Module 10: Translation Today

Lecture 38: Machine Translation

The Lecture Contains:

Previous Next

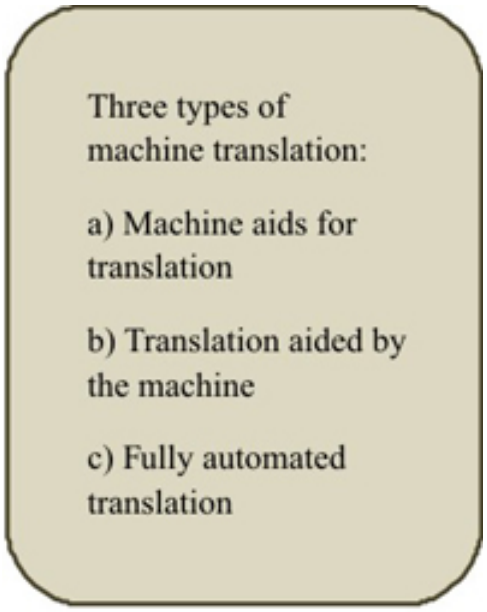
Introduction

If translation is not just about written languages it is also not about human beings alone. Recent advances in technology especially computer technology, have made it possible to have machines involved in the process. Translation, as we have seen, is perhaps as ancient as the activity of writing itself. However, the idea of getting this also done by machines is very recent. The fast pace of life necessitated the invention of machines that could take over a range of activities, one of which was also translation.

According to Harold L. Somers, the idea of mechanized translation goes back to the 17th century, but it can be precisely dated as 1933 ("Machine Translation", 140). This was the year when patents were given to two scientists – the Russian Petr Smirnov-Troyanskii and the Armenian French Georges Artsrouni to work on their ideas for a machine that could undertake translation. However, the eventual invention of mechanized translation was the offshoot of research related to defence matters after the World War II. The invention of the electronic computer after World War II, which was initially used to calculate ballistic firing tables and in code-breaking, helped a great deal in furthering this idea. The credit for pioneering machine translation is usually given to Warren Weaver of the US. He is called the founding father of Machine Translation (MT).

Previous Next

Role of Machines

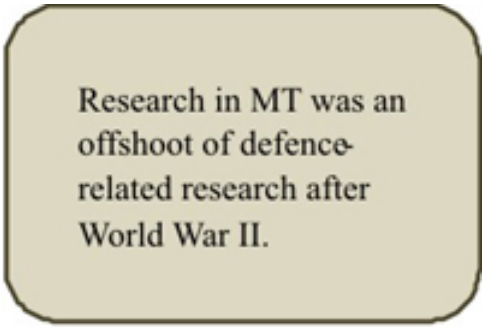


What is machine translation? Very simply put, it is the use of machines in translation. It is also called computer aided translation (CAT). Although there are no rigid boundaries, three types of machine translation have been identified generally: machine aids for translation, translation aided by the machine(also known as CAT), and machine translation. Machine translation is a blanket term used for all these activities. It is a field that is still now in flux and occupies an area that spreads over translation studies, linguistics and computer science. According to Minako O'Hagan, "Technology-based solutions to translation needs are a natural consequence of the shortened timeframe available for translation and increasing budgetary constraints resulting from GLOBALIZATION [sic], as well as progressive digitization of source content" ("Computer-aided Translation", 48).In other words, the fast pace of our lives today along with huge advances in the field of technology have made it possible or necessary, to depend on machines for translation.

As pointed out earlier, there are three types of MT. The machine aids for translators are literally what the term suggests, the help that can be provided by machines. These include word processors, dictionary management tools, and "various look-up facilities which support the translator but do not actually perform the translation task" ("Machine-aided Translation", 134).If you are a regular user of the computer, think of the number of times you might turn to dictionary.com or other internet sources to look for word meanings. Or the way in which you use your spellcheck on the computer. Today they are also called CAT tools which mean any form of computerized tool that can help in translation. They range from word processors to internet search engines, multilingual electronic dictionaries etc. In machine-aided translation the actual task of translation is done by systems but it depends on human input at various stages. Machine translation is different from this in that it is fully automated and does not have any human help in the translation activity. The final output might be passed on to a translator for post-editing. Human intervention is the last step in this process of machine translation.

Previous Next

History of Machine Translation



The idea of using computers to help in translation, as observed earlier, was an accidental by-product of war-time research into the possibility of code breaking, or deciphering the secret codes of the enemy. Although William Weaver was the first to submit a proposal for machine translation, it was Yehoshua BarHillel in the US who became the first researcher in MT in 1951. The desire to gain an upper hand in military matters motivated countries like the US to be liberal in funding such research. In 1954 came the first demonstration of MT when a Russian-English translation system was jointly developed by a team of researchers. Since this was based in Georgetown University, this came to be known as the Georgetown experiment. This initial success spurred the US to invest further in MT research. Other countries like Britain, Japan and the Soviet Union were also involved in research at their own levels in this area. The 1950s saw the height of the Cold War between the US and the Soviet Union, and it was but natural if both countries attempted to outdo the other in military research.

However, the Automatic Language Processing Advisory Committee (ALPAC) report published in 1964 in the US put a halt to the burgeoning interest in MT research. The report gave a negative feedback on MT, saying that "MT was slower, less accurate and twice as expensive as human translation, and that there was no immediate or predictable prospect of useful MT" ("Machine Translation", 140). This was a dampener, and government funding tapered off in the US. However, there were research groups in various parts of the developed world which were working using private funds. The TAUM group in Montreal managed to develop a translation system that was able to replace human translators in the translation of weather bulletins from English to French. SYSTRAN, a translation system developed by Peter Toma in California was used by NASA to translate from Russian to English. These private ventures again revived the waning interest in MT research leading to the Commission of the European Communities (CEC) to fund the Eurotra in 1978, which is the largest MT project ever undertaken till date..

MT research took off again after this and commercial MT systems began to make their appearance. There was more interest in this in Europe than in the US. The 1970s also saw rapid changes in the fields that contributed to MT – in Linguistics and Computer Science. There was immense speculation especially in the field of Computer Science, about the prospects of Artificial Intelligence or the 'intelligence' that was required of a machine to translate a text.Computer scientists work in tandem with linguists to understand the process of language cognition, which is a very significant component in developing artificial intelligence. This has contributed to the development of MT research..

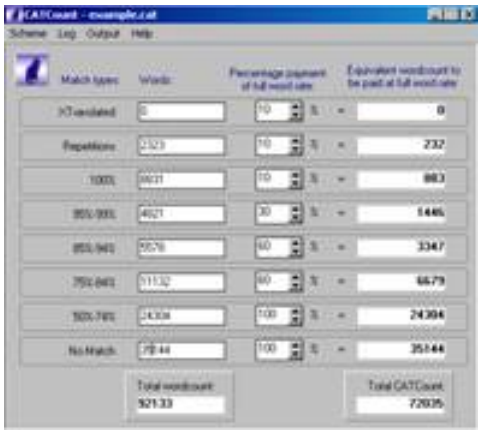
Translation Memory

Translation that was done with the help of the computer had mainly three tasks to perform. These were editing, terminology management and the translation itself. Special software programmes were developed for editing, with translators in mind. One of them, for example, helped translators who had to overwrite the source text or rewrite the source text extensively.This programme prevented the whole text from being overwritten accidentally.

Terminology management means "collecting subject-specific terminology, entering the terminology in a machine-readable glossary or terminological database, and making sure that all this can be accessed from the translation editor during the actual translation process" ("Machine-aided Translation", 135). Software was developed for this function also in machine-aided translation.

Besides this, there were other programmes that helped the translator to choose equivalent words and phrases and help her in the decision making process in translation. The main tool thus developed was translation memory (TM). This was one of the earliest technologies to be developed in the 1970s in connection with translation, and began to be sold at a commercial level from the 1990s. This tool is capable of advising the translator regarding the translation of complete sentences or larger parts of the text. It "allows the translator to store translations in a database and 'recycle' them in a new translation by automatically retrieving matched segments (usually sentences) for re-use" ("Computer-aided Translation", 48). Translation Memory has a database that contains selections from source and target texts, arranged in pairs. These pairs are called Translation Units (TU). The way TM functions is like this – it divides the source text into various segments. Each of these source text segments is compared to a source text segment that is already stored in the database. When it finds a matching segment, it retrieves the relevant translation unit for it. In ordinary terms, the translation memory database searches for suitable translations for the source text. In doing so, it compares the source text with the selections it already has in its database. If it finds a suitable translation, it informs the translator about it.

Types of Units



The relevant TU that is retrieved will match the source text segment. What is a text segment? From the TM perspective, a segment is a smaller unit of the source text like a sentence or a word. The basis of TM technology is the splitting of texts into various segments and then trying to align them. The process of splitting up a text into segments or smaller units is called segmentation. The most common unit is the sentence, but there are systems that consider headings, lists etc also as segments. Problems in alignment or matching can occur if the SL and TL do not match in terms of segments. For example, a language like English which is a Latin-based script and Chinese will be very difficult to align. These languages differ in segmentation or the way in which they can be split up into smaller units.

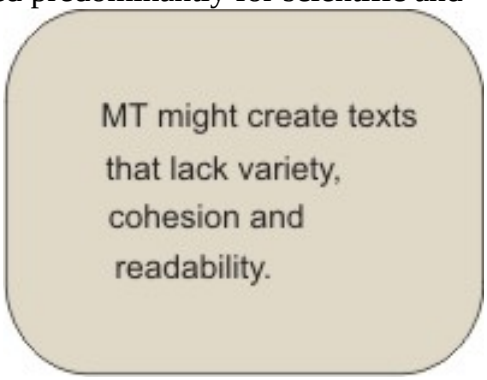
But the ways in which it matches the source text varies. The matches are classified as ‘exact’, ‘full’ and ‘fuzzy’. An exact match means that the source text to be translated is identical to a source text segment that is stored in TM. This does not require any change even in formatting and can be used as it is. A full match implies that the segment stored in the memory is almost identical, with a few variables like date and time that can be changed easily. A fuzzy match is one where the source text is similar to the one stored in memory, but the matching segment can be used only with a bit of editing.A fuzzy match comes closest to the human process of choosing an equivalent for a word or sentence in translation. In computer terminology, this is a much more sophisticated technology than the exact or full matches.

TM technology has developed a great deal since its inception. It used to be able to retrieve only sentences initially, but second generation technology was able to retrieve fuzzy matches. Third generation TM technology that is aiming for matches at the sub-sentence level is emerging now. This is because translation theorists have pointed out that a translator in actual practice might not consider the sentence as a basic unit. But as O’Hagan observes, an ideal TU is yet to be identified.

Previous Next

Impact of TM

Translation tools help the translator, but how is it received? CAT is used predominantly for scientific and technical translation, and so the mechanical nature of finding 'matches' in the stored memory is successful. However, it is successful only to a certain extent. Even when TM is able to come up with exact matches for source text segments, the translator will have to consider the text as a comprehensive whole and judge if the segment will fit. TM is capable of creating what is called a 'sentence salad' effect, "when sentences are drawn (without adequate contextual information) from various translation memories created by different translators with different styles"("Computer-aided Translation", 50). Sometimes the cohesion and logical flow of arguments in the translation can be compromised for the sake of using TM. In their attempt to find suitable matches for source text segments, translators avoid using certain grammatical constructions. This will yield consistent matches, which is good as far as technical translations are concerned. But the end result might be a text that is monotonous, and lacks in readability and coherence.



Michael Cronin draws our attention to an aspect of translation that can easily be forgotten in the eagerness to automate the process, which is that translation is not merely a synchronic process, but a diachronic one. He observes: "…students are increasingly adept at using the tools which allow for rapid information retrieval or the quick dispatch of text but deep, historical knowledge of languages and cultures which takes time and effort is not always valued in a culture of informational ubiquity" (21).Computers can provide exact matches for words, but do not have an awareness of the history behind works and the factors that have moulded them. This makes for a thin translation that is word perfect but nothing more.

Cronin quotes Reinhard Schäler to point out that one problem that has been noted in connection with the rise and spread of MT is the translators' declining concern about the quality of translations. Schäler points out that "accuracy and consistency of the translation (features associated with MT output) take precedence over style, readability and naturalness, all associated with the traditional values and reference system of the human translators" (qtd in Cronin, 22). MT without human intervention can have undesirable consequences.

Previous Next

The future

However, we cannot afford to be cynical about the role of machines in translation. We have to accept the fact that translation and translators are located in the technologically advanced world today, and can afford to ignore technology only at their peril. Cronin is of the view that translators should ideally be poised between the sciences and the humanities, as practitioners of a 'third culture' (112). They interact with, and negotiate several disciplines that occupy an 'in-between' space, an area with fuzzy boundaries. They travel constantly between languages and cultures, but are also bound by the disciplinary peculiarities. Cronin's term for these "translation agents of the new millennium" is "translational cyborgs who can no longer be conceived of independently of the technologies with which they interact" (112). They do not constantly draw upon material from scientific or technical matter, but their identity is modified by the technological developments in translation methods. They are termed cyborgs because they are informed by technology as well. However, they refuse to be taken over completely by machines.

There were many people who predicted that once the concept of MT gained ground, we can expect the slow demise of human agency in translation. This has not happened; in fact, statistics reveal that there are more human translators than ever before. Cronin observes that the problem with such predictions is that it prefers substitutive thinking over relational thinking: "That is to say, the notion of the machine fully replacing the translator or becoming a wholly adequate substitute for the translator is considerably less plausible than the emergence of translational cyborgs where the levels of interaction between humans and machines are deeper and more extensive, with the strengths of each relating to the other in an optimal and mutually complementary fashion" (116). This is a much more positive outlook which conceptualizes a future which does not deny or downplay technological development and where machines and humans cooperate to produce faster (and perhaps better?) translations.

Assignments

1. What are the various ways in which machines or computers can aid the activity of translation?
2. What does the future hold for machine translation?

References

Cronin, Michael. *Translation and Globalization.* London: Routledge, 2004.

Freigang, Karl-Heinz. "Machine-aided Translation". *Routledge Encyclopedia of Translation Studies*. Ed.

Mona Baker. 1998. London: Routledge, 2001: 134-136.

Somers, Harold L. “Machine Translation”. *Routledge Encyclopedia of Translation Studies*. Ed. Mona Baker. 1998. London: Routledge, 2001: 140 – 143.

O’Hanagan, Minako. “Computer-aided Translation”. *Routledge Encyclopedia of Translation Studies*. Eds. Mona Baker and Gabriela Saldanah. 1998. London: Routledge, 2009: 48 – 51.

Previous Next