawa, G., Ferreyra, M., Ahmed, R., Ezzat, O., & Pottie, K. (2013). Using machine translation in clinical practice. Canadian Family Physician, 59(4), 382-383.
8. Wołk, K., & Marasek, K. (2015). Enhanced Bilingual Evaluation Understudy. arXiv preprint arXiv:1509.09088.