MACHINE TRANSLATION TECHNOLOGIES USED
IN ONLINE TRANSLATION INDUSTRY

E.I. Gimazitdinov, D.A. Morel Morel

Russia, Belgorod State National Research University

The article gives an overview of such machine translation technologies as Rule-based Machine Translation, Statistical Machine Translation, Hybrid Machine Translation as well as of Translation Memory. Peculiarities of their implementation in online translators by different specialized IT companies are also considered.

For more than 50 years of its history the machine translation has grown out of a fantastic project into an amazing reality.

The idea to use computer for translating was suggested in the USA in 1947 after the appearance of the first computers. The first public presentation of the machine translation took place in 1954 and was called Georgetown experiment. The experiment was conceived and aimed to draw public and government attention. It was a paradox but it was based on rather a simple system that only included six grammar rules and its vocabulary contained 250 records. The system was specialized in organic chemistry and operated on the basis of IBM <u>701</u> mainframe. Despite the system simplicity the experiment had a great impact as researches began in different countries such as England, Bulgaria, Germany, Italy, China, France, Japan [Hutchins 1986].

Nowadays, there are many commercial machine translation projects such as PROMT, Systran, Linguatec, Google, Microsoft, IBM, Atril as well as a lot of online translators including Reverso, WorldLingo, InterTran, Windows Live Translator, Ectaco, Google Translate, Yandex.Perevod, PROMT. Set up by Peter Tom in 1968 Systran Company is considered to be one of the oldest machine translation companies. The team headed by Professor Piotrovsky has made a great contribution to the machine translation development in Russia. His laboratory founded the well-known PROMPT Company that designed the first Russian commercial machine translation program on the basis of Piotrovsky's ideas.

Thus, there are three basic machine translation technologies now:

1. Rule-based Machine Translation (RBMT).

This technology is based on the vocabulary data and the grammar rule analysis of specific languages. Such principles of operation are built up on the linguistic description of two different languages (bilingual dictionaries and other data resources containing morphological, grammar and semantic information), formal grammars as well as translation algorithms. Translation quality depends on the volumes of linguistic databases (dictionaries) and the description level of natural languages. In other words, a maximum number of grammar structure peculiarities of both so-called input and output languages should be taken into account [Islamov, Fomin 2013].

There are two types of rule-based technologies. Transfer principle technologies include morphological, syntactic and semantic analyses of any text in the input

language, transformation into the structure of the output language, the synthesis of the text in the output language. The second type known as Interlingua principle technologies relies on the analysis of the input text based on the metalanguage terms and metastructure synthesis of the text in the output language.

Some of the advantages of RBMT technologies are syntactic and morphological accuracy, reliable and predictable results, and possibility to choose a required item and to adjust it.

These technologies also have disadvantages such as much effort and long terms of designing, constant necessity to support language databases.

The companies producing and supplying RBMT technologies are PROMT, Systran, Linguatec.

2. Statistical Machine Translation (SMT).

This technology is based on the selection of the best possible sentence using data taken from bilingual text combination. Such translation systems are built up on comparing large corpora of parallel texts. Parallel text corpus means texts containing sentences in one language and their translations in another language. Statistical Machine Translation profits from a feature "self-training" that means the more parallel text corpora are available to the system, the better result will be provided [Islamov, Fomin 2013].

SMT technologies have a number of advantages: good translation results, easy sentence build-up when sufficient amount of parallel corpora is provided, the opportunity to use the technology for any language pairs.

However, there are some disadvantages such as a limited number of existing parallel corpora, a failure to use morphology and syntax properly, other linguistic problems such as doubling and repetition, blank and omission, information substitution.

The companies producing and supplying SMT technologies are PROMT, Google, SDL Language Weaver, Microsoft, Asia Online, IBM.

3. Hybrid Machine Translation (HMT).

In connection with the fact that SMT and RBMT technologies have disadvantages and have got their development limits machine translation designers expect a technological breakthrough due to the development of hybrid translation technology. This system is based on the combination of RBMT and SMT methods. Such an approach allows to use all the advantages of both technologies (grammar accuracy of RBMT and good translation results of SMT) [Federmann 2012].

Companies-producers of Hybrid Machine Translation technologies are PROMT, Systran.

Thus, there are advantages and disadvantages of machine translation technologies. For example, any content usage and fast translation are considered the strong sides while their disadvantages include the necessity of preliminary setting to provide high efficiency, the dependence of translation quality and settings on the input text quality.

Besides the machine translation technologies mentioned above, there are programs the operation principle of which is based on Translation Memory (TM) technology.

Translation Memory is a bilingual database of frequently used sentences. The basis of this system includes the principle "not to translate the same text twice". This technology compares the document to be translated with the data of an advance created translation database. While scanning the whole text the system finds the segments that have been translated once and uses the translation from Translation Memory database [Grabovsky 2004].

The following advantages of this technology can be pointed out: the opportunity of re-using the translations made earlier, the minimal use of post-editing. On the other hand, one of its disadvantages is the requirement to accumulate preliminary bases. Also, re-using depends on similarity between the translated content and TM bases.

Companies-producers of TM technology are PROMT, SDL Trados, Atril (Déjà Vu systems), OmegaT.

TM and MT technologies are usually used together to translate large volumes of standard documents as each of them fulfils different tasks within the framework of some general assignment: TM bases provide extraction and substitution of the content translated earlier while the new content is translated due to MT system. Thus, the advantage of these technologies combination is the provision of high operation speed and minimal post-correction when translating any content.

In order to understand the operation principles of the largest online-translators such as PROMT, Google Translate и Yandex we should consider the peculiarities of each of them.

The former researchers of the engineering linguistic laboratory of Leningrad State Pedagogical University designed online-translator PROMT. It operates by using the information contained in dictionary entries which are sections of linguistic databases.

PROMT translation was based on the RBMT technology. At the end of 2010 PROMT presented Hybrid Machine Translation technology combining RBMT and SMT methods. The main idea of this combination is that the program creates many sentence variants instead of the only translation variant. Besides, a sentence can have a few hundreds of variants due to word polysemy, various constructions and statistic processing results. Then possible language model allows choosing the best of suggested variants. Thus, the combination of RBMT and SMT systems provides all the advantages of both technologies. In addition, such an approach gives an excellent opportunity to teach and improve the online-translator [Sokolova 2013].

Online-translator Google Translate was designed by the company Google in the middle of 2000s. It is aimed at sites and texts "on the fly" translating.

This online-translator is based on SMT methods. Google Translate uses machine translation self-training algorithm based on the language analysis of texts. The peculiarity of this online-translator is its unique approach to translating process. In comparison with other systems Google Translate does not analyze grammar and vocabulary rules. On the contrary, this system looks for language similarities between the text being translated and tremendous vocabulary database consisting of the translation variants made by humans. Self-training statistic algorithms for building up translation models are also used in this technology. Such method gives an opportunity

to improve the quality and level of the output text as well as to exclude "non-combined combinations" frequent in other translation systems. Now Google Translate operates in 65 languages, translating not only from the most frequently used English, German or Italian but even from Latin or Philippine [Adams 2010].

Another online translator is Yandex. Perevod – the web-service of Yandex Company – has been designed for translating a text segment or web page. This system uses the self-educated algorithm of SMT technology and it makes up its vocabularies of correspondences on the basis of the analysis of million texts translated [O'Hear 2011]. First, the computer compares the input text with vocabulary database, then with language model base trying to define the meaning of an expression according to its whole content.

Three main components can be pointed out. They are the translation model, the language model and the decoder.

The translation model is a table where all the words and phrases included in this system and given in one language have all their possible translation variants in another language. Also, the probability of these translating variants is indicated (every pair of languages has its own table). The creation of the translation model is divided into three stages: first, parallel documents are selected, then pairs of sentences are found there and, finally, pairs of words and word combinations are selected.

The system compares not only separate words but also word combinations consisting of two, three, four or five words following each other. The translation model in Yandex translator contains hundred millions of pairs of words and word combinations for every pair of languages.

Another component of this system is the language model. For its creation the system studies hundred thousands of different texts in the required language and makes up the list of all the words and word combinations used in them indicating the frequency of their use. This is the system knowledge about the language a text should be translated into.

The function of decoder is to translate. It selects all the translation variants for every sentence of the input text by combining the phrases from the translation model and sorting them according to decreasing their possibility. For example, the computer user wants to translate the phrase "to be or not to be". Let us suppose that the translation combination "быть или не бывает" is suggested as the highest probability among all the variants taken from the translation model. The translation variant "быть или не быть" follows it and takes the second place.

Then decoder evaluates all the combination variants by means of the language model. In the example given above the model indicates to the fact that "быть или не быть" is used more frequently than "быть или не бывает". Finally, the decoder selects the sentence with the highest combination probability (due to the translation model) and the frequency of its use (according to the language model).

In conclusion we would like to emphasize the following.

As in general online translators provide rather bad quality of translation it is obvious that a person who knows the language perfectly is required for checking and correcting the final result. In addition, we should take into account that any

documents implying some legal responsibility are supposed to be under a human being's checkout.

Despite all discovered drawbacks machine translation is very popular and has a lot of advantages. It gives opportunities to understand the general content of the text very fast and to analyze multilanguage information from the Internet quickly, to optimize translation of large volumes of texts on the same topic, to have rapid business e-mail correspondence with foreign partners as well as to communicate in social networks and various forums, to retrieve some information from a large volume of texts on the basis of linguistic characteristics of words, to economize one's time and money.

REFERENCES:

1. Adams, T. (2010). Can Google break the computer language barrier? // The Guardian. – URL: http://www.theguardian.com/technology/2010/dec/19/google-translate-computers-languages.

2. Federmann, C. (2012). Hybrid Machine Translation Using Joint, Binarised Feature Vectors // Proceedings of the 10th Conference of AMTA. – San Diego: AMTA. – URL: http://amta2012.amtaweb.org/AMTA2012Files/papers/126.pdf.

3. Grabovsky, V.N. (2004). Tekhnologiya Translation Memory // Mosty, 2. – Pp. 57-62.

4. Hutchins, W.J. (1986). Machine Translation: Past, Present, Future. – Chichester (UK): Ellis Horwood. – 382 p.

5. Islamov, R.S., Fomin, A.G. (2013). Analiz sovremennyh sistem mashinnogo perevoda tipa SMT i RBMT // Filologicheskie nauki. Voprosy teorii i praktiki, 3(21)/1. – Pp. 69-73.

6. O'Hear, S. (2011). Anyone speak Russian? Yandex rolls out Yandex.Translate // TechCrunch. – URL: http://techcrunch.com/2011/03/16/anyone-speak-russian-yandex-rolls-out-yandex-translate/.

7. Sokolova, S. (2013). How the Computer Translate // PROMT Technologies. – URL: http://www.promt.com/company/technology/pdf/e_how_computer_translates_sokolova.pdf.

PSYCOLOGICAL AND PEDAGOGICAL HELP TO PUPILS WITH SPEECH IMPARATMENTS IN THE COMPREHENSIVE SCHOOL

M.M. Godovnikova
Scientific adviser L.V. Godovnikova

Russia, Belgorod State National Research University

This article is devoted to psychological and pedagogical help to pupils with speech imparatments in the comprehensive school. The problem which is discussed about complex psychological and pedagogical help during educational process for providing for effective adaptation of the child with speech imparatments in schools.

One of the major tasks facing experts, working in most educational institutions, the organization of psychological and pedagogical help to pupils who experience school difficulties. Obviously that category of children of risk group includes children with speech imparatments.

It is necessary to do complex psychological and pedagogical help during educational process for providing for effective adaptation of the child with speech