Z E S Z Y T Y N A U K O W E UNIWERSYTETU RZESZOWSKIEGO
SERIA FILOLOGICZNA
ZESZYT 93 / 2016 STUDIA ANGLICA RESOVIENSIA 13

doi: 10.15584/sar.2016.13.9

Karolina PUCHAŁA-LADZIŃSKA
University of Rzeszow
karolinapuchala1@gmail.com

MACHINE TRANSLATION: A THREAT OR AN OPPORTUNITY FOR HUMAN TRANSLATORS?

Abstract: The motivation behind any kind of development is always the wish to make a certain area of life easier. Hence, it is not surprising that attempts have been made to eliminate the barrier which has been a nuisance for people since time immemorial – the language barrier. Nowadays linguists and computer scientists worldwide strive to create miniature computers that could serve as universal translators, capable of translating between any two languages. This idea, once belonging to the realm of dreams only, is something that Google, among others, intends to turn into a reality.
The purpose of the article is to explain the idea of machine translation (MT), present a brief outline of its history, evaluate the progress that has been made in this field of translation studies, as well as discuss the future of machine translation and try to answer the question whether it is likely that human translators will one day be replaced with artificial intelligence.

Key words: machine translation (MT), Natural Language Processing (NLP), hypothesis of strong artificial intelligence.

The history of machine translation: a brief outline

Translation is a fine and exacting art,
but there is much about it that is mechanical and routine […][1]

Although the idea of a translating machine may seem to be quite recent, the first ideas of mechanizing translation date back to the 17th century (Hutchins 2005; Kozłowski 2002).[2] However, it was not until the 20th century that they could be

[1] Kay, M. 1997. "The Proper Place of Men and Machines in Language Translation", [in:] *Machine Translation* 12, pp. 3-23.

[2] Information included in this section comes from Hutchins, J. 2005. "The History of Machine Translation in a Nutshell". Retrieved on: 10 January 2015. Available at: http://www.hutchinsweb.

put into practice. The first inventors to apply for patents for translating machines were Georges Artsrouni and Petr Troyanskii, the latter of whom introduced not only the idea of an automatic bilingual dictionary, but also an Esperanto-based, language-independent system for coding grammatical structures and conducting the processes of analysis and synthesis. However, ideas of Troyanskii remained unknown until the end of the 1950s.

The year 1954 saw the introduction of the system designed by IBM and Georgetown University. Despite applying very restricted vocabulary and simple grammar, the system aroused much public interest which led to the acquisition of considerable funds from the US for researching machine translation. It was around that time that machine translation became the area of worldwide interest.

The early systems employed in machine translation comprised large bilingual dictionaries, in which every source language entry had one or more corresponding entries and some word order rules in the target language. The general optimism lasting for about a decade in the field of machine translation began to taper off as more and more semantic barriers were encountered, for which no simple solutions could be suggested. Besides, the quality of the target texts produced by the working systems, such as Mark II or the one from the Georgetown University, left a lot to be desired. Consequently, the lack of progress made the US government establish the Automatic Language Processing Advisory Committee (ALPAC) which, in its report produced in 1966, concluded that machine translation is slower, less accurate and more expensive than human translation and, as a result, there seems to be no point in further investment in MT. Instead, ALPAC proposed the development of tools which were supposed to assist translators in their work, such as automatic dictionaries.

In the aftermath of the ALPAC report, the research in machine translation was discontinued in the US for more than a decade, which also slowed down the progress in this area in the Soviet Union and Europe. Nevertheless, MT research was continued in Canada, Germany and France, the result of which was, among others, the successful system METEO in Canada used for rendering weather reports (developed in 1970 at the Montreal University).

In the 1960s MT research in the US and the Soviet Union was focused on developing systems that could translate scientific and technical documents from English into Russian and vice versa. Because these documents were addressed to a limited number of recipients, mostly scientists, their translations were to be communicative and as long as they could correctly convey the source text message, it was not really important that they sounded unnatural.

Starting from the mid 1970s the demand for machine translation changed as a result of rapid international trade. Thus, administrative and commercial demands of multilingual communities stimulated the development of MT in Europe, Canada

---

me.uk/Nutshell-2005.pdf, and Kozłowski, S. 2002. „Co to jest tłumaczenie maszynowe" retieved 10 January 2015 from http://kf.mish.uw.edu.pl/kog/kog_seb.pdf.

and Japan. There was a need to develop cost-effective machine translation systems capable of rendering commercial and technical documentation between the main languages used in international trade.

In the 1980s numerous MT systems were developed: SYSTRAN (performing translation between multiple languages), LOGOS (German-Russian and English-French), a system for internal use within the Pan American Health Organization (Spanish-English and English-Spanish), METAL (German-English) and several systems for the English-Japanese language pair (created by Japanese corporations). Also, the research on some more advanced systems was continued. The predominant strategy around that time was translating by means of language-independent representation.

However, the end of the decade was an important turning point. A group of researchers from IBM laboratories published the results of the experiments conducted on the system called CANDIDE, which used only statistical methods. At the same time some of the Japanese research teams started using corpus-based methods (corpora of translation examples), initiating the so-called "example-based" translation, as opposed to "rule-based" translation favoured before. Significant progress in the early 1990s was achieved also thanks to research on speech translation, including speech recognition and synthesis. Around that time a change of focus in MT research could be observed: from purely theoretical research to practical applications, and this trend continued throughout the 1990s. The use of MT by large corporations has grown rapidly, particularly in the field of software localization (adaptation of computer programs and games to the target language recipients), the sales of MT software for personal computers has increased substantially and MT has been offered by a growing number of online services, thus becoming easily available for any user having an Internet access.

Nowadays, a growing demand for translation can be observed, especially in the areas of international commerce and administration. The European Union itself is like a Babel Tower with 28 participating countries and so it is not surprising that the European Commission employs 3.5 thousand translators translating documents, legal acts, bills and other texts from and into 23 official languages of the EU.[3] Together with new states entering the EU, the demand for translation is growing.

Machine translation: the devil is in the detail

As explained by Kozłowski (2002:64), machine translation is an automatic translation of a written text or speech from one language to another and constitutes one of the most important applications of Natural Language Processing (NLP) – a field which combines the elements of IT and linguistics.

[3] Available at: http://europa.eu/about-eu/facts-figures/administration/index_pl.htm. Retrieved on: 10 January 2015.

Large amounts of money have already been invested into the research on machine translation, with a view to developing a system capable of rendering all kinds of texts. However, the results achieved so far are still not satisfactory and it is still unknown whether machines will ever be able to translate as well as humans.

The idea of a machine imitating human mind theoretically seems probable, especially when taking into consideration the hypothesis of strong artificial intelligence proposed by John Searle, whereby human mind is working in complex algorithms and, therefore, it is possible to construct a machine that could successfully perform any intellectual task that a human being can (Penrose 1996:28-38). This hypothesis implies that it is possible to build a machine that could perform all the creative tasks, which have so far been an exclusively human domain, such as writing poetry, creating art, or composing music. Even if one day this hypothesis proves to be true, one question remains: would machines be able to do all these creative tasks equally well as human beings?

Translation is a challenging and difficult task as it is a creative process and in order to be able to perform it, one not only needs linguistic skills, but also the general knowledge about the world and different cultures. For a machine to be capable of translating the way human translators do, it would need to have all this knowledge uploaded into its software. Another question arises here: is it possible to collect all the possible facts comprising the knowledge of the world and upload it into a machine?

As noticed by Kozłowski (2002:66), it appears that the problem of translating a text from one language into another is far too complex to be fully automated. The range of this automatization would undoubtedly depend on the similarity of languages and proximity of cultures – it would probably be easier to create a machine that would effectively translate between two Indo-European languages, rather than between two geographically and culturally distant languages (for instance Polish and Chinese).

Due to its various imperfections, machine translation has not enjoyed much respect and has often been a subject of ridicule. There is no shortage of examples in which machines proved that no matter how many words their databases have, they are still unable to translate like human beings. The frequently quoted instance is the one in which the English sentence *The spirit is willing, but the flesh is weak* (from the Gospel of Matthew) was translated by one of the first machine translators into Russian, and then back-translated into English with the end result: *The vodka is strong, but the meat is rotten* (Hutchins 1995:17). A similarly unsuccessful effect was achieved when trying to get the machine to translate the idiomatic phrase *Out of sight, out of mind* into Japanese, which the machine dutifully did, but the end result, when translated back into English, was: *Invisible idiot* (Hutchins 1995:18).

These two examples, as well as all the other of the numerous instances of machine mistranslations, illustrate the fact that, so far, machines have been unable to recognize idiomatic phrases and they remain insensitive to various nuances in

human communication, and without these two significant skills, the output that they produce in translation is likely to be unsuccessful in delivering the source language message. This, in turn, precludes effective communication.

Types of machine translation

Hedden (2000) divides machine translation into the following types:

— Machine Aided Human Translation (MAHT), also known as Computer Aided Translation (CAT) – the translation is performed by a human translator who uses a computer system in order to precipitate the entire process. MAHT systems provide the translator with general dictionaries, dictionaries of specialist terminology, dictionaries of synonyms and frequently used expressions, the functions of spellcheck and grammar check, thus allowing the translator to focus on the more creative part of the task. Such systems are frequently integrated with text editors. Although this type of translation belongs to the field of machine translation, it is the human translator who is of paramount importance here, with the actual machine serving only as an aid.

— Human Aided Machine Translation (HAMT) – the source text is modified before, during or after being translated by the computer. For instance, a person modifies the form of the text before translation in order to make it easier to process by the computer, or introduces some extra information into the source text to facilitate its further analysis performed by the computer. Types of text translated in this way are instruction manuals, which need to be rendered into various languages. The systems which are given an unmodified text in the source language frequently need human assistance during translation, which entails that the human translator removes all the semantic ambiguities. Once the text is no longer ambiguous, such systems then prove to be very fast and effective in translating it from the source language into multiple target languages. However, as noticed by Kozłowski (2002:68), those systems are not very popular among the users, as the users need to mechanically answer the questions posed by the computer and thus do not have the full control over the end result.

— Fully Automated Machine Translation (FAMT) – the program receives a source text and produces a target text without any human interference. Usually such an automatically-translated target text is of a poor quality but it is still sufficient for certain applications in which access to the information is of primary importance and accuracy of translation recedes into the background, for example browsing the Internet sites or reading e-mails. Nevertheless, this type of translation does not enjoy much popularity among human translators who prefer to translate the entire text themselves rather than correct the imperfect computer-generated translation.

Machine translation: how does it work?

Machine translators differ in the way they work. As can be read in *Wikipedia,*[4] there are four main modes of operation:

— rule-based machine translators – they are fed information about the linguistics of the two languages and hence, when performing translation, they make use of the morphological, syntactic and semantic rules of the given languages. The main problem related to this kind of translators is that all the exceptions from the general rules and all the possible instances of ambiguity need to be accounted for and fed into the source language analyzer for the program to be able to cope. If the source text contains errors, the system is unable to translate it unless these errors were also entered into the analyzer. Since such translators remain insensitive to linguistic and contextual nuances, they will encounter problems translating homonyms, for instance the Polish word zamek which, depending on the context, may be translated into English as castle, lock or zip;

— transfer-based machine translators – they create a translation from an intermediate representation which stimulates the meaning of the original sentence;

— statistical machine translators – translations are generated on the basis of statistical methods making use of bilingual text corpora (for instance EUROPARL – the record of the European Parliament). However, such corpora are still rare in some fields and for some language pairs. One of the companies to use this type of systems is Google, which, having previously used rule-based systems, switched to statistical translation method in 2007. Google Translate and other similar systems compare and detect patterns in millions of documents which have previously been translated by humans and make intelligent guesses based on the findings. They also calculate how often certain words occur together in a given language. The statistical method solves the problem of ambiguity: if the Polish word zamek occurs close to any form of the verb mieszkać, then it has to be translated into English as castle, and if zamek occurs next to the word naprawiać, then its English equivalent would be lock. The quality of translation performed by such systems improves together with the increased number of human-translated documents in a particular language. Thus, those systems are largely dependent on the availability of parallel texts;

— example-based machine translators – the corpus used contains texts that have already been translated. In order to translate a given sentence, the system selects sentences with similar components from this corpus. These similar sentences are then used to translate the components of the original sentence and then put together to form a complete sentence in the target language.

It seems that statistical machine translators have been most successful so far. However, they are surely not free from imperfections – they rely heavily on the

[4] Availavle at: http://en.wikipedia.org/wiki/Machine_translation. Retrieved on: 18 January 2015.

information from their databases but, so far, this information cannot be adjusted to the specific character of the text at hand and such a customization is necessary in order to produce an effective, good-quality translation. As observed by Osmałek (2014),[5] machines are, as yet, unable to recognize linguistic nuances and in a situation when a human translator easily finds a synonym appropriate to the given context, the machine simply translates the word. While machines often produce target texts of an acceptable accuracy level, such texts will always lack the "human factor", which is indispensable in certain text types, for instance advertisements. Therefore, it seems that whenever the only purpose of the text is to convey general information, machine translation will do. But when the source text has an aesthetic value and/or is supposed to make a certain effect on the reader or persuade the reader to do something, employing an experienced human translator seems to be necessary, as only a human translator is able to ensure that the translated text still fulfils all the functions of the source text.

Google Translate: becoming better and better

As stated in *Wikipedia,* Google Translate is a translation service created by Google Inc., able to provide an instant translation of a word, phrase, fragment of a text or an entire website from and into 90 languages.[6] The languages supported by the service range from the fairly common ones like English, German or Spanish, to such "exotic" ones like Uzbek, Sesotho or Sinhala. In 2013 the company stated that its translator served 200 million people daily. However, users of Google Translate do not always need to manually enter a piece of text; as observed by Schultz (2013),[7] a confusing Japanese sign only has to be photographed with a smartphone for the Google application to automatically recognize the text in the image and provide its translation. Nowadays users only need to point their camera at a sign or text to see the translated text overlaid the screen. A recently added feature, as stated in the *Daily Mail*,[8] is a conversation mode to translate speech in real time, which means that the application is now able to recognize speech in various languages. In practice this entails that as the users speak to the application, a written translation appears on the screen.

As already mentioned, Google Translate is based on statistical analysis – the system depends on the corpus of equivalent words and expressions in the source

[5] Information retrieved on: 9 January 2015. Available at: http://blog.supertresc.pl/dlaczego-tlumaczenia-maszynowe-sa-nieprzydatne#.VqDSt1K2on.

[6] Available at: http://en.wikipedia.org/wiki/Google_Translate. Retrieved on: 19 January 2015.

[7] Available at: http://www.spiegel.de/international/europe/google-translate-has-ambitious-goals-for-machine-translation-a-921646.html. Retrieved on: 9 January 2015.

[8] Available at: http://www.dailymail.co.uk/sciencetech/article-2908490/Google-Translate-app-adds-conversation-mode-translate-SPEECH-real-time.html. Retrieved on: 19 January 2015.

and target languages and the relevant data are selected from this huge database by means of statistical models. However, as already stated, until 2007 Google was using rule-based models but it soon turned out that rules are too inflexible and frequently prove too much for the computer to handle. For this reason, Google switched to statistical models, whereby algorithms search through huge amounts of information, gather relevant data and learn at the same time. For instance, when the program is required to translate a sentence from one language into another, what it does is search for matching phrases which already exist and, based on these findings, works out how best to compose the target sentence. For some languages and texts this proves to work quite well; the problem might be with some less known languages. Because the accuracy of Google translation largely depends on the size of the database between the source and the target language, this accuracy may suffer when translating between some more "exotic" languages in which there are not many texts available. Also, such machine translations are likely to be more accurate when translating between structurally similar languages, i.e. the ones which have similar grammar and word structure. As stated by Franz Josef Och, the former head of machine translation in Google, the combination of English and Spanish works very well, but translating between English and Japanese is not nearly as effective.[9]

But even if Google translates between fairly common, structurally similar languages, there are still certain elements that it cannot cope with, for instance text ambiguity, and although it is able to produce a comprehensible translation, it will not be considered successful as all the elements comprising the text's aesthetics will simply be lost in translation.

An important aspect of Google Translate is that it constantly updates its databases thanks to the users' feedback, adding new words, phrases and expressions, particularly technical ones. Nevertheless, despite constantly becoming better and better and being able to produce a translation within seconds, it is still human translators who have the upper hand in the field of translation, especially when the text to be translated fulfils functions other than only the communicative one.

The future of machine translation

However, Google Translate is not the only project aimed at making fast translations available to everyone. In fact, it faces growing competition from other corporations. For instance, Facebook already acquired a company which developed a system for automated translation and the social network intends to use

[9] Schultz, 2013. "Translate This: Google's Quest to End a Language Barrier". Retrieved on: 9 January 2015. Available at: from http://www.spiegel.de/international/europe/google-translate-has-ambitious-goals-for-machine-translation-a-921646.html.

the system to facilitate communication among its users who come from more than 200 countries and are more than a billion in number.[10] Also the European Union has designed its own free of charge system for automated translation – MT@EC,[11] which has been developed for translating EU-related texts. It is also predicted that due to budget cuts, the EU will need to eliminate about 10 percent of its human translators in the foreseeable future (Schultz 2013).

Yet another corporation to launch an automatic translation system was Microsoft. It developed a program which is not only capable of instant translation, but it also recognizes speech and can translate spoken utterances simultaneously as the person is speaking. A truly innovative feature offered by this system is that it is able to provide a translation not in a machine-like monotone, but in the speaker's own voice assembled from the previously recorded speech samples, as demonstrated in 2013 by Rick Rashid, the head of Microsoft Research, at a conference in China. As he was speaking in English, the computer was translating his words simultaneously into Mandarin and delivering the translated text in Rashid's own voice (Schultz 2013).

Taking all these successful projects into account, as well as the continuous research and innovations in the field of machine translation, it is not surprising that human translators may worry about their job security. After all, machines outdo humans in at least two aspects as regards translation: they can do it much faster and much cheaper, and these two factors are particularly desirable nowadays, when saving time and cutting costs are priorities for most companies. Therefore, some translators fear that too much progress in the field of machine translation will eventually threaten their positions.

So far, however, machine translations have far too many shortcomings to be beneficial in all spheres of life. The output produced by Google Translate and similar systems will suffice only for limited purposes – to find out about the general meaning of the source text message. Nonetheless, creativity and intellect typical of human beings are indispensable elements of translation and, as yet, no software is able to imitate them. It remains to be seen whether machines will one day emulate the human brain so closely that human translators will become redundant altogether.

In the meantime, however, machine translators should be treated as translation aids, where the human translator adopts the role of a post-editor. This means that, instead of translating a text from scratch, the translator checks, proofreads and revises the translation already performed by a machine.[12] One considerable advantage of such a "cooperation" is that it increases the translator's productivity. Obviously, such a "machine and human" translation cannot be effectively applied

[10] Data as of the year 2013, according to Schultz, T. "Translate This: Google's Quest to End a Language Barrier". Retrieved on: 9 January 2015. Available at: http://www.spiegel.de/international/europe/google-translate-has-ambitious-goals-for-machine-translation-a-921646.html.

[11] Available at: https://webgate.ec.europa.eu/fpfis/mwikis/thinktank/index.php/MT@EC.

[12] Retrieved on: 21 January 2015 from http://www.amtaweb.org/mt-for-translators/.

to all text types, but it proves to work well with more formal, organized and structured texts with repetitive patterns and predictable use of terminology, for instance: contracts, annual corporate reports, software documentation, product manuals, to name but a few.[13]

Therefore, it seems that the fears of human translators of being one day replaced with machines are so far unfounded. Nevertheless, it seems inevitable that the role of the translator will change in the future. As machine translators are becoming more and more advanced, human translators might no longer be "proper" translators but rather "editors", correcting the texts previously translated by machines.

References

Hanke, K. 2012. „Coraz lepsze, lecz ciągle niedoskonałe tłumaczenie maszynowe". Retrieved on: 9 Janary 2015. Available at: http://www.goethe.de/ins/pl/lp/kul/dup/uwe/pl9923031.htm.

Hedden, T.D. 2000. "Machine Translation: a Brief Introduction". Retrieved on: 21 January 201. Available at: http://ice.he.net/~hedden/intro_mt.html.

Kozłowski, S. 2002. „Co to jest tłumaczenie maszynowe?". Retrieved on: 10 January 2015. Available at: http://kf.mish.uw.edu.pl/kog/kog_seb.pdf, pp. 63-93.

Hutchins, J. 1995. "The Whisky Was Invisible or Persistent Myths of MT", [in:] *MT News International* 11, pp.17-18.

Hutchins, J. 2005. "The History of Machine Translation in a Nutshell". Retrieved on: 10 January 2015. Available at: http://www.hutchinsweb.me.uk/Nutshell-2005.pdf.

Kay, M. 1997. "The Proper Place of Men and Machines in Language Translation", [in:] *Machine Translation* 12, pp. 3-23.

Osmałek A. 2014. „Dlaczego tłumaczenia maszynowe są nieprzydatne?". Retrieved on: 9 January 2015. Available at: http://blog.supertresc.pl/dlaczego-tlumaczenia-maszynowe-sa-nieprzydatne#.VK_7HnuoMnI.

Penrose, R. 1996. *Nowy umysł cesarza*. Warszawa: Wydawnictwo Naukowe PWN.

Schulz, T. 2013. "Translate This: Google's Quest to End the Language Barrier". Retrieved on: 9 January 2015. Available at: http://www.spiegel.de/international/europe/google-translate-has-ambitious-goals-for-machine-translation-a-921646.html.

Wilks, Y. 2009. *Machine Translation. Its Scope and Limits.* New York: Springer.

[13] More examples of such text types can be found at: http://www.amtaweb.org/mt-for-translators/.