

РАЗДЕЛ II. ПРИКЛАДНАЯ ЛИНГВИСТИКА  
SECTION II. APPLIED LINGUISTICS

UDC 81`33

DOI: 10.18413/2313-8912-2024-10-2-0-2

Viktoriya V. Dyomochkina<sup>1</sup>   
Dmitry Yu. Gruzdev<sup>2</sup>   
Elena V. Lukyanova<sup>3</sup> 

Machine translation in hindsight

<sup>1</sup> Military University,  
14 B. Sadovaya St., Moscow, 115432, Russia  
*E-mail:* [dyomochkina.viktoria@yandex.ru](mailto:dyomochkina.viktoria@yandex.ru)  
ORCID: 0000-0002-4820-9693

<sup>2</sup> Military University,  
14 B. Sadovaya St., Moscow, 115432, Russia  
*E-mail:* [gru@inbox.ru](mailto:gru@inbox.ru)  
ORCID: 0000-0003-2874-6309

<sup>3</sup> Belgorod State National Research University,  
85 Pobedy St., Belgorod, 308015, Russia  
*E-mail:* [lukyanova@bsu.edu.ru](mailto:lukyanova@bsu.edu.ru)  
ORCID: 0000-0002-1479-2227

*Received 31 March 2024; accepted 15 June 2024; published 30 June 2024*

**Abstract:** The paper expands on the analysis of key projects adorning the machine translation (MT) hall of fame and their role in addressing practical tasks. The most successful initiatives suggest that the fledgling MT was contingent on the level of entropy, a.k.a. random nature of natural languages: the lower the indicator, the higher the predictability of the text, and by implication the efficiency of the system. It accounts for the success of the first Georgetown-IBM experiment and Canada's METEO-1. The latter grew into a full-fledged system that for almost a quarter of the 20<sup>th</sup> century, provided English-French-English translations of weather bulletins, boasting high language predictability. Although, in between them the 1964 ALPAC report sowed a seed of doubt of the MT validity, it never aimed at killing the research area at all. On the contrary, it highlighted technics and applications, where the technology had demonstrated promising results, including raw MT, post-edited MT, and M-AT. The authors note a cyclic nature of the development of MT-powered methods and technologies. Today's combination of resources and the way they are used are different very little from those employed in the past century. What makes them stand apart is the maturity of MT technologies, which made it through rule-based, direct, corpus-based, and knowledge-based translation to SMT and eventually to NMT. It has been established that the improved performance comes at a cost of more elaborate and larger data sets, tagged, marked up and annotated for automated use in language models. Taking advantage of these as well as artificial intelligence




(AI), the authors venture into modeling basic text processing scenarios in a bilingual environment. This results in recommendations as to future paths for the improvement of MT technologies in the hands of professional translators by fine-tuning language models individually and pursuing post-editing (PEMT) and pre-editing practices paving the way for more complex transformations and lower equivalence levels.

**Keywords:** Machine translation; Tagging; Annotation; MT quality; Entropy; PEMT; CAT

**How to cite:** Dyomochkina, V. V., Gruzdev, D. Yu. and Lukyanova, E. V. (2024) Machine translation in hindsight, *Research Result. Theoretical and Applied Linguistics*, 10 (2), 21-45. DOI: 10.18413/2313-8912-2024-10-2-0-2

УДК 81`33

DOI: 10.18413/2313-8912-2024-10-2-0-2

Дёмочкина В. В.<sup>1</sup>   
Груздев Д. Ю.<sup>2</sup>   
Лукьянова Е. В.<sup>3</sup> 

Машинный перевод вчера и сегодня

<sup>1</sup> Военный университет  
ул. Б. Садовая, 14, Москва, 115432, Россия  
E-mail: [dyomochkina.viktoria@yandex.ru](mailto:dyomochkina.viktoria@yandex.ru)  
ORCID: 0000-0002-4820-9693

<sup>2</sup> Военный университет  
Ул. Б. Садовая, 14, Москва, 115432, Россия  
E-mail: [gru@inbox.ru](mailto:gru@inbox.ru)  
ORCID: 0000-0003-2874-6309

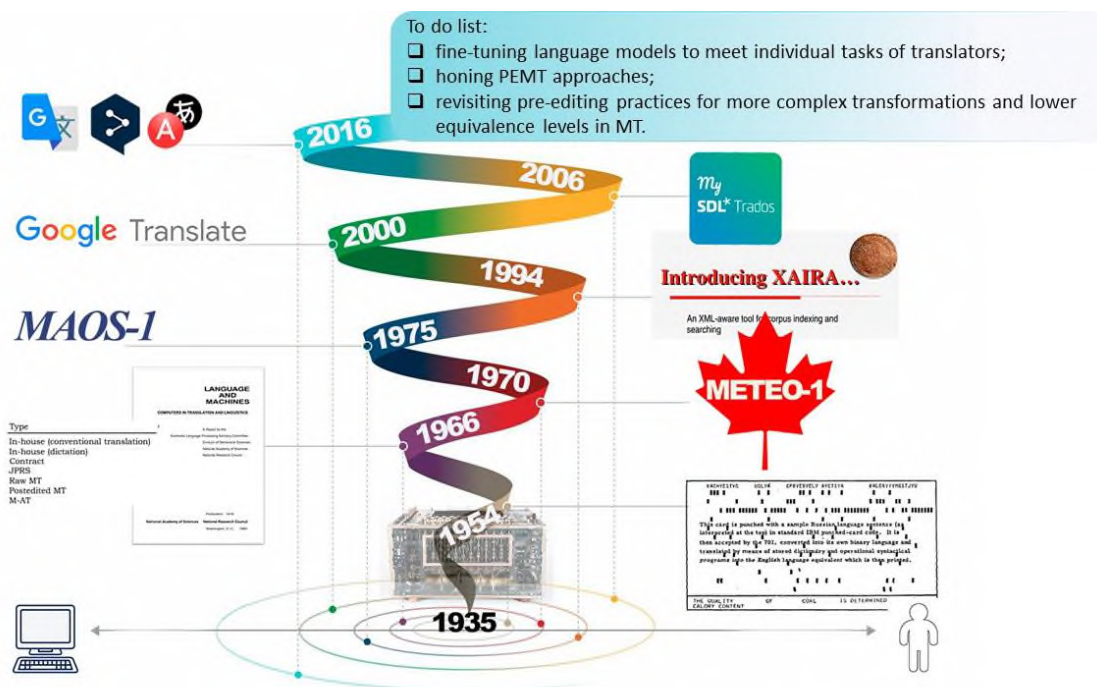
<sup>3</sup> Белгородский государственный национальный исследовательский университет  
ул. Победы, 85, Белгород, 308015, Россия  
E-mail: [lukyanova@bsu.edu.ru](mailto:lukyanova@bsu.edu.ru)  
ORCID: 0000-0002-1479-2227

*Статья поступила 31 марта 2024 г.; принята 15 июня 2024 г.;  
опубликована 30 июня 2024 г.*

**Аннотация.** Работа посвящена анализу ключевых проектов на всем треке развития машинного перевода (МП) и их роли в решение практических задач. На примере наиболее успешных инициатив установлено, что в период становления МП фактором, определяющим работоспособность систем, была энтропия текстов: чем меньше энтропия, тем выше предсказуемость, следовательно, выше эффективность технологии. Именно это объясняет успех Джорджтаунского эксперимента и проекта МЕТЕО-1, с помощью которой метеорологическая служба Канады четверть века обеспечивала автоматический перевод с английского языка на французский и обратно погодных сводок, благодаря высокой формализации текста. Доклад Консультативного комитета по автоматической обработке языка ALPAC, опубликованный в 1964 году, поставил под сомнение валидность полноценного машинного перевода, но не самой технологии, успехи которой отмечены в проектах при участии людей. Это направление, в частности, развилось в отдельную ветвь автоматизации

перевода, так называемый накопительный перевод. Авторы отмечают циклический характер развития методов и механизмов перевода с привлечением решений из области МП. Современная комбинация ресурсов и порядка их использования отличается немногим от представлений прошлого века. Разница заключается в возможностях технологий, которые проделали длинный путь от перевода, основанного на правилах и корпусе до современного статистического и нейронного перевода. Установлено, что для повышения качества последних требуется значительное увеличение объемов параллельных корпусов, должным образом обработанных, размеченных и подготовленных для автоматического обучения языковых моделей. С учетом прогресса и приобретенных преимуществ МП авторы моделируют базовые ситуации обработки текстов, в том числе с использованием ИИ. По результатам практической апробации различных методов работы с МП предложены дальнейшие пути совершенствования технологии в интересах профессиональных переводчиков, в том числе через самостоятельное обучение языковых моделей, развитие практики постредактирования (РЕМТ) и прередактирования для подготовки почвы для более глубоких трансформаций и достижения низких уровней эквивалентности. Массивы продолжают оставаться в центре внимания при обучении современных технологий МП. В этой связи выделяется два основных направления: совершенствование подготовки корпусов для восприятия компьютером посредством автоматизации основных типов разметки и повышение обеспеченности низкоресурсных языков.

**Ключевые слова:** Машинный перевод; Разметка; Аннотация; Качество МП; Энтропия; Постредактирование; Автоматизированный перевод



**Информация для цитирования:** Дёмочкина В. В., Груздев Д. Ю., Лукьянова Е. В. Машинный перевод вчера и сегодня // Научный результат. Вопросы теоретической и прикладной лингвистики. 2024. Т. 10. № 2. С. 21-45. DOI: 10.18413/2313-8912-2024-10-2-0-2

## **Introduction**

From the first practical attempts of P. Troyansky and G. Artsrouni to automate translation, machine translation (MT) endeavors have never disappeared from the agenda of computational linguistics for a single day (Hutchins, 2004a). Prior to the era of personal computers, efforts were redirected to other fields for long periods due to the technical limitations of computers. Today, MT is one of the fastest growing areas in the field of computational linguistics and artificial intelligence (AI). With the advent of modern technologies based on neural networks and deep learning, accuracy, quality, and alternatives in MT providers have increased exponentially. For the average consumer, knowledge in this area is limited to the last dozen years. It was in this period that it became feasible to take the human out of the loop when it came to familiarizing with a text in a foreign language. Meanwhile, back in the days when computer linguistics was just taking shape, there were many milestones and working solutions, considerably ahead of their time and the state of computer technology. We believe that some of them have retained their relevance to this day, thus it is reasonable to consider the possibility of bringing them to maturity, prompted by advanced hardware. Given this, the paper will analyze the most prominent projects in the MT field to determine their applicability and viability in today's environment. The goal calls for a phased approach covering the following tasks:

1. Shortlisting milestone projects.
2. Analyzing their application.
3. Considering pros and cons of each project.
4. Determining the primary reason for their success.
5. Testing the findings in modern hardware/software combinations.

## **Background**

The history of MT is an endeavor to develop a fully-automatic general-purpose high-quality machine translation system (FGH-MT), which still remains elusive not least because of the creative nature of translation. However, this turns out to be relative, for “MT is completely possible in some fields like science, technology, law, and medicine (Fumani and Reza, 2007).” On the other hand, “achieving paraphrase level or dynamic equivalence between the source and target language still appears to be a far-fetched dream (Oladosu et al., 2016).” This will demand compound transformations, often resulting in a lower level of equivalence, which is the case in literary translation. Replete with inferences, such texts prove to be tough for MT. (López et al., 2010) sums it up: “The more general the domain or complex the style of the text, the more difficult it is to achieve a high-quality translation.”

The MT quality varies and depends among other things on the type of software, which breaks down into two main categories: single and hybrid approaches (Oladosu et al., 2016). The former is based on one method, rule-based, direct, corpus-based, and knowledge-based approaches to MT, while the other is a combination of the statistical method (SMT) and the rule-based approach, which includes a word-based model, a phrase-based model, a syntax-based model, and forest-based model (Gashaw and Shashirekha, 2019). Most single approaches demonstrate poor performance in large scale applications and produce a “shallower representation of knowledge resulting in lower quality (Gashaw and Shashirekha, 2019).”

In the 1990s, in the spot light of MT researches were focused on SMT projects, which sprouted from the Candide Project at IBM in the late 1980s (Brown et al., 1993).

The model also marked the departure from word-based to phrase-based MT.

25 years and numerous efforts later, the MT community arrived at neural MT (NMT). This is the latest breakthrough, which received a wide-spread recognition. The quality of general translation sky-rocketed, as evidenced in accounts of the use of Google, DeepL, and Yandex (O'Brien et al., 2014). It does provide for basic translation in language pairs that do not have parallel corpus in a process called “zero-shot translation” (Costa-jussa et al., 2017).

This said, NMT has many challenges, such as “domain mismatch, size of training data, rare words, long sentences, word alignment, and beam search (Gashaw and Shashirekha, 2019).” The new model still relies on parallel corpora with the-more-the-better being still fundamental for its training, therefor ad-hoc domains remain beyond its grasp (Hurskainen, 2018).

Another significant hurdle is that the data set needs to be prepared for processing by language models. Mostly it boils down to

attaching labels “showing the role of each word in a sentence” (Hurskainen, 2018). Done through mark-up, tagging, and annotation, the process still relies on manual disambiguation.

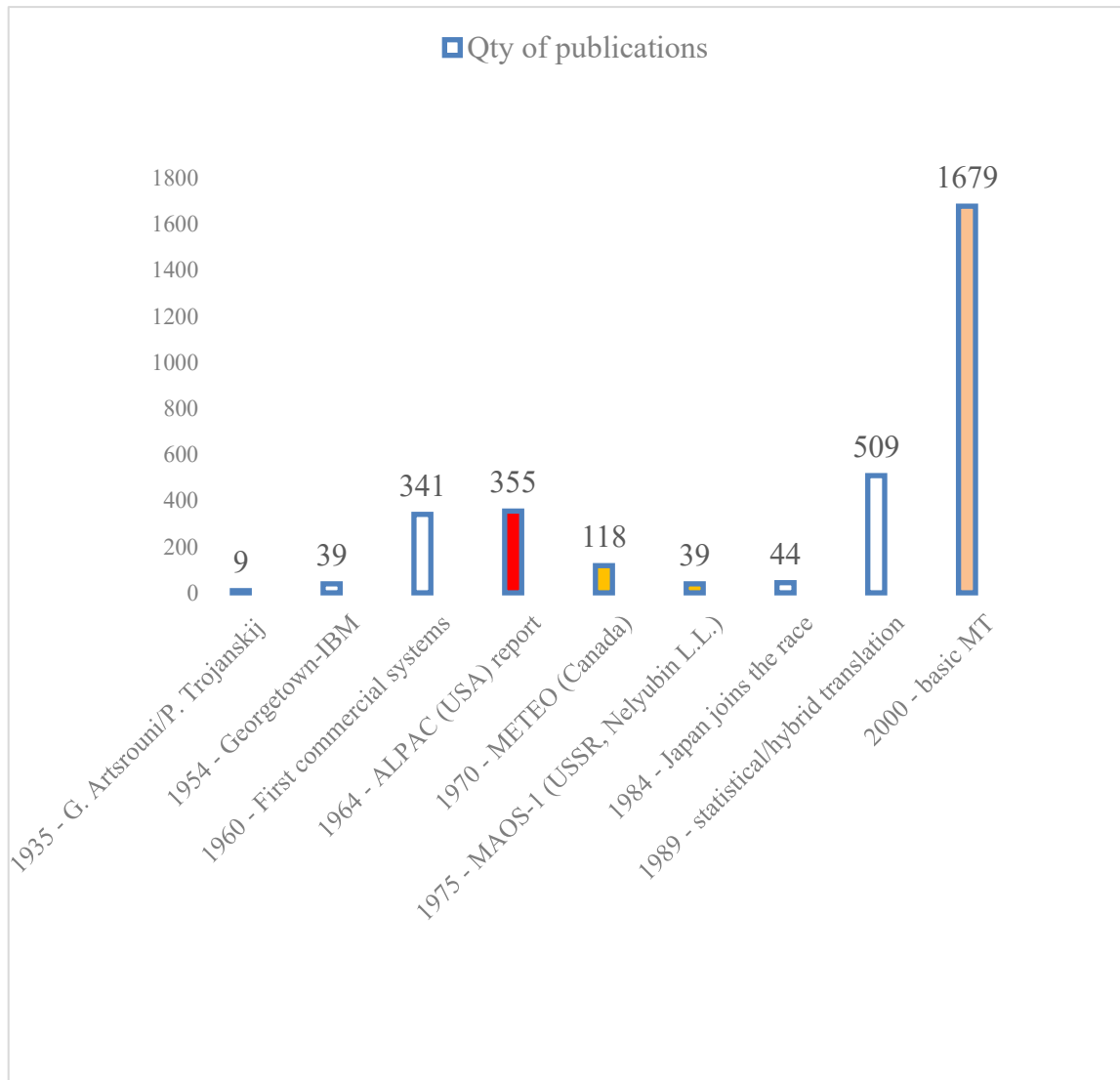
From the application perspective, machine translation remains a human-in-the-loop technology as well. True, the computer now shoulders a larger share of the burden, but the final touch still belongs to a human translator (Bharati et al., 2003). This collaboration transpires today through pre- and post-editing and computer-assisted solutions (O'Brien et al., 2014).

### **Materials and methods**

Machine translation has been on the agenda of linguists for more than half a century. Its path has been cloudless by no means. The roller-coaster trajectory, common to any research effort, was also conditioned by the limitations of the nascent computer technology, which made it impossible to fully implement the concept at certain stages (see Figure 1).

Figure 1. Density of MT researches<sup>1</sup>

Рисунок 1. Плотность исследований в области машинного перевода



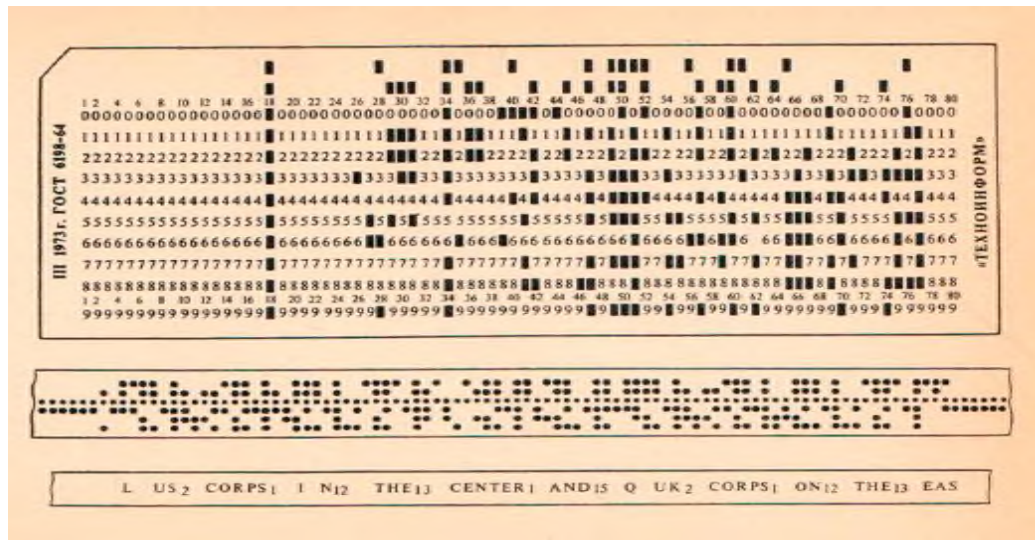
Not only did the computer's own ability to convert information to other languages seriously limit the research potential, but the ways of inputting raw data as well. In the first 30-40 years, scientists had to make do with punch cards, which took some time to prepare (see Figure 2) (Nelyubin, 1983).

Despite the labor-intensive process of preparing the material, MT visionaries continued to move towards their goals. The first results were a real breakthrough for their time. For example, the Georgetown-IBM experiment, a joint effort between the

eponymous university and IBM in 1949-1954, managed to translate about 60 sentences with 60% accuracy. The project was based on the use of grammar rules and dictionary databases to translate from Russian into English (Daems et al., 2017). Short sentences were chosen for demonstration in line with to the level of development of the system. Another thing that stands out, however, is the area from which the text samples were taken. Most of the simple statements were about processes in organic chemistry (see Table 1) (Hutchins, 2004b).

<sup>1</sup>Machine Translation Archive, available at <https://aclanthology.org/www.mt-archive.info/srch/authors.htm> (as of March 2017).

**Figure 2.** Information on punch card, punch tape and control tape  
**Рисунок 2.** Информация на перфокарте, перфоленте и контрольной ленте



**Table 1.** Sentences, selected for the demonstration of the results of the Georgetown-IBM experiment (Ornstein, 1955)

**Таблица 1.** Предложения для демонстрации результатов Джорджтаунского эксперимента

№ п/п	Russian	English
1.	Качество угля определяется калорийностью.	The quality of coal is determined by calory content.
2.	Крахмал вырабатывается механическим путем из картофеля.	Starch is produced by mechanical methods from potatoes.
3.	Железо добывается из руды химическим процессом.	Iron is obtained from ore by chemical process.
4.	Динамит готовится химическим процессом из нитроглицерина с примесью инертных соединений.	Dynamite is prepared by chemical process from nitroglycerine with admixture of inert compounds.
5.	Обработка улучшает качество нефти.	Processing improves the quality of crude oil.
6.	Международное взаимопонимание является важным фактором в решении политических вопросов.	International understanding constitutes an important factor in decision of political questions.

For the purpose of demonstration, general examples were selected as well. Though there were few of them so that it is not possible to draw unambiguous conclusions as to the quality of translation in comparison with technical samples. We can only note some lexical errors. For example, in line 6 of Table 1, *solving* and *issues* should be used instead of *decisions* and *questions*, respectively. The program coped better with the organic chemistry sentences. Thanks to

the terminology, the developers eliminated lexical ambiguity. The direction of translation was also beneficial: the English language, into which the translation was performed, is characterized by a polysemantic vocabulary, which reduced the variability of translation of individual lexical units.

Fifteen years later, still with technology lagging behind MT ideas, a second milestone project, dubbed METEO, was successfully implemented to translate weather bulletins

from English into French and back (Nirenburg, 1993). In contrast to the American experiment, the Canadian system evolved to a full-fledged system, translating 80,000 words of weather reports per day until 2001.

At the same time, Soviet and Russian scientist L. L. Nelyubin carried out a similar project, called MAOS-1/2/3 (*Machinny Avtomatizirovanny Otrasleyov Slovar*, Russian for Automated Ad-hoc Dictionary) (Nelyubin, 1975). The system provided translation of NATO military documents from English into Russian. The Canadian and

Soviet experience was in many ways similar to each other and to the Georgetown-IBM experiment. The three milestones were "doomed" to success due to the lack of ambitions to create a one-stop shop in MT: the American team could not think of this, being pioneers in the field, while their colleagues from Canada and the USSR chose narrow fields to see their projects mature to useful MT tools. The second important factor was the choice of relatively limited sublanguages designed to convey messages in template documents (see Table 2).

**Table 2.** Sample weather bulletin on the official Canadian government website and the U.S. Army Air Forces Combat Order template

**Таблица 2.** Образцы погодной сводки на официальном сайте правительства Канады и шаблон боевого приказа военно-воздушного корпуса СВ США

Weather bulletin <sup>2</sup>	U.S. Army Air Forces Combat Order Template <sup>3</sup>
<p>FPCN11 CWUL 191945                      Forecasts for Western Quebec issued by Environment Canada at 3:45 p.m. EDT Tuesday 19 March 2024 for tonight Wednesday and Wednesday night.                      The next scheduled forecast will be issued at 5:00 a.m. EDT Wednesday.</p> <p>Metro Montréal - Laval.</p> <p>Tonight..Mainly cloudy. Periods of light snow beginning this evening. Wind west 20 km/h becoming light this evening. Low minus 1. Wind chill near minus 6.</p> <p>Wednesday..Periods of light snow ending in the morning then mainly cloudy. A few rain showers beginning in the afternoon. Wind becoming southwest 20 km/h gusting to 40 in the morning. High plus 5. UV index 3 or moderate.</p> <p>Wednesday night..A few rain showers or flurries ending in the evening then mainly cloudy. Wind west 30 km/h gusting to 50. Low minus 7. Wind chill minus 15 overnight.</p> <p>Vaudreuil - Soulanges - Huntingdon.</p> <p>Tonight..Mainly cloudy. Periods of light snow beginning this evening. Wind southwest 20 km/h becoming light this evening. Low minus 1. Wind chill near minus 6.</p> <p>Wednesday..Periods of light snow ending in the morning then mainly cloudy. A few rain showers beginning in the afternoon. Wind becoming southwest 20 km/h gusting to 40 in the morning. High plus 5. UV index 3 or moderate.</p> <p>Wednesday night..A few rain showers or flurries ending in the evening then mainly cloudy. Wind west 30 km/h gusting to 50. Low minus 9. Wind chill minus 16 overnight.</p>	<p>29. PREPARATION.—<i>a. Basic form.</i>—There are three main parts to the form: heading, body, and ending. (See fig. 10.)</p> <p>(1) <i>Heading.</i></p> <p style="padding-left: 40px;">Name of unit                      Place                      Date and time of issue</p> <p>Field order number                      Maps: Full description of map(s) used in operation.</p> <p>(2) <i>Body.</i></p> <ol style="list-style-type: none"> <li>1. a. (1) Ground information of the enemy.                      (2) Air information of the enemy.</li> <li>b. (1) Information of supporting or supported ground forces.                      (2) Information of supporting or supported air forces.</li> <li>2. The decision of the commander and so much of the plan and scheme of maneuver as applies to the entire command, including where applicable: route out, route back, assembly, unit initial point, axis of attack, method of attack, maneuver after attack, and rally. The questions as to <i>what, where, when, how, and why</i> should be answered.</li> <li>3. Detailed instructions to each of the next lower subordinate units of the command, using a <i>separate lettered subparagraph for each</i>. Include loadings—ammunition, bombs, and chemicals.                     <ol style="list-style-type: none"> <li>a.</li> <li>b.</li> <li>z. Instructions applicable to <i>two or more units</i>—may include where applicable: loadings, secondary objectives, rendering reports, operation of runways, and preparations for subsequent operations.</li> </ol> </li> <li>4. Instructions regarding supply and administration necessary for the operation, which may include supply, traffic, transportation, airdromes, personnel, and miscellaneous. If instructions are too voluminous, a separate administrative order may be used.</li> <li>5. Instructions for:                     <ol style="list-style-type: none"> <li>a. Signal communications—signal operation instructions (SOI) number, time effective, liaison frequencies.</li> <li>b. Axes of signal communications—supported ground forces.</li> <li>c. Command posts.                             <ol style="list-style-type: none"> <li>(1) Ground—state location, when necessary, of supporting or supported units.</li> <li>(2) Air—position of commander, designation of deputy.</li> </ol> </li> </ol> </li> </ol> <p>(3) <i>Ending.</i></p> <p style="text-align: right;">Signature (commanding officer, executive officer, or staff officer).</p> <p>Authentication                      "Official"                      Signature of S-3</p>

<sup>2</sup> Public Text Bulletins, retrieved from [https://www.weather.gc.ca/forecast/public\\_bulletins\\_e.html](https://www.weather.gc.ca/forecast/public_bulletins_e.html)

<sup>3</sup> FM 1-75 Army Air Forces Field Manual, Combat Orders. 1942-06-16.



This is accounted for by entropy, a measure of randomness in a situation (Weaver, 1949). First coined by Ludwig Boltzmann and J. Willard Gibbs in the statistical thermodynamics formulation in the 1870s, the term was later adopted in Information Theory by Claude Shannon and Ralph Hartley in the 1940s<sup>4</sup>.

In linguistics it determines the probability of the next word in a sentence. For example, in Table 3 predicates (2) preceded by the aviation engine type (1) in the role of a subject (1) dictate the use of an aircraft equipped with the engine in sentence 1 or the amount of thrust it can develop in sentence 2. The level of entropy gets higher at the beginning of a sentence and drops down with every next word. For the sake of comparison, the predicates (1) alone can be followed by any other word, including a preposition. The phenomenon is at the core of the sentence builder ESL activity, where the probability of the word sequence is of the essence.

Due to the complex nature of the task at hand, luck did not always favor scientists. There were protracted periods of search, which sometimes led to conclusions that further research was not worthwhile. The Georgetown-IBM experiment was followed by a decade-long lull that led to the negative recommendations by the U.S. Advisory Committee on Automatic Language Processing (ALPAC) in 1966. The report rated machine translation unsatisfactory in three areas: cost, quality and time. Experimentally it was found that familiarization with a research article on physics, translated by a machine, required 21% more time, while the degree of assimilation of the material fell by 29%. At the same time, the accuracy of information perception dropped by 10%. As a result of post-editing the indicators were improved to

11%, 13% and 3%, respectively.<sup>5</sup> Bringing the MT output to a satisfactory one required additional time, which sometimes reached several weeks.<sup>6</sup> In terms of cost-effectiveness, machines did not do better than humans, either. For example, a translation, contracted to a third party, was \$50 cheaper with quality indicators approaching *good* compared to satisfactory MT outputs with post-editing (see Figure 3).

The reason for such outright criticism of the fledgling MT was accounted for by the inflated expectations. In contrast to the successful METEO and MAOS, the U.S. Air Force's Department of Foreign Technology and the National Atomic Energy Laboratory in Oak Ridge, USA, set out to fully automate the translation of research texts, which, although rich in terminology, were still highly entropic due to the use of an unlimited sublanguage. Standing out in this report is MAT (machine-assisted translation), in which a machine was assigned the task of vocabulary management, but the translation was still performed by a human. The results of this approach were rated *excellent* with a 25% reduction in cost relative to conventional translation.

Negative feedbacks did not kill the MT domain, but rather prompted dips in research intensity. In the era of conventional media, it took some time for information to spread, in average 10 years for critical decisions to take its toll. For example, even the success in Canada, which had international publicity, did not bring research intensity out of the downward trend triggered by the ALPAC report (see Figure 1). The lack of a complete break in the line of MT researches is accounted for by the gradual involvement of other states, supporting their national languages. For example, Japan entered the race the lowest point of the MT popularity.

<sup>4</sup> Entropy in thermodynamics and information theory, retrieved from [https://en.wikipedia.org/wiki/Entropy\\_in\\_thermodynamics\\_and\\_information\\_theory](https://en.wikipedia.org/wiki/Entropy_in_thermodynamics_and_information_theory)

<sup>5</sup> ALPAC (1966) Languages and machines: computers in translation and linguistics, 34.

<sup>6</sup> ALPAC (1966) Languages and machines: computers in translation and linguistics, 21.

**Table 3.** Entropy in natural languages  
**Таблица 3.** Энтропия в естественном языке

1.	The (1) <u>Pratt &amp; Whitney GTF™</u> (2) <u>powers</u> the (3) <u>Airbus A320neo</u> . <sup>7</sup>
2.	The (1) <u>PW1000G</u> (2) <u>generates</u> (3) <u>15,000 to 33,000 lbf</u> . <sup>8</sup>

**Figure 3.** Comparative evaluation of the cost of human, machine and automated translation<sup>9</sup>  
**Рисунок 3.** Сравнительная оценка стоимости перевода, выполненного человеком, машиной и в автоматизированном режиме

ESTIMATES OF COSTS AND  
 QUALITY FOR VARIOUS TYPES  
 OF TRANSLATION

Type	Quality	Cost for 11,000,000 Russian Words
In-house (conventional translation)	Good	\$ 440,000
In-house (dictation)	Good	440,000-
Contract	Fair to good	350,000
JPRS	Fair	240,000
Raw MT	Unsatisfactory	80,000 +
Postedited MT	Fair	400,000
M-AT	Excellent	310,000
Analysts proficient in Russian	-	0

**Results and discussions**

From the very first attempt at practical implementation of MT, scientists and developers understood the centrality of the data sets needed to train systems. Therefore, the initial task was to work on quantitative indicators, which presented a certain challenge in the conditions of immature text recognition technologies. However, the analysis of the first complex MT systems pointed to the need to enhance the computer's perception of natural language. The basic perception of text in the form of groups of symbols did not provide much for text transformations, something to consider when

it comes to achieving adequate translation.

In order to track the MT research trajectory in hindsight, it was decided to upgrade the MT research repository collected by John Hutchins into an adhoc corpus<sup>10</sup>. Using the AntConc<sup>11</sup> corpusmanager, a frequency list of the corpus of 1,058,811 tokens (18,800 words) and concordances for a group of selected key words were compiled (see Table 4). To extend the coverage of queries, all words were reduced to their base with the wildcard \* replacing the endings to open options for wider variations (Gruzdev and Kodzhebash, 2023).

<sup>7</sup> GTF engine: Powering the Airbus A320neo, Airbus A220, and the Embraer E-Jets, retrieved from <https://www.mtu.de/engines/commercial-aircraft-engines/narrowbody-and-regional-jets/gtf-engine-family/>

<sup>8</sup> Pratt & Whitney PW1000G, retrieved from [https://en.wikipedia.org/wiki/Pratt\\_%26\\_Whitney\\_PW1000G](https://en.wikipedia.org/wiki/Pratt_%26_Whitney_PW1000G)

<sup>9</sup> ALPAC (1966) Languages and machines: computers in translation and linguistics, 80.

<sup>10</sup> Machine Translation Archive, available at <https://aclanthology.org/www.mt-archive.info/srch/authors.htm> (as of March 2017).

<sup>11</sup> Anthony, L. (2020). AntConc (3.5.9) [Computer Software]. Tokyo, Japan: Waseda University, retrieved from <https://www.laurenceanthony.net/software> (Accessed 15 February 2022).

**Table 4.** Queries based on selected words (green – corpus development, yellow – corpus quality improvement, orange – corpus employment)

**Таблица 4.** Запросы на основе отобранных лексических единиц из частотного списка (зеленый – разработка корпуса, желтый – совершенствование качественного состава массивов, оранжевый – использование корпуса)

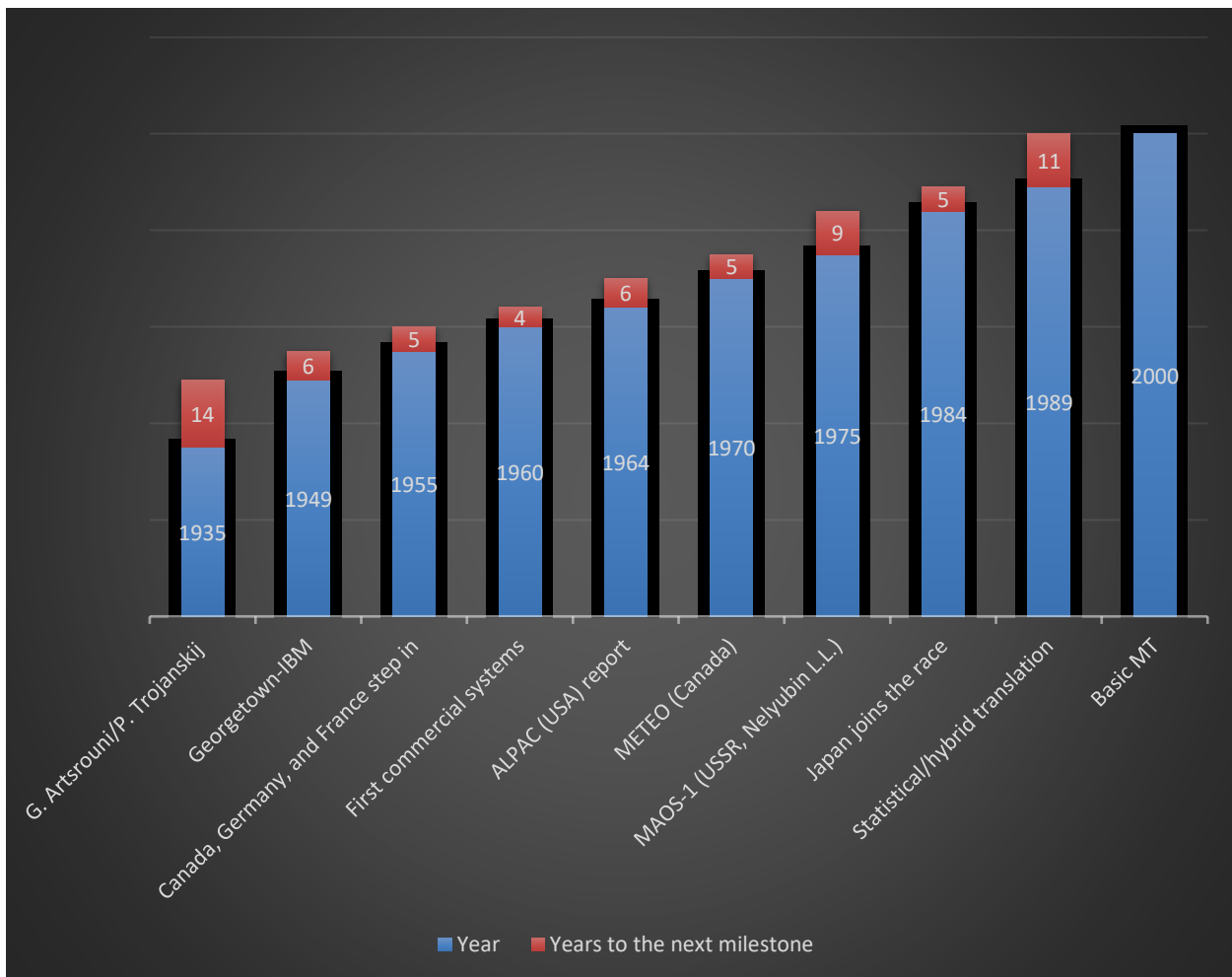
Group	Query	Seq. number in the list	Frequency
3	Acquisition	835	152
1	Align*	857	147
2	Ambigu*	1,827	63
2	Annotat*	1,657	71
1	Corpus	118	932
2	Disambigu*	427	303
2	Discrimin*	515	255
1	Match*	954	130
3	Minin*	571	232
2	Pars*	4,465	20
3	Retriev*	172	688
2	Segmentat*	665	198
2	Tag*	1,533	79

In total, three complex queries were generated: (1) corpus compilation, (2) corpus annotation, and (3) linguistic information mining. The concordances were further subjected to additional processing in MS Excel to calculate the density of word occurrences every five years, the intervals being established on the basis of the average gap between significant events over the entire research trajectory (see Figure 4). Because the corpus was compiled from the publishing data of research papers, each line contained the

publishing date. We took advantage of this feature of the concordances and applied the function COUNTIF to the entire range. For example, =COUNTIF(A1:A15972;"=\*1940\*") counts the quantity of all papers published in 1940.

The results of processing the three concordances against the background of the aggregate total of MT papers per each interval have been summarized in Figure 5 for further analysis.

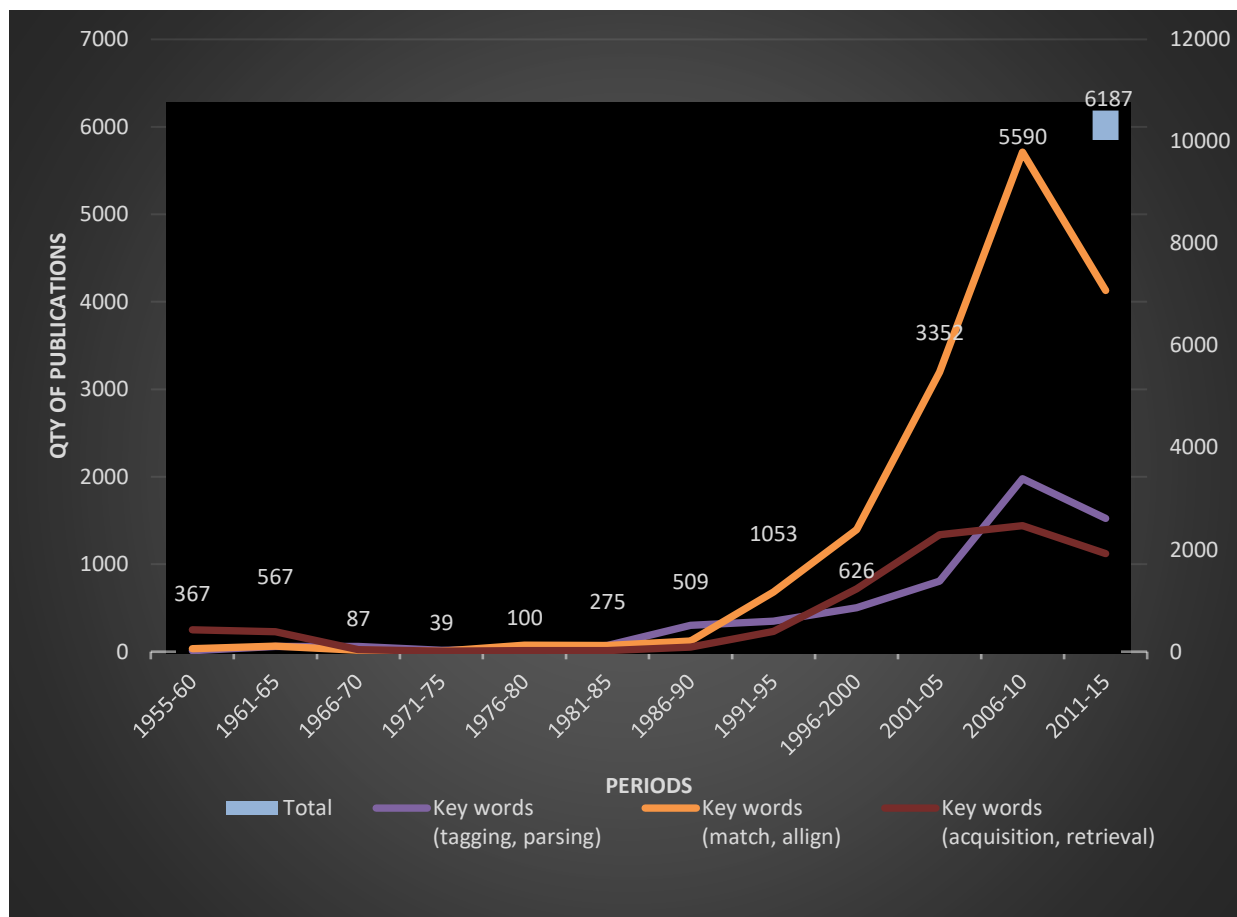
**Figure 4.** Milestones in the development of machine translation  
**Рисунок 4.** Вехи в развитии машинного перевода



**Figure 5.** Nature of MT research in different periods  
**Рисунок 5.** Характер исследований в области МП в разные периоды

After the first corpora were created, scientists moved on to practical steps to extract the necessary linguistic information. In 1966, it became clear that the corpora still needed to be prepared to expand the horizon of applied MT researches. As a result, data mining improvement was excluded from the MT agenda for the next 20 years. Throughout the period, the corpus did not undergo major changes, being basically a large database of textual information selected according to

certain criteria. By 1990, researches already had developed an idea of how corpora could be used, and computers were taught to “understand” texts. This was a watershed moment, marking the beginning of the modern phase of MT development. All three paths scrutinized in the three concordances got an impetus: development of new types of corpora, improvement of information retrieval methods and teaching computers to understand all levels of language.



By this time, a solution to bring the perception of texts by a computer closer to human abilities had taken shape. It was based on annotation and markup and encompassed the main language levels, namely: phonetic, prosodic, orthographic, semantic, morphological, syntactic, and anaphoric (Gruzdev and Kodzhebash, 2023). At the initial stage, all manipulations were done manually, but by the end of the century it had been realized that automation of the process was needed to handle multimillion-token corpora. However, by 2010 the problem had not been completely addressed with only POS

(part of speech) reaching accuracy of 95% in the automatic mode (Névolet al., 2010).

It is the markup and annotation that allows the silicon brain to correctly transfer information from one language to another. Without it, the computer has a two-dimensional picture of the text. The computer can fathom the language levels only after the information about the role of each lexical unit has been integrated into the data set. Table 5 shows an example of a sentence translation by the SMT version of Google Translate at the dawn of annotation.<sup>12</sup>

<sup>12</sup> Google Translate Translation methodology. Available from [https://wiki.edunitas.com/IT/114-10/Google-Translator\\_2167\\_eduNitas.html](https://wiki.edunitas.com/IT/114-10/Google-Translator_2167_eduNitas.html)

**Table 5.** Translation done by SMT Google Translate

**Таблица 5.** Перевод, выполненный версией Google Translate, основанной на статистическом подходе

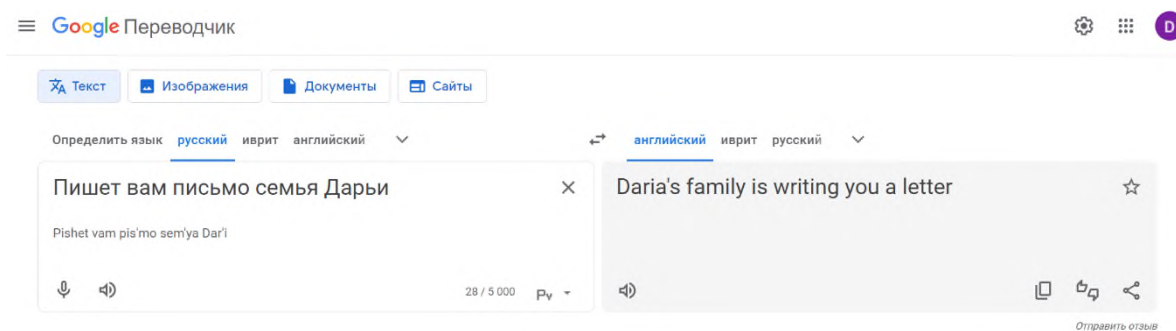
Original	Translation
Пишет вам письмо семья Дарьи.	You wrote a letter to family Darya

The author of the article, covering the example, attributed the distortion in translation to the inherent English rule of preserving the direct word order and ignorance of the grammar of the source language. Expanding on the latter, we would

like to draw the attention to the lack of developed markup, primarily syntactic and anaphoric, in the SMT version of the online translator, which would let the computer correctly identify sentence members and match pronouns.

**Figure 6.** A translation of the sentence «Пишет вам письмо семья Дарьи» in the modern version of Google Translate

**Рисунок 6.** Пример перевода предложения «Пишет вам письмо семья Дарьи» в современной версии Google Translate



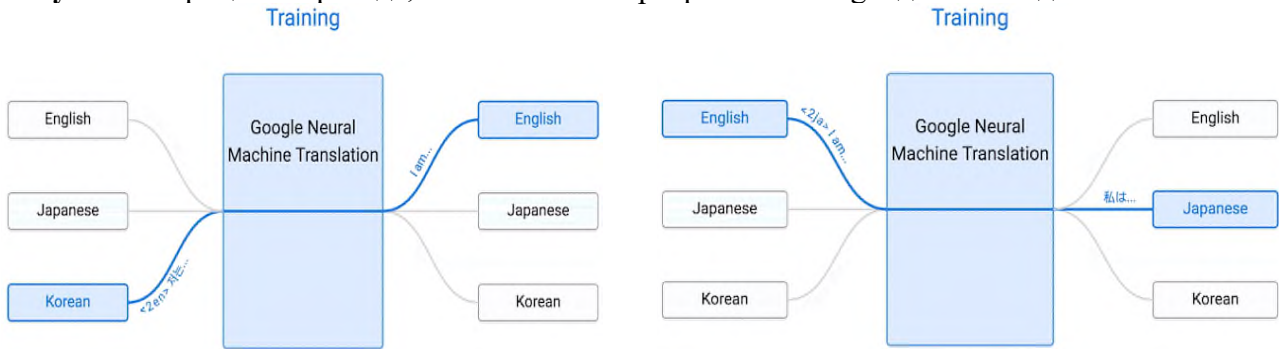
When Google switched to NMT (Neural Machine Translation), such errors were finally addressed (see Figure 6). The new model lifted the role of an intermediary language from English. Thus, the model that translates from English to Korean and from English to Italian can also communicate text from Italian to Korean without any substantial data in this language combination, i.e. perform zero-shot translation (see Figures 7, 8) (Schuster et al., 2016; Quach, 2016; Suxia Lei and You Li, 2023).

The share of MT studies for language pairs involving English is still predominant, but most of the efforts date back to 2015 (see Figure 9). Despite the introduction of the new model that can translate from scratch in any language pair, the focus has shifted to rare languages, most of them being low-resource languages. The languages lack annotated texts, speech data and other linguistic resources needed for MT development and training.<sup>13</sup> These are essential for adjusting language models generated by AI.

<sup>13</sup> POEditor Low-resource languages: A localization challenge, retrieved from <https://poeditor.com/blog/low-resource-languages/>

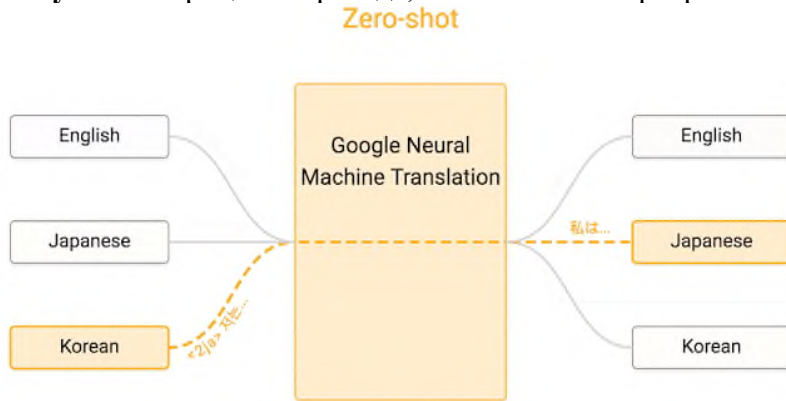
**Figure 7.** Translation process performed by Google through 2016

**Рисунок 7.** Процесс перевода, выполняемый программой Google до 2016 года



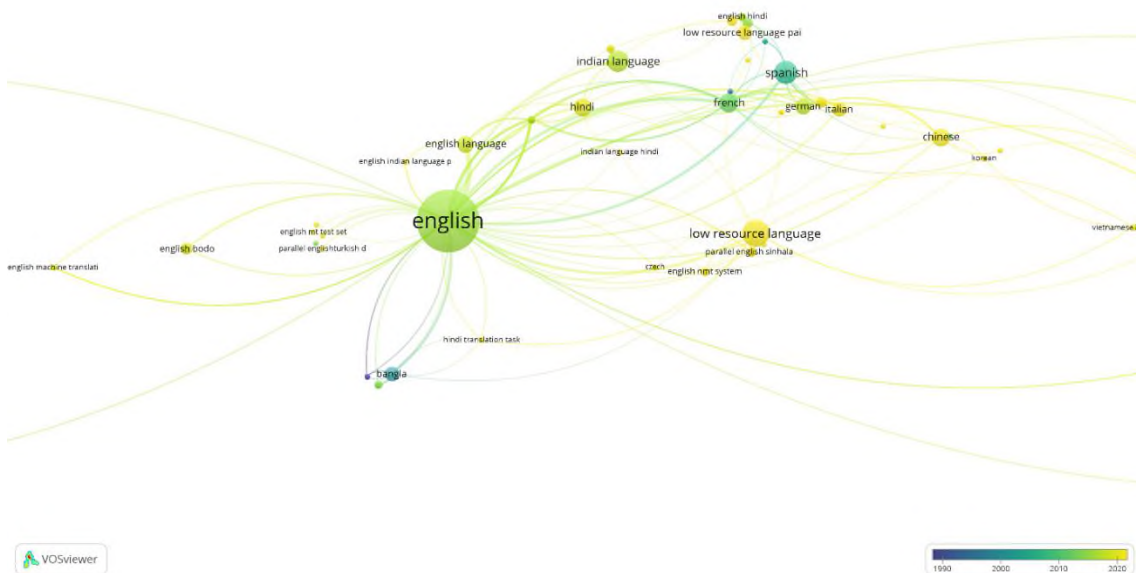
**Figure 8.** Translation process performed by Google after 2016

**Рисунок 8.** Процесс перевода, выполняемый программой Google, после 2016 года



**Figure 9.** Language focus of MT researches<sup>14</sup>

**Рисунок 9.** Языковая ориентация исследований МП



<sup>14</sup> Based on a Lens selection of 1094 MT papers covering a period of 1990-2015, registering the most intensive MT development spike (key words: MT, machine translation, annotation, tagging, NLP, natural language processing, parsing, corpora, corpus, parallel corpus, comparable corpus, rule-based, neural MT, NMT).

The fact that AI was brought into the task of corpus processing and preparation has not completely exhausted the issue of automating semantic markup. The current solution is based on a multiple vector embedding approach, which already allows the hardware to "grasp" the meaning, though has a long way to go (Tripathi and Kansal, 2020). This explains the low efficiency of MT when working on fiction and poetry, where the aesthetic impact becomes paramount in text pragmatics. In order to achieve this, the most complex transformations are of the essence. However, these linguistic manipulations are contingent on understanding of the underlying meaning, exactly what we observe in Table 6, containing two translations of a fragment of Sonnet No. 130 by W. Shakespeare:

My mistress' eyes are nothing like the sun;

Coral is far more red than her lips' red;

If snow be white, why then her breasts are dun;

If hairs be wires, black wires grow on her head.

Simple generalization (1) and omission (4) were lost on Google Translate, let alone complex transformations – explication (2) and explication (3). Basic features of poetry –

rhyme and meter – are missing in the MT output.

This said, the MT progress is more tangible in the field of technical translation. Thanks to the absolute priority given to conveying the meaning contained in research texts, MT technologies are used as a non-alternative means of translation into major foreign languages without post-editing by humans (see Figure 10).

Major corporations are among those, who favor the approach. However, this is a desperate measure to catch up with the ever-growing information flow, no human can ever meet without getting into overheads. It is noteworthy that there is no distortion in the fragment, and translation inaccuracies do not interfere with the main message.

In contrast to the poor experience of the U.S. Air Force Foreign Equipment Department with MT, highlighted in the 1966 ALPAC report, modern systems bring the performance to a satisfactory-to-good level without post-editing. Nevertheless, MT has not eclipsed their human counterparts, and the same translation techniques as in the U.S. report (see Figure 2) are still relevant today, but at a higher technological development level: conventional translation, STT-powered dictation translation, CAT (Computer-Assisted Translation), and PEMT (post-editing machine translation).

**Table 6.** Translations of Sonnet 130 by W. Shakespeare, by DeepL and S. Y. Marshak

**Таблица 6.** Переводы сонета 130 У. Шекспира, выполненные DeepL и С. Я. Маршаком

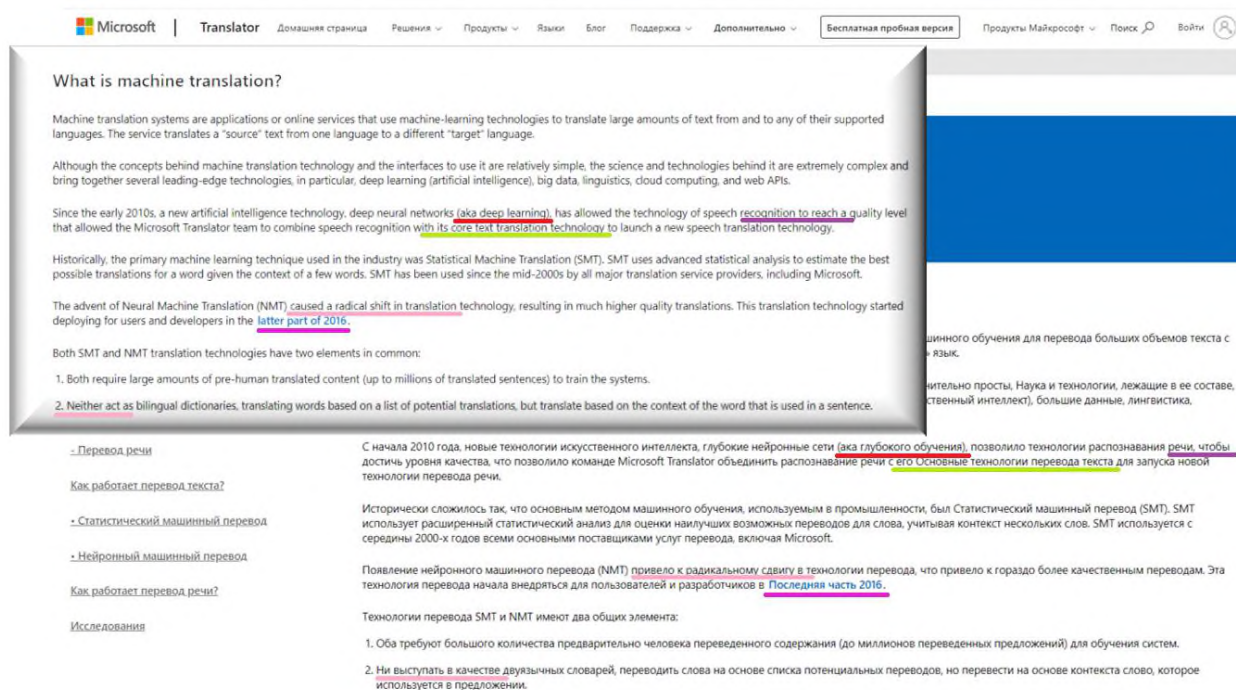
DeepL	S. Marshak <sup>15</sup>
Глаза моей госпожи совсем не похожи <b>(1) на солнце;</b> Коралл куда <b>(2) краснее</b> , чем губы ее; Если снег бел, <b>(3) то грудь ее темна;</b> <b>(4) Если волосы – проволока</b> , то черная проволока растет на ее голове.	Ее глаза на <b>(1) звезды</b> не похожи <b>(2) Нельзя</b> уста кораллами <b>назвать</b> , <b>(3) Не белоснежна</b> плеч открытых кожа, И черной проволокой вьется прядь.

<sup>15</sup> Shakespeare, W. Sonnet 130, rendered into Russian by S. Marshak, retrieved from [https://engshop.ru/shekspir-sonet-130-na\\_anglieskom/](https://engshop.ru/shekspir-sonet-130-na_anglieskom/)



Figure 10. MT-powered multilingual service on the Microsoft website

Рисунок 10. Мультиязычное сопровождение сайта корпорации «Майкрософт» на основе машинного перевода



The practice of dictating a translation and then converting it into a text has gained new perspectives with the advent of specialized software replacing yet another player in the process, a typist. Today, the translator is no longer dependent on this additional link in the translation chain.

Automation became a plausible alternative while modern MT systems were in their infancy. This path led to the development of CAT tools, a.k.a. Translation Memory (TM). As MT reached an acceptable quality level, the two systems merged creating a cumulative effect (Quintana and Castilho, 2022). For example, Trados, one of the CAT leaders, now could outsource translation of fragments missing in the TM. As a result, all texts, regardless of their entropy, could be translated at a certain level of automation, turning the software into a true virtual translator's workbench. Tantamount to the recognition of MT by translators, it was yet another way of putting MT into practice.

Post-editing of MT is still on the agenda and forms another alternative to the use of

new technologies in professional activities. However, under today's conditions, this approach looks more cost-effective. Given the significant improvement in the quality of MT, editing time is measured in days rather than weeks. This path led to the establishment a separate practice, known as PEMT (Post-Editing of Machine Translation).

The need for additional manipulations of the MT output is dictated by additional pragmatic tasks. For example, non-periodical publications pursue not only to spread information, but preserve and popularize the printed heritage as well, while ad-materials should emphasize the advantages of the product in an unambiguous way.

The machine is not yet capable of accomplishing these tasks due to the immature automated semantic markup, preventing MT from performing critical transformations, a part and parcel of adequate translation (see Table 7) (Volkart and Bouillon, 2022).

**Table 7.** Comparison of human and MT translations of the text S-400 Air Defense Missile System  
**Таблица 7.** Сопоставление переводов фрагмента текста «Зенитная ракетная система С-400», выполненных человеком и системой МП

Original	DeepL	Translation
<p>Зенитная ракетная система С-400</p> <p>На первом этапе основным преимуществом С-400 над системами предыдущих модификаций стала более высокая степень автоматизации боевой работы, использование (1) <i>современной элементной базы</i>, возможность интеграции в различные уровни управления не только ВВС, но и других (2) <i>видов вооруженных сил</i>, а также увеличение числа (3) <i>одновременно сопровождаемых</i> и <i>обстреливаемых целей</i>. Хотя в 2007 году официально было заявлено, что дальняя граница зоны поражения ЗРС С-400 может достигать 400 км, до недавнего времени в боекомплект входили только зенитные управляемые ракеты семейства 48Н6, поступившие на вооружение в начале 1990-х вместе с (4) <i>ЗРС С-300ПМ</i>. Максимальная дальность поражения крупных аэродинамических целей (4.1) <i>ЗУР 48Н6Е3</i> на средних высотах составляет 250 км.</p>	<p>S-400 air defense missile system</p> <p>At the first stage, the main advantage of the S-400 over the systems of previous modifications was a higher degree of combat automation, the use of modern (1) <i>element base</i>, the ability to integrate into various levels of control not only of the Air Force, but also of other (2) <i>branches</i> of the armed forces, as well as an increase in the number of (3) <i>simultaneously escorted and fired targets</i>. Although in 2007 it was officially announced that the S-400 SAM's long-range kill zone can reach 400 kilometers, until recently its ammunition included only 48N6 family anti-aircraft guided missiles, which entered service in the early 1990s together with the (4) <i>S-300PM SAM</i>. (4.1) <i>The 48N6E3 SAMs</i> have a maximum range of 250 km at medium altitudes against large aerodynamic targets.</p>	<p>S-400 SAM</p> <p>By the time the S-400 achieved an IOC its major advantages over predecessors were a far superior automated performance, (1) <i>advanced hardware</i>, increased (3) <i>simultaneous engagement capability</i>, and provisions for integration into various command echelons of not only the AF but also other (2) <i>services</i>. Though it was stated in 2007 that the system boasted a maximum engagement range of 400km, until recently, its ammunition load comprised obsolete (4.0) <i>48N6 SAMs</i>, phased into service along with the (4) <i>S-300PM</i> back in the early 1990s. Thus, the (4.1) <i>48N6E3</i> can reach large mid-altitude threats at no more than 250km.</p>

At the present stage, restricting MT to a bounded sublanguage is less relevant for improving the end quality, since the technology is now developed as a general-purpose system. Therefore, the issues of lexical polysemy and terminology have not been exhausted, which can be traced in fragments 1-3 in Table 7. For example, the term “*сопровождать*” in the MT version is translated *to escort* (provide protection, move alongside), whereas in military discourse it means *to keep in sight (to track)*. However, the verb in question is part of a broader concept, for which there is an equivalent in English *to have multiple engagement capability*. Access to this lexical unit would improve the quality of the translation text Vs the original, however, the only way to implement the solution in the translation is by adopting an explication.

The given examples of inaccuracies will only make it difficult to grasp the information in a fast manner, while fragment 4 may well be misleading. According to DeepL’s translation both the S-300PM and the 48N6E3 are missiles (SAM stands for Surface-to-Air missiles), not a system and missile, respectively. Due to the MT limitations in performing semantic analysis, we assume that one efficient way to circumvent the problem is to go for text pre-editing. In the early days

of MT, this was a common practice for reducing text to a simple, computer-legible form (Tomasello, 2020). Like post-editing, all preparatory manipulations with the text were performed manually, which led to the increase in time consumed by the task, rendering the only at the time MT advantages negligible at best. Given the growing popularity of chatbots in solving complex tasks without human’s intervention, we went for testing the AI capabilities in simplifying the source text for subsequent machine translation (see Figure 11).

Compared to the original, the edited text became more compact, all complex sentences were reduced to a number of simple ones (see Table 8), and fragments 1 and 3 were rehashed, cutting chances of ambiguity. Fragment 4 underwent a major transformation, which resulted in the omission of precision information. Technically, this transformation refers to an omission, but it was uncalled for.

The translation of the edited text inherited only one inaccuracy in fragment 2, which survived AI-powered pre-editing (see Table 9). We attribute it to the fact that the lexical unit refers to a non-alternative term in this context. As expected, fragment 4 sustained the logic of the original at the expense of the precision information.

**Figure 11.** Pre-editing of the source text by Yandex GPT

**Рисунок 11.** Упрощение исходного текста с помощью Yandex GPT

Упрости текст для машинного перевода: "На первом этапе основным преимуществом С-400 над системами предыдущих модификаций стала более высокая степень автоматизации боевой работы, использование современной элементной базы, возможность интеграции в различные уровни управления не только ВВС, но и других видов вооруженных сил, а также увеличение числа одновременно сопровождаемых и обстреливаемых целей. Хотя в 2007 году официально было заявлено, что дальняя граница зоны поражения ЗРС С-400 может достигать 400 км, до недавнего времени в боекомплект входили только зенитные управляемые ракеты семейства 48Н6, поступившие на вооружение в начале 1990-х вместе с ЗРС С-300ПМ. Максимальная дальность поражения крупных аэродинамических целей ЗУР 48Н6ЕЗ на средних высотах составляет 250 км. В целом зенитный ракетный дивизион С-400 сохранил структуру С-300П, включая многофункциональную РЛС, пусковые установки, автономные средства обнаружения и целеуказания. Все боевые средства ЗРС размещаются на самоходных колесных шасси повышенной проходимости, имеют встроенные системы автономного электропитания, топопривязки, связи и жизнеобеспечения. Для обеспечения длительной непрерывной работы предусмотрена возможность электропитания от средств внешнего электроснабжения. В состав средств управления боевой работой ЗРС С-400 входят пункт боевого управления 55К6Е и РЛС обнаружения 91Н6Е."

На начальном этапе С-400 имел несколько преимуществ перед предыдущими системами. Во-первых, он имел более высокую степень автоматизации в бою. Во-вторых, он использовал современные технологии. В-третьих, он мог быть интегрирован в разные уровни управления, не только в ВВС, но и в другие виды вооруженных сил. Также увеличилось число целей, которые можно отслеживать и атаковать одновременно.

В 2007 году было заявлено, что дальность поражения С-400 составляет 400 километров. Однако до недавнего времени только определенные типы ракет входили в его арсенал. Эти ракеты были разработаны в начале 1990-х и использовались вместе с С-300ПМ. Они могут поражать цели на расстоянии до 250 километров.

Зенитный ракетный комплекс С-400 имеет схожую структуру с С-300Р. Он включает в себя многофункциональную радиолокационную станцию, пусковые установки и автономные системы обнаружения и наведения. Все компоненты системы размещены на мобильных колесных платформах.

**Table 8.** Comparison of the original with its simplified version produced by Yandex GPT  
**Таблица 8.** Сопоставление оригинала с его упрощенной версией после обработки в Yandex GPT

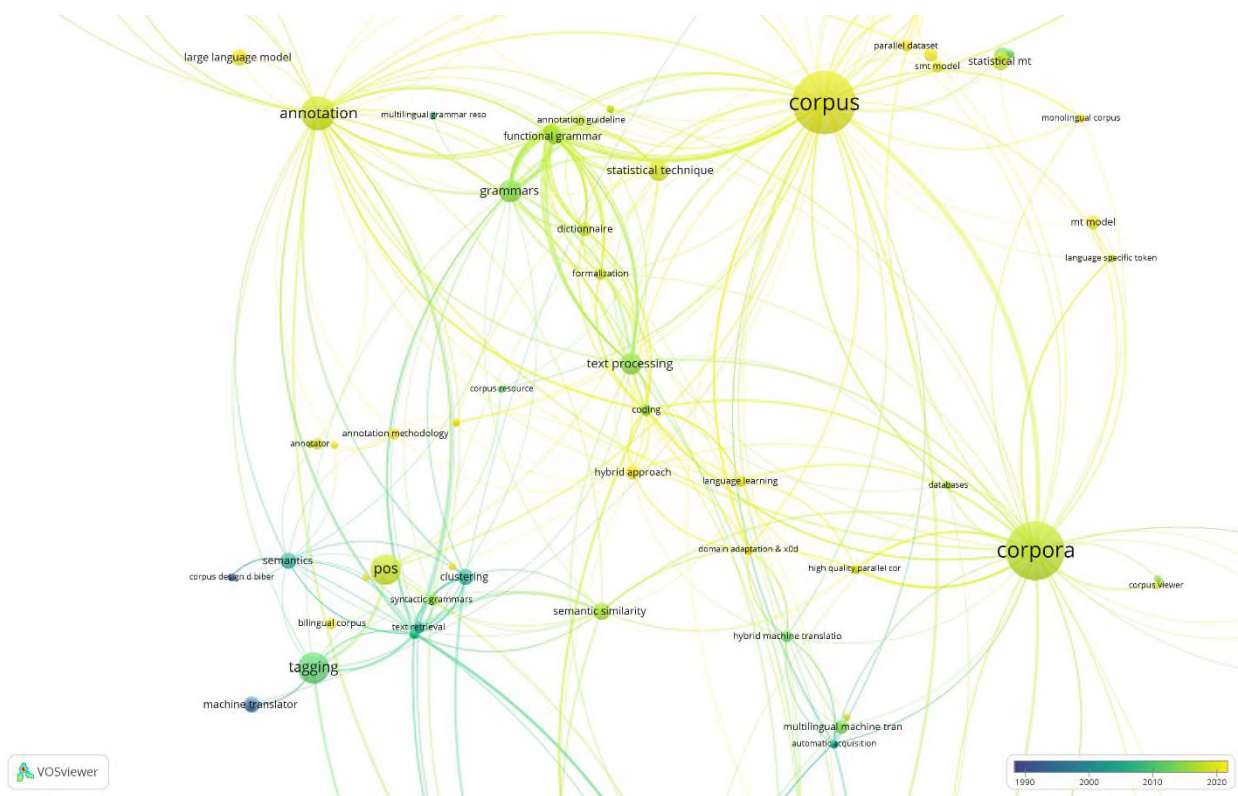
Yandex GPT	Original
<p>На начальном этапе С-400 имел несколько преимуществ перед предыдущими системами. Во-первых, он имел более высокую степень автоматизации в бою. Во-вторых, он использовал <b>(1) современные технологии</b>. В-третьих, он мог быть интегрирован в разные уровни управления, не только в ВВС, но и в другие <b>(2) виды вооруженных сил</b>. Также увеличилось число целей, которые можно <b>(3) отслеживать и атаковать одновременно</b>. В 2007 году было заявлено, что дальность поражения С-400 составляет 400 километров. <b>(4) Однако до недавнего времени только определенные типы ракет входили в его арсенал. Эти ракеты были разработаны в начале 1990-х и использовались вместе с С-300РМ. Они могут поражать цели на расстоянии до 250 километров.</b></p>	<p>На первом этапе основным преимуществом С-400 над системами предыдущих модификаций стала более высокая степень автоматизации боевой работы, использование <b>(1) современной элементной базы</b>, возможность интеграции в различные уровни управления не только ВВС, но и других <b>(2) видов вооруженных сил</b>, а также увеличение числа <b>(3) одновременно <u>сопровожаемых</u> и обстреливаемых целей</b>. Хотя в 2007 году официально было заявлено, что дальняя граница зоны поражения ЗРС С-400 может достигать 400км, до недавнего времени в боекомплект входили только зенитные управляемые ракеты семейства 48Н6, поступившие на вооружение в начале 1990-х вместе с <b>(4) ЗРС С-300ПМ</b>. Максимальная дальность поражения крупных аэродинамических целей <b>(4.1) ЗУР 48Н6Е3</b> на средних высотах составляет 250км.</p>

**Table 9.** Comparison of the human translation with MT of the original and its simplified version  
**Таблица 9.** Сопоставление перевода, выполненного человеком, с машинными переводами оригинала и его упрощенной версией

DeepL	Yandex GPT-DeepL	Human translation
<p>At the first stage, the main advantage of the S-400 over the systems of previous modifications was a higher degree of combat automation, the use of modern <b>(1) element base</b>, the ability to integrate into various levels of control not only of the Air Force, but also of other <b>(2) branches</b> of the armed forces, as well as an increase in the number of <b>(3) simultaneously <u>escorted</u> and fired targets</b>. Although in 2007 it was officially announced that the S-400 SAM's long-range kill zone</p>	<p>At the initial stage, the S-400 had several advantages over previous systems. Firstly, it had a higher degree of automation in combat. Second, it utilised <b>(1) modern technology</b>. Third, it could be integrated into different levels of command, not only in the Air Force but also in other <b>(2) branches</b> of the armed forces. It also increased the number of targets that could be <b>(3) tracked and attacked simultaneously</b>. In 2007, the S-400 was claimed to have a</p>	<p>By the time the S-400 achieved an IOC its major advantages over predecessors were a far superior automated performance, <b>(1) advanced hardware</b>, increased <b>(3) simultaneous engagement capability</b>, and provisions for integration into various command echelons of not only the AF but also other <b>(2) services</b>. Though it was stated in 2007 that the system boasted a maximum engagement range of 400km, until recently, its ammunition</p>

DeepL	Yandex GPT-DeepL	Human translation
<p>can reach 400 kilometers, until recently its ammunition included only 48N6 family antiaircraft guided missiles, which entered service in the early 1990s together with the (4) S-300PM <u>SAM</u>. (4.1) The 48N6E3 <u>SAMs</u> have a maximum range of 250 km at medium altitudes against large aerodynamic targets.</p>	<p>range of 400 kilometres. (4) <i>However, until recently, only certain types of missiles were part of its arsenal. These missiles were developed in the early 1990s and used alongside the S-300PM.</i> They can hit targets up to 250 kilometres away.</p>	<p>load comprised obsolete (4.0) 48N6 <u>SAMs</u>, phased into service along with the (4) S-300PM back in the early 1990s. Thus, the (4.1) 48N6E3 can reach large mid-altitude threats at no more than 250km.</p>

Figure 12. Current trends in the MT development<sup>16</sup>  
 Рисунок 12. Современные тенденции в области развития МП



<sup>16</sup> Based on a Lens selection of 1560 MT papers covering a period of 1990-2020 (key words: MT, machine translation, annotation, tagging, NLP, natural language processing, parsing, corpora, corpus, parallel corpus, comparable corpus, rule-based, neural MT, NMT).

The obtained results suggest that complex transformations should be carried out in the original text prior to its machine translation. In addition to eliminating complexities, pre-editing is recommended for correcting misprints and errors, as well as removing ambiguity, both being equally detrimental to the MT output.

At the current stage of MT development, the main trends have undergone all but minor changes (see Figure 12). Corpora and their architecture are still relevant for training language models, which are gaining prominence in feasibility studies testing their suitability for fine-tuning MT for specific translation domains. Information processing for corpora expansion is no longer on the list, since it is now possible to convert speech to text in addition to optical recognition of graphical materials. Annotation and corpus markup systems have made a significant progress; however, both have a long way to go before they mature.

### Conclusions

As a result of analyzing the most successful MT projects and their impact on modern realities, we have arrived at a number of significant conclusions. First, MT was successfully put to use far before computers became widespread. The effectiveness of the technology depends on the users' understanding of current limitations. Before 1990, there were a few of them, all of which were fully taken into account by the developers of METEO-1. Second, the MT evolution traces a spiral trajectory. All of the MT applications noted in the 1966 ALPAC report are still relevant today. Third, modern MT has already been recognized by professional interpreters as a full-fledged tool, as evidenced by the PEMT emergence. Fourth, not all the issues in this area have been solved. One of the most urgent is semantic markup, designed to teach the computer to comprehend the text. This drawback does not allow, among other things, to improve the quality of translation of technical texts. We believe that a feasible solution is the implementation of such an

advantage of specialized literature as low entropy. This approach calls for training language models, which is already possible at the current stage of MT development. Fifth, due to the improvement in the MT quality, giant enterprises have started to abandon post-editing, a major hindrance curbing the translation rate. However, aesthetic impact on the reader, high fidelity and contextualized translation are still beyond the MT grasp. Therefore, AI-powered pre- and post-editing practices should be utilized. Sixth, deep transformations and lower levels of equivalence are not attainable when applying MT to raw texts. Technologies should be led to desired translation solutions in pre-editing. As for future steps, it is advisable to delve into language models and their training, PEMT technologies as the most efficient way of translation in the conditions of rapid growth of information flow, as well as ways of adjusting equivalence and transformations when addressing translation problems in MT.

### References

- Bharati, A., Chaitanya, V., Kulkarni, A. and Sangal, R. (2003). Anusaraka: Machine Translation in Stages, *ArXiv, cs.CL/0306130*. (In English)
- Brown, P. F., Della Pietra, S. A., Della Pietra, V. J. and Mercer, R. L. (1993). The Mathematics of Statistical Machine Translation: Parameter Estimation, *Computational Linguistics*, 19 (2), 263–311. (In English)
- Costa-jussa, M., Escolano, C. and Fonollosa, J. (2017). Byte-based Neural Machine Translation, *Proceedings of the First Workshop on Subword and Character Level Models in NLP*, 154–158. DOI: 10.18653/v1/W17-4123 (In English)
- Daems, J., Vandepitte, S., Hartsuiker, R. and Macken, L. (2017). Translation Methods and Experience: A Comparative Analysis of Human Translation and Post-editing with Students and Professional Translators, *META*, 62 (2), 245–270. <https://doi.org/10.7202/1041023ar> (In English)
- Fumani, F. Q. and Reza, M. (2007). Ambiguity in Machine Translation, *Ketabdari Va Etella'resaani*, 9, 21–38. (In English)
- Gashaw, I. and Shashirekha, H. L. (2019). Amharic-Arabic Neural Machine Translation,

*Computer Science & Information Technology (CS & IT)*, 55–68.

<https://doi.org/10.5121/csit.2019.91606>

(In English)

Gruzdev, D. Yu. and Kodzhebash, D. O. (2023). POS-powered queries for neat and lean concordances in ad-hoc corpora analysis, *Teoreticheskaya i prikladnaya lingvistika* [Theoretical and Applied Linguistics], 9 (4), 35–48. DOI: 10.22250/24107190\_2023\_9\_4\_35 (In English)

Hurskainen, A. (2018). Can machine translation assist in Bible translation?, *SALAMA – Swahili Language Manager: Technical reports on LT*, 62, Helsinki. (In English)

Hutchins, J. (2004a). Two precursors of machine translation: Artsrouni and Trojanskij, *International Journal of Translation*, 16 (1), 11–31. (In English)

Hutchins, J. (2004b). The first public demonstration of machine translation: the Georgetown-IBM system, 7th January 1954. (In English)

Volkart, L. and Bouillon, P. (2022). Studying Post-Editese in a Professional Context: A Pilot Study, *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, Ghent, Belgium. European Association for Machine Translation, 71–79 (In English)

López, V. F., Corchado, J. M., De Paz, J. F., Rodríguez, S. and Bajo, J. (2010). A SomAgent statistical machine translation, *Applied Soft Computing*, 11 (2). <https://doi.org/10.1016/j.asoc.2010.08.018> (In English)

Nelyubin, L. L. (1975). Informacionno-statisticheskie i inzhenerno-lingvisticheskie osobennosti yazyka i teksta v usloviyah spekkommunikacii [Information-statistical and engineering-linguistic features of language and text in conditions of special communication], D. Sc. Thesis, Applied linguistics, Leningrad. (In Russian)

Nelyubin, L. L. (1983). Translation and applied linguistics, Vysshaya Shkola Publishers, Moscow, USSR. (In Russian)

Névél A., Dogan R. and lu Zh. (2010). Semi-automatic semantic annotation of PubMed queries: A study on quality, efficiency, satisfaction, *Journal of biomedical informatics*, 44. 310–8. DOI: 10.1016/j.jbi.2010.11.001. (In English)

Nirenburg, S. (ed.) (1993). *Progress in machine translation*, IOS Press, Amsterdam, Netherlands. (In English)

O'Brien, S. Balling, L. W., Carl, M., Simard, M. and Specia, L. (eds.) (2014). *Post-editing of Machine Translation: Processes and Applications*, Cambridge Scholars Publishing. (In English)

Oladosu, J., Esan, A., Adeyanju, I., Adegoke, B., Olaniyan, O. and Omodunbi, B. (2016). Approaches to Machine Translation: A Review, *FUOYE Journal of Engineering and Technology*, 1 (1), 120–126. <https://doi.org/10.46792/fuoyejet.v1i1.26> (In English)

Ornstein, J. (1955). Mechanical Translation: New Challenge to Communication, *Science*, 22 (3173), 745–748. DOI: 10.1126/science.122.3173.745 (In English)

Quach, K. (2016). Google's neural network learns to translate languages it hasn't been trained on: First time machine translation has used true transfer learning, *The Register*, available at: [https://www.theregister.com/2016/11/17/googles\\_neural\\_net\\_translates\\_languages\\_not\\_trained\\_on/](https://www.theregister.com/2016/11/17/googles_neural_net_translates_languages_not_trained_on/) (Accessed 10 June 2024). (In English)

Quintana, R. C. and Castilho, S. (2022). A review of the Integration of Machine Translation in CAT tools, *Proceedings of the International Conference “New Trends in Translation and Technology NeTT 2022”*, Rhodes Island, Greece, 214–221 (In English)

Tripathi, S. and Kansal, V. (2020). Machine Translation Evaluation: Unveiling the Role of Dense Sentence Vector Embedding for Morphologically Rich Language, *International Journal of Pattern Recognition and Artificial Intelligence*, 34 (1), 2059001. <https://doi.org/10.1142/S0218001420590016> (In English)

Schuster, M., Johnson, M. and Thorat, N. (2016). Zero-Shot Translation with Google's Multilingual Neural Machine Translation System, *AI Blog*, available at: <https://ai.googleblog.com/2016/11/zero-shot-translation-with-googles.html> (Accessed 10 June 2024). (In English)

Suxia Lei and You Li. (2023). English Machine translation System Based on Neural Network Algorithm, *Procedia Computer Science*, 228, 409–420. <https://doi.org/10.1016/j.procs.2023.11.047> (In English)



Tomasello, L. (2020). Neural Machine Translation and Artificial Intelligence: What Is Left for the Human Translator?, Master's degree theses, University of Padua, Italy. *(In English)*

Weaver, W. (1949). The Mathematics of Communication, *Scientific American*, 181 (1), 11-15. *(In English)*

*Все авторы прочитали и одобрили окончательный вариант рукописи.*

*All authors have read and approved the final manuscript.*

*Конфликты интересов: у авторов нет конфликтов интересов для декларации.*

*Conflicts of interests: the authors have no conflicts of interest to declare.*

**Viktoriya V. Dyomochkina**, Ph.D. Student at the English Department, Military University, Russia.

**Демочкина Виктория Викторовна**, адъюнкт кафедры английского языка (основного)

ФГКВОУ ВО «Военный университет имени князя Александра Невского» МО РФ, Россия.

**Dmitry Yu. Gruzdev**, Ph.D. in Linguistics, Associate Professor, Deputy Head of the English Department, Military University, Russia.

**Груздев Дмитрий Юрьевич**, к.ф.н., доцент, заместитель начальника кафедры английского языка (основного) ФГКВОУ ВО «Военный университет имени князя Александра Невского» МО РФ, Россия.

**Elena V. Lukyanova**, Ph.D. in Linguistics, Associate Professor, Department of Foreign Languages and Professional Communication, Belgorod State National Research University, Russia.

**Елена Викторовна Лукьянова**, к.ф.н., доцент, доцент кафедры иностранных языков и профессиональной коммуникации, Белгородский государственный национальный исследовательский университет, Россия.