The format has the following features:

- it is possible to involve a large number of participants in the process;
- it is possible to consider and repeat narrow questions of the material of the topic being studied;
- there is an opportunity to hold debates in the form of a role-playing game.

For the three formats described, it is important to announce topics well in advance so that participants have the opportunity to prepare. Firstly, it reduces the stress level of participants who, due to their individual characteristics, experience difficulties in the face of the need to speak out in public. Secondly, the preparatory stage makes debates more informative and will allow participants to gain new knowledge.

So, debates are a teaching method, which can be defined as a communicative situation modelled on the basis of a controversial issue, where participants hold different points of view and prove them to the opponent, while refuting their position. Most of the debate formats described by modern researchers are used to work with students who have well-developed skills of speaking in a foreign language. The article describes the debate formats that have been successfully applied in classrooms of students at the initial stage of learning a foreign language.

References

1. Нежурина, Н. Ю. Технология организации и проведения дебатов в процессе формирования коммуникативно-речевой компетенции / Н. Ю. Нежурина, Э. П. Комарова // Вестн. Воронеж. гос. техн. ун-та. – 2014. – Т. 10. – № 3-2. – С. 171–174.
2. Snider, A. Many Sides: Debate Across the Curriculum / A. Snider, M. Shnurer. – New York : International Debate Education Association, 2006. – 305 p.

UDC 372.881

FROM THE HISTORY OF MACHINE TRANSLATION AND COMPUTATIONAL LINGUISTICS

ИСТОРИЯ РАЗВИТИЯ МАШИННОГО ПЕРЕВОДА И КОМПЬЮТЕРНОЙ ЛИНГВИСТИКИ

Stepanov D.A.
Vitebsk State Technological University, Belarus
e-mail flvstu@gmail.com
Степанов Д.А.
Витебский государственный технологический университет, Республика Беларусь

Keywords: computational linguistics, machine translation, rule-based MT, corpus-based MT, example based MT, automated translation.

Ключевые слова: компьютерная лингвистика, машинный перевод, метод, основанный на правилах, корпусный метод, автоматизированный перевод.

Abstract. The article discusses the current state of such a scientific area of linguistics as computational linguistics, and one of its key objects - machine translation, the history of its emergence and development.

Аннотация. В статье рассматривается современное состояние такого научного направления языкознания, как компьютерная лингвистика, и один из его ключевых объектов – машинный перевод, история его возникновения и развития.

In our time, it is difficult to imagine human life without the use of digital technologies – computers and gadgets. Computerization has become an integral parameter of the functioning of society, it undoubtedly affects the quality of development of any social institutions and the standard of living in general. Taking into account the globalization processes in the modern world, the frequency and clarity of the transmission of information in different languages, here we refer to translation, have become an important part of communication both within the Internet and in building real connections in intercultural communication. Computational linguistics began to attract the attention of scientists from various scientific fields in the 1940s and 1950s of the 20th century, even before the creation of powerful electronic computing systems.

On the one hand, we mean "computational linguistics" as "one of the areas of applied linguistics, in which computer programs, computer technologies of organization and data processing. On the other hand, this is the area of application of computer language models in linguistics and related disciplines" [1].

In foreign linguistics, various aspects of computational linguistics in different periods were considered in the works of N. Chomsky, V.H. Yngve "A model and a hypothesis for language structure" (1960), M. Halliday "System and function in language" (1976), J. Bresnan, R. Kaplan "A competence-based theory of syntactic closure. The mental representation of grammatical relation" (1982), T. Vinograd "Computer Software for working with language" (1984), V. Raskin "Natural Language Processing for Information Assurance and Security: Overview and Implementations" (2000), Y. Wilks "Computational linguistics: History" (2006), A. Clark, S. Lappin "The handbook of computational linguistics and natural language processing" (2012).

In Russian linguistics the following works are worth mentioning: O.S. Kulagina "Research on machine translation" (1979), Yu.N. Marchuk "Problems of machine translation" (1983), G.V. Chernov "Machine translation and applied linguistics. Problems of creating an automatic translation system" (edited by G. V. Chernov, 1986), N. D. Andreev "The main directions of the experimental laboratory of machine translation" (1986), N. N. Leontiev "Knowledge base and automatic translation (project multilingual information and reference system)" (1989).

We believe that the links between computational linguistics and academic linguistics in its classical sense are not as numerous and obvious as it might seem at first. In the middle of the 20th century. two prominent scientists W. Yngve and N. Chomsky disagreed about the representation of the mathematical aspects of the structures of the language and its representation in computer encoding. The polarity of opinions lay in

the nature of tree representations (in short, the modeling of specific meanings and the processing of words in complex sentences), as well as the role of processing procedures and resources in the calculation of syntactic structure [1]. W. Yngve argued that such computations should take into account the memory size limits for intermediate structures, which, in his assumption, correspond to the inherent limitations on language processing. N. Chomsky, in turn, to describe such processes, resorted to such a term as "language performance", which is used to describe "the actual use of the language in specific situations".

In the 1960s attempts have been made to program Chomsky's transformational grammars for parsing sentences: the largest and longest attempt was made at International Business Machines (IBM) in New York. The attempts were generally unsuccessful in the sense that the programmed grammars parsed almost nothing beyond the sentences for which they were designed, and even in the case of sentences, gave a large number of interpretations between which it was impossible to choose. The latter was the lot of almost all parsers until recent statistical developments, including the original Harvard Kuno and Oettinger parsers, which were developed in 1962, and parsers based on the more complex linguistic grammars of the 1970s and 1980s.

In the 1980s the formalism of lexical-and-functional grammar, created by scientists R. Kaplan and J. Bresnan (1982) [2], and the functional unification grammar of M. Kay (1984), were developing thanks to the grammar of M. Halliday, also known as a unifying logical paradigm for grammar processing, which appeared with the Prolog programming language.

It must be said that in scientific and popular science literature, machine translation is often erroneously identified with computer-aided translation. However, these concepts shall be delimited.

Machine translation is one of the areas of computational linguistics that studies automatic translation from one language to another using automatic programs (Google Translate, DeepL, Yandex Translate). Automated translation is a translation performed by the translator using computer tools: applications, online dictionaries (Reverso Context, Abby Lingvo), etc. The concept of machine translation arose in the second half of the 20th century and was first used by Warren Weaver, an American scientist, who put forward the concept of MT in 1947 after the development of the first computer. Weaver's idea was based on the concept of "Interlingua" (Interlingva). “Interlingua is an international auxiliary language developed in 1937–1951. International Auxiliary Language Association (IALA))” [2].

The main methodological principles of machine translation allow us to distinguish two main types of it: rule-based MT (RBMT) – a method based on rules, and corpus-based MT (CBMT) – a method based on corpora, respectively, it is customary to talk about two approaches to machine translation: rule-based and corpus-based. In addition, machine translation can be carried out using methods such as tree-based method (tree-based method), example-based (based on an example) and rule-based statistical (rule-based statistical method). Currently, the rule-based statistical method (RBSMT) is most often used. Rule-based machine translation methods use bilingual dictionaries

and handwritten rules to translate texts from the source language into the target language. However, manual translation of writing rules is a time-consuming task. In addition, rules are difficult to maintain, as well as difficult to transfer from one domain to another and from one language to another. Thus, rule-based systems are difficult to scale for open domain translations and multilingual translation.

In 1954, Georgetown University, in collaboration with the now well-known computer manufacturer IBM, completed the first Russian-English machine translation experiment using an IBM-701 computer, showing that there is a future for machine translation: the problem of machine translation was relevant for the next ten years. years, but its active development stopped with the release of the report of the Automatic Language Processing Advisory Committee (ALPAC) in 1966. In 1965, Natural Learning Processing specialists held the first International Conference on Computational Linguistics, which was devoted to parsing and translation based on rule-based ma-chine translation. Since the 1970s RBMT methods have become more thorough. In 1978, SYSTRAN, one of the first machine translation companies, released a commercial translation system that became one of the best-known examples of a commercially successful rule-based system at the time. Google used SYSTRAN until 2007 [3].

In the early 2000s with the advent of bilingual corpora of natural languages, linguists in Russia and foreign countries began to pay great attention to the problems of corpus linguistics. Machine translation uses three corpus methods: Example based machine translation (EBMT), Statistical machine translation (SMT) and Neural machine translation (NMT). If we talk about the history of the issue, then back in the mid-1980s. The EBMT method was proposed, which allows translating source texts by extracting pairs of similar sentences from a bilingual corpus. Translation results using EBMT methods are of high quality if similar pairs of sentences can be found. The practical implementation of the theory of N. Chomsky was associated with the problem of the limitedness of programmed texts, which did not have great functionality and processed the context and meaning of language units only within one sentence. Later, machine translation introduced such methods as rule-based MT (RBMT) – a rule-based method, and corpus-based MT (CBMT) – a corpus-based method. First, the rule-based method is actively used to create bilingual dictionaries. However, its complexity lies in the fact that the manual translation and writing of the rules are carried out directly by the researcher. Thus, rule-based systems are labor intensive for open domain translations and multilingual translation.

After some time, when scientists began to actively develop corpus linguistics, several corpus types of machine translation were formed: based on examples (Example Based Machine Translation / EBMT), statistical machine translation (Statistical Machine Translation / SMT) and neural machine translation (Neural Machine Translation / NMT). Despite the complexity and significant drawbacks in working with corpora, these methods have contributed to the improvement of machine translation and were able to advance computational linguistics.

In general, there is still a long way to go to achieve high quality machine translation. It is necessary to develop new methods that could combine symbolic rules and neural networks to further improve the quality of translation. Fortunately, the use of machine translation in real-world applications continues to provide more and more data, driving the rapid development of new machine translation methods.

References

1. Баранов, А. Н. Компьютерная лингвистика / А. Н. Баранов // Большая российская энциклопедия [Электронная версия]. – URL: https://bigenc.ru/linguistics/text/2087783.
2. Clark, A., Fox C., Lappin S. The Handbook of Computational Linguistics and Natural Language Processing. – John Wiley & Sons, 2012. – Vol. 118.
3. Wang, H. et al. Progress in Machine Translation / H. Wang [et al.] // Engineering. – 2022. – Vol. 18. – Issue 11. – P. 143–153.