Application of modern machine translation systems in teaching foreign languages

E V Fibikh and N V Kuznetsova
Reshetnev Siberian State University of Science and Technology, 31 Krasnoyarsky Rabochy Av., 660037 Krasnoyarsk, Russia.

E-mail: fibichev@mail.ru, natalie_ivleva@mail.ru

Abstract. The paper considers the concept of machine translation and its future trends in the modern world. It specifies types of modern translation systems and covers issues the humankind faces with them. It concludes that the future development of this industry is directly connected with the development of artificial intelligence.

1. Introduction
Dramatic improvement in the growth and introduction of information and communication technologies in the educational process, in particular the use of e-learning in the sphere of educational services, has happened over the last few years. The transition process from the traditional learning to the learning on the basis of computer technologies has been evolving gradually over the last decades. With the appearance of the huge archives, presented on machine-readable media, as well as the growth of information and knowledge base, the thought of using this material for teaching purposes, where the only form that could keep up with the intellectual content generation can be electronic resources, has appeared more and more often. In this situation, electronic teaching serves as a new technological environment for the knowledge delivery from open sources [1].

As a result of the wide expansion of the international open educational resources and access to them, machine translation has become more extensively used in recent years. In this regard, machine translation systems and computer dictionaries are very convenient working tools to save time and optimize the process of understanding information in a foreign language. This article is devoted to the analysis of machine translation systems aimed at the translation process automation.

2. Specifications of machine translation systems
The relevance of searching for the ways of maximal translation process automation in teaching foreign languages is caused, on the one hand, by provision of maximal efficiency of work with texts in a foreign language; on the other hand, by the widespread use of CLIL technologies in the process of teaching foreign languages, as one of the key modern methods [2].

Machine translation systems (MTS) of texts from one natural language into the other simulate the work of a human translator. Their usefulness depends on the degree to which the objective laws of the language and thinking are taken into consideration. These laws have not been studied at the proper level yet. Therefore, to solve the problem of machine translation, it is necessary to take into account the experience of international communication and the experience of translation activities accumulated by mankind. In the process of translation, the basic semantic units are not individual words, but
phraseological phrases expressing concepts. These very concepts constitute elementary mental images. Only by using them, one can build more complex images corresponding to the translated text.

Modern machine translation systems are divided into three large groups:

1. Based on the rules;
2. Based on examples;
3. Statistical.

Let us consider rule-based machine translation systems. This concept covers machine translation systems based on linguistic information about the source and target languages mainly derived from (bilingual) dictionaries and grammars covering the basic semantic, morphological, syntactic regularities of each language. This approach to machine translation is also called classical.

On the basis of these data, the source text is gradually, sentence by sentence, converted into the target language. Such systems are often opposed to example-based machine translation systems. The basic principle of such systems operation is the connection of the structure of the input and output sentences. The received translation is typically of no good quality. However, it works on simple examples.

The translation from English to German will look like this:
A girl eats an apple. / Ein Mädchen isst einen Apfel.

These systems are divided into three groups:

- literal translation systems;
- transfer systems;
- interlinguistic systems.

The literal translation systems are rarely used today because of the poor translation quality. The words of the source text are converted (as they are) into the words of the target text. Often such transformation occurs without lemmatization and morphological analysis. This is the simplest method of machine translation. It is used to translate long word lists (such as catalogues). It can also be used to compile interlinear translation for TM (Translation Memory) systems.

Both transfer systems and interlinguistic systems have the same general idea. For translation it is necessary to have a mediator which comprises the meaning of the translated expression. In interlinguistic systems, the mediator does not depend on the pair of languages, while in transfer systems it does.

Transfer systems operate on a very simple principle: rules are applied to the input text, which align the structure of the source and target languages. The initial stage of the work includes morphological, syntactic (and sometimes semantic) analysis of the text to create an internal representation. Translation is generated from this representation using bilingual dictionaries and grammatical rules.

A more “abstract” internal representation can sometimes be built on the basis of the primary representation, which was derived from the source text. This is done in order to emphasize the parts important for translation, and to discard non-essential parts of the text. While constructing the target text, the transformation of the levels of internal representations occurs in the reverse order [4-6].

Using this strategy, one gets a fairly high quality of translations, with the accuracy of around 90% (although this is highly dependent on the language pair). The operation of any transfer translation system consists of at least five parts:

- morphological analysis;
- lexical categorization;
- lexical transfer;
- structural transfer;
- morphological generation.

Morphological generation. Words of the original text are classified by speech figures with the focus on their morphological characteristics and word stems.
Lexical categorization. Some words can have multiple meanings in any text that causes ambiguity within the analysis. Lexical categorization helps derive word context.

Lexical transfer. Word stems are translated through bilingual dictionaries. The process is similar to consecutive translation.

Structural transfer. Words are connecting in a sentence coherently.

Morphological generation. Based on output data of structural transfer word forms are created in the translated text.

One of the main features of transfer systems in machine translation is the stage, at which the transitional text in the original language is transferred to the transitional form of the text in the target language. It may work at any level out of two in the linguistic analysis or both.

Levels:

Syntactic transfer.

This level is characterized by transferring syntactic structures between the original and translated languages. It is suitable for languages belonging to the same group, for example, in roman languages, between Italian, Spanish, Catalanian, French, etc.

Semantic transfer.

This level is characterized by semantic performance. It depends on the original language. This performance may consist of structures which present a meaning. Translation requires structural transfer. This level is usually used for translation between languages from different language groups.

Interlinguistic machine translation is one of the classical approached in machine translation. The original text is transformed into abstract performance which does not depend on a language. Translated texts are created through that performance. The main advantage of this approach is to create a way to add a new language in the system. It is possible to prove mathematically that within this approach it is possible to create a new interpreter for this system which will cheapen the system itself in comparison with the system of transfer transition. Moreover, this approach provides an opportunity to create text retelling, paraphrasing of the original text within one language. It is also a simple form of translation between different language groups, for example, Russian and Arabic.

To be able to use the system of interlinguistic machine translation it is important to have:

- dictionaries for text analysis and generation;
- description of language grammar;
- definition knowledge base;
- definition projection rules for languages and performance.

It is impossible to design a database for a wide range of specializations which is the most difficult part. Those databases which are designed for narrow areas have calculation complexity.

### 3. Role and trends of artificial intelligence in machine translation

Nowadays, we still do not have any approach which would work accurately for two languages. Many experts reckon it to be a far-fetched idea. The most essential problem is projection of interlanguage performance. It must be both abstract, independent of certain languages, and reflect features of any existing language. On the other hand the task on idea eliciting is not solved even in the area of artificial intelligence. Nevertheless, artificial intelligence is considered to be one of the most promising areas in machine translation development.

Artificial intelligence has infiltrated numerous aspects of our lives in recent years, thanks to improvements in the field of machine learning, where computers ostensibly program themselves. This drive towards digital self-learning has led to major breakthroughs in our day-to-day interactions with machines, most notably the rise of digital home assistants such as Amazon Echo, and the recently launched Google Lens, which identifies objects based on visual cues from your phone’s camera.
One of the main advantages of artificial intelligence in translation is that it has the capability of quickly capturing and understanding language nuances that might be challenging to accomplish through other forms of machine translation. As a result, mainstream tools, such as virtual assistants, can be made available in countless languages, especially with companies such as Google, Microsoft, Amazon, Apple, and IBM continuing to develop artificial intelligence technology.

4. Trends in machine translation

Machine translation has been around for over 60 years but has only recently started being taken seriously by the language industry. This is largely due to the advent of neural machine translation (NMT). Even though this artificial intelligence-based approach has its pitfalls, its successes mean that machine translation is here to stay. For the same reason, post-editing machine translation services (PEMT) are in increasing demand. The global market of machine translation is expected to reach $980 billion by 2022 (figure 1).

The dynamics presented in the graph does not mean it is necessary to become a post-editor, or a post-editing agency, right now. It just means you have to adapt your service offering, preparing yourself for the new reality [7].

![Figure 1. Global market of machine translation by 2022.](image)

On the other end of the spectrum, we have transcreation. This service might not account for the largest part of the translation but it is growing, thanks to more and more businesses’ understanding the value of approaching their international clients in the right way. The concept of transcreation involves both localization and creative writing. As a language professional or an LSP, you are expected to adapt texts or even whole marketing campaigns so that they sound and feel as if they were conceived in the target language.

This new translation trend encourages translators to feed their creative side, and LSPs to look for ones who are able to do so. Together with the previous point about machine translation and PEMT, this presents an interesting dichotomy that we may observe unfolding in the language industry over this and subsequent years.

Video content is taking over the Internet. Cisco expects that 80% of Internet traffic will come from videos by 2021. This also means higher demand for translation. For example, on Facebook, some 85% of videos are played back without sound, so subtitles for the target language audience are essential.
When translating videos, it is often important to keep timestamps of the original video and to present the result in the requested format. So having a good grasp of subtitling tools is an important value you can bring to the table.

Forbes magazine expects that the e-learning market will reach $325 billion by 2025. At the same time, most of the e-learning content is currently available in English only. As the English-speaking market reaches saturation, more and more online educators will be seeking translation services for their e-learning courses.

It’s worth noting that most e-courses include not only content, but also the technology that is used to deliver it to students. Therefore, if you are an LSP that wants to offer translation of e-trainings, it is important that you learn the software behind them. Here’s a case study of Idea Translations, a company who made developing and managing e-courses a part of their service offering.

Last but not least, there is a bigger trend that will define not only 2019, but the whole vector of industry development in the years to come. It can be expressed in two words: content economy. The content economy is an emerging paradigm for how modern businesses drive growth in the virtual marketplace. Content has become the lifeblood of both sales and marketing. The sheer amount of content that companies generate to stay relevant is growing at an unprecedented rate. It is fair to say that these days every successful business is, to one degree or another, a content creation business. The only way translation companies can adapt to the changing landscape — and they have to adapt if they want to stay alive and thrive — is to become a connected part of the content delivery loop. It entails a lot of commitment and requires tough choices to be made, but it’s the only way to ride the crest of the content wave instead of being swept away by it.

5. Conclusion
In conclusion, it is obvious that machine translation together with artificial intelligence technology is gaining popularity thanks to availability and comfort in using online translators and mobile applications. However, it is necessary to draw students’ attention to the purpose of such products, their assignment and ways of usage.

References
[1] Kuznetsova N V and Fibikh E V IOP Conference Series: Materials Science and Engineering 122 012008
[2] Kuznetsova N V, Fibikh E V and Piskorskaya S Yu 2017 Experience of Reshetnev Siberian State University of Science and Technology in CLIL-technology implementation Perpektovy Nauki 9(96) 81-5
[3] Bakulov A D, Leontyeva N N and Shalyapina Z M 2007 Local systems of machine translation. Artificial intelligence (Moscow: Radio and communication)
[4] Vannikov Yu V 2010 Text adequacy and types of adequate translation (Moscow)
[5] Gryaznukhina T A, Darchuk N P and Klimenko N F 2000 The use of computers in linguistic research (Kiev: Naukova Duma)
[6] Kapanadze O G 2009 Modern foreign machine translation systems (Moscow: VTsP)
[7] Lynan T 1996 Communication in the Language Classroom (Oxford: Oxford University Press)