CONTEMPORARY TENDENCIES IN DEVELOPMENT OF MACHINE TRANSLATION FROM ENGLISH INTO UKRAINIAN

O. Novikova
Candidate of Philological Science, Associate professor of Department of translation and linguistic training of foreigners, Oles Honchar Dnipro national university
novikovaolga19911991@gmail.com
orcid.org/0000-0001-9119-5154

I. Suima
Candidate of Philological Science, Associate professor, Associate professor of Department of translation and linguistic training of foreigners, Oles Honchar Dnipro national university
Suima-irina@i.ua
orcid.org/0000-0002-2209-8614

K. Shevchyk
Candidate of Philological Science, Associate professor, Associate professor of Department of translation and linguistic training of foreigners, Oles Honchar Dnipro national university
shevchikekaterina3@gmail.com
orcid.org/0000-0002-4423-228X

Introduction. As a separate scientific discipline, translation studies emerged in the second half of the 20th century predominantly due to globalization efforts and intensification of inter-ethnic relations. By the end of the millennium, a new discipline was developing at an ever-increasing rate. Machine translation quickly became not just a theoretical discipline but a cornerstone of scientific cooperation, sitting on the crossroads of computer science, engineering and linguistics (Darwish, 2001; Forcada, 2017: 291; Gordin, 2015: 147; Hearne, 2011: 207; Jurafski, 2009: 276). Translations of texts from one language to another can be categorized as routine work, but only partially. On the one hand, in the work of any translator there are quite a few elements of formalism, but on the other hand, no translation can be done fully formally without any creative decision-making.

Today machine translation is one of types of human activity. Machine translation can greatly facilitate global communication, accelerating the translation process, despite the imperfect quality of the source text.

Most often the results of online tools require post-editing and can only be effectively used by those who already speak the target language to some extent. The need for a competent translation is growing every year. Today, the search for an algorithm to deliver this quality of translation is one of the most important questions in computer science and linguistics, therefore informing the scientific relevance of this work.

The aim of this paper is to analyze different approaches to the machine translation systems, their characteristics, efficacy and the quality of their output on the examples from Google Translate, Microsoft Translator and Yandex.

To achieve this aim, the following tasks were set:
1) to identify the most capable algorithms of MT in use today;
2) to compare the results of translations made by online translators;
3) to analyze typical stylistic, lexical and grammatical errors that appear in the translation;
4) to identify the advantages and disadvantages of online translators;
5) to provide recommendations for improving machine translation.

To solve these tasks, we use such methods over the course of this work: descriptive,
comparative, analysis, experiment and the method of linguistic interpretation of the results obtained.

**Methods and methodology of investigation.** At the present stage of research, there are two main incentives for the development of machine translation. The first is purely scientific; it is determined by the complexity and intricacy of machine translation models. As a type of linguistic activity, translation affects all levels of language – from grapheme recognition to conveying the content of individual sentences and text as a whole. There is a need to accelerate the process of and increase the volume of translation, thus increasing the requirements for translation as an industrially applicable product. The second incentive is social. It is driven predominantly by the growing role of translation in the modern world as a prerequisite for the provision of interlingual communication, the volume of which is increasing every year.

The first generation of machine translation systems was based on sequential translation algorithms, that could on translate word-by-word, phrase-by-phrase. The capabilities of such systems were determined by the available vocabulary sizes, which directly depended on the amount of addressable computer memory. Translation of the text was carried out in separate sentences with meaningful connections between them not being taken into account. Such systems are called direct translation systems. Later on, they were replaced by subsequent systems, in which the translation from language to language was performed at the level of syntactic structures. The translation algorithms used a set of logical operations, with the following steps (Hutchins, 2005):

1) analyzing the translation sentence;
2) constructing its syntactic structure according to the rules of grammar of the source language;
3) transforming it into a syntactic structure of the original sentence according to the target language grammar;
4) synthesizing the original sentence, substituting the right words from the dictionary. Such systems are called T-systems (from the word «transfer»).

Building machine translation systems based on obtaining some meaningful representation of the input sentence through its semantic analysis is considered to be the aim of machine translation. It should then be followed by a synthesis of the sentence in the target language according to the obtained meaningful representation. Such systems are called I-systems (from the word «interlingua»). It is generally believed that the next generations of machine translation systems will belong to the class of I-systems (Hutchins, 1986: 135).

The first experiments with creating machine translation programs showed that it was necessary to solve problems one by one. There were too many difficulties and inaccuracies in the rules of formalization and creation of algorithms for text analysis, no agreement on the contents of dictionaries or on linguistic patterns to be used in machine translation. It turned out that traditional linguistics had neither the factual materials nor the ideas and concepts needed to build machine translation systems that would use the inner content to reconstruct the text.

At the end of the twentieth century, the statistical approach to machine translation started to develop. Ostensibly, such translation is not based on rules, but on statistics. The main method of such a translation is to train the machine by providing a sufficiently large (sample size of hundreds of thousands) number of parallel texts – containing the same information in different languages. Let us consider the methods of statistical translation by looking at the example model used by Yandex in the Yandex.Translate system as analyzed by M. Vozniuk. It consists of three stages: a translation model, a language model, and a decoder (Vozniuk, 2011: 143). The translation model for a pair of languages is a table consisting of all the words and phrases of the source language known to the machine and their translations into the target language, indicating the likelihood of such a conversion.

Statistics-based machine translation (SBMT or SMT) takes into account not only individual words, but also phraseology consisting of several words. Next is the language model, namely the language model for target language. It is a list consisting of all words and phrases found in the provided texts along with the frequency of their use. The system then proceeds directly to the
translation process that the decoder deals with. Each sentence of the source text sorts through all viable translation options, combining phrases from the translation model, and arranging them in descending order of probability. Thus, the language model tells the decoder which version of the translation is more suitable for the given phrase, based on statistical data.

The main advantage of statistical systems is that their quality does not lag behind the development and mobility of the language: if any changes occur in the language, the system immediately recognizes this and learns independently (Chan, 2015: 385). Statistical systems also have high smoothness, that is, the output text is similar to the speech delivered by a person. However, the existence of such a system requires serious technical resources and high-quality parallel texts. Another significant drawback of such a system is the lack of sensitivity to the fine structure of the text, and, as a consequence, a large number of grammatical errors may be contained in the output text.

Nowadays machine translation is a rapidly developing area, especially considering the innovations brought by the neural networks. Since 2014, according to the BLEU metrics, the average evaluation of a translated text and how it compares to an authored translation significantly increased to 0.4, which indicates that MT systems are slowly closing the gap between automated and manual translations.

It is best to keep in mind, however, that while quality is the key metric for evaluation of various target language texts, we should keep in mind that these two types of translation may serve different purposes.

**Results and discussion.** In this article we will be analyzing stylistic, lexical and grammatical errors made by modern NMT systems when tasked with translating a text of a belles-lettres style. For our research we decided on using Kurt Vonnegut’s 1973 novel «Breakfast of Champions or Goodbye Blue Monday». This novel was chosen due to Vonnegut’s style being much more laconic and literal than that of a typical novelist. This feature of author’s idiomatic style, complete with generous usage of simple stylistic devices such as inversion, deadpan and others, allows us to test NMT systems on a genuine literary work, all while not overtaxing said systems with too much contextual knowledge. In total, 50 samples from the novel were chosen for the analysis.

Even in the first example we can already see what will quickly become a trend going forward: Microsoft’s translation utility is not properly equipped to deal with inflections and grammatical categories which differ from English (such as cases, grammatical gender etc.)

This example clearly shows that in the situation with a colloquialism, all three of the NMT systems default to literal translation (squat – присідання, присадкуватий). To have squat is generally understood to mean «to have nothing at all», but as the neural; network was not made aware of general context of the phrase, it fell back on the old word-for-word translation techniques.
This is a first example where we can see a clear divergence between our results. Whereas Google’s system leaned toward a more literal translation («багато всього»), Yandex decided to use paraphrase for its translation, which gave it a more natural feel («все ще було в достатку»).

The most peculiar, however, is Microsoft’s translation: it simply omitted the names of the protagonists from the beginning of the sentence, as if unsure how to translate them. Name translation will be a more interesting point in our next example:

Here we can more clearly see the issue with translation of names. Both Google and Microsoft translated «Trout» as «форель», which, while lexically correct is completely wrong from a contextual standpoint, as it is the last name of the novel’s protagonist.

Moreover, Microsoft was unable to identify a correct vector of context for a phrase «for short», which led to a rather incongruent translation by the system.

This example mainly falls victim to literal translation, where «piece of evil nonsense» was translated word-for-word by both Google’s and Microsoft’s systems as «імітаток зла» and «імітаток злих дурниць». Yandex successfully evaded this particular issue, but all three chose to directly translation word-combination «beacon of freedom», whereas, in our opinion, a generalization would have worked much better (i.e. «символ свободи»).
<table>
<thead>
<tr>
<th>Original text</th>
<th>Google Translate</th>
<th>Yandex.Translate</th>
<th>Microsoft Translator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actually, millions of human beings were already living full and imaginative lives on the continent in 1492.</td>
<td>На справді мільйони людей вже повною мірою живуть на континенті у 1492 році.</td>
<td>Насправді мільйони людей вже жили повним і творчим життям на континенті в 1492.</td>
<td>Насправді, мільйони людських істот вже живуть повним і творчим життям на континенті в 1492.</td>
</tr>
</tbody>
</table>

This is a rather peculiar sample, as the underlying statistical data may explain the issue with this paragraph. Two out of three algorithms translated «imaginative» as «творчий». As this word if oftentimes used as a synonym for creative, they just substituted the closest word they have for that vector. We would suggest «насичений» as a contextual synonym here, as it has slightly wider application.

<table>
<thead>
<tr>
<th>Original text</th>
<th>Google Translate</th>
<th>Yandex.Translate</th>
<th>Microsoft Translator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Here was the core of the bad ideas which Trout gave to Dwayne: Everybody on Earth was a robot, with one exception – Dwayne Hoover.</td>
<td>Тут було серцевина поганих ідей, які Траут дав Двейну: Усі на Землі були роботами, за одним винятком – Двейн Гувер.</td>
<td>Ось у чому була суть поганих ідей, які Траут вселяв Двейна: всі на Землі були роботами, за одним винятком – Двейн Гувер.</td>
<td>Тут був основний з поганих ідей, які форель дала Двейн: все на землі був робот, з одним винятком – Двейн Гувер.</td>
</tr>
</tbody>
</table>

Here Google clearly misjudged the meaning of the word «core», which can mean «серцевина» in the context of talking about wood and trees, where wooden core is the innermost part of the tree. Even Microsoft Translator managed the give roughly the correct idea, although the pervasive problems with inflections and syntax errors make this text mostly unreadable. All systems are constrained by inability to completely change the structure of the sentence without altering its meaning or intent – which can be perfectly seen as all three of them use dash to show who the narrator is talking about. A full transformation could be beneficial here to make the sentence more natural-sounding: «Ось у чому була суть поганих ідей Траута, якими він годував Двейна: усі на Землі були роботами, за винятком самого Двейна Гувера…».

<table>
<thead>
<tr>
<th>Original text</th>
<th>Google Translate</th>
<th>Yandex.Translate</th>
<th>Microsoft Translator</th>
</tr>
</thead>
<tbody>
<tr>
<td>His high school was named after a slave owner who was also one of the world's greatest theoreticians on the subject of human liberty.</td>
<td>Його середня школа була названа в честь рабовласника, який також був одним з найбільших теоретиків світу на тему свободи людини.</td>
<td>Його Вища школа була названа на честь рабовласника, який також був одним з найбільших теоретиків у світі з питання людської свободи.</td>
<td>Його середню школу було названо на честь рабської власниці, яка також була одним з найбільших теоретиків світу з тематики людської свободи.</td>
</tr>
</tbody>
</table>

This example mostly displays that even with major issues in regards to syntax, Microsoft Translator can still produce reasonably accurate results. Previously «high school» was understood and translated as «середня школа», which is represented by Google’s and Microsoft’s translation. Nowadays, though, it is mostly translated verbatim as «старша школа». Also of note is the final part of the sentence, with both Yandex («одним з найбільших теоретиків у світі з питання людської свободи») and Microsoft’s «одним з найбільших теоретиків світу з тематики людської свободи» showing great results for translating structures with prepositions.
This paragraph shows us the downside of NMT systems. Yandex’s translation of «bad chemicals» is most likely due to rather frequent occurrence of a phrase «bad chemical reaction» in the dataset used to train the system, which led to its decision to translate it directly, omitting the word chemical as it deemed it unnecessary. Microsoft Translator translated verb «to stick» as «дотримуватися», which makes sense as it is oriented on business users and international communication. This meaning of the word is more common than the one the author intended. This shows us that Microsoft still leans heavily on their statistical model, while they are still in the process of implementing full capabilities of neural networks.

<table>
<thead>
<tr>
<th>Original text</th>
<th>Google Translate</th>
<th>Yandex.Translate</th>
<th>Microsoft Translator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dwayne's bad chemicals made him take a loaded thirty-eight caliber revolver from under his pillow and stick it in his mouth. This was a tool whose only purpose was to make holes in human beings.</td>
<td>Погані хімічні речовини Дуейна змусили його взяти з-під подушки завантажений револьвер тридцяти восьми калібрів і засунути його в рот. Це був інструмент, єдиною метою якого було зробити діри в людях.</td>
<td>Довгі неможливості Дуейна змусили його взяти заряджений револьвер 38 калібру під подушку і дотримуватися його в рот. Це був інструмент, чия мета полягала в тому, щоб зробити отвори в людських істотах.</td>
<td></td>
</tr>
</tbody>
</table>

This is the only example in our testing whereby all three systems had the same output. From this one may induce that NMT can now handle translation of simple sentences quite well.

<table>
<thead>
<tr>
<th>Original text</th>
<th>Google Translate</th>
<th>Yandex.Translate</th>
<th>Microsoft Translator</th>
</tr>
</thead>
<tbody>
<tr>
<td>The clothes were conservative and neat, in Harry's opinion.</td>
<td>Одяг був консервативним і акуратним, на думку Гаррі</td>
<td>Одяг був консервативним та акуратним, на думку Гаррі</td>
<td>Одяг був консервативним і акуратним, на думку Гаррі</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Original text</th>
<th>Google Translate</th>
<th>Yandex.Translate</th>
<th>Microsoft Translator</th>
</tr>
</thead>
<tbody>
<tr>
<td>So poor Harry spent a wretched Veterans'</td>
<td>Так бідний Гаррі після цього бідолаха Гаррі</td>
<td>Так цьо після цього бідолаха Гаррі</td>
<td>Так бідні Гаррі провів жалюгідний</td>
</tr>
</tbody>
</table>

Right from the outset we can see that only one (Yandex) of three NMT systems has successfully translated «the body bag» idiom. The other two systems failed to recognize it as well as to set correct inflections for the words according to their cases. All three systems failed the translation of the last sentence in this example. «I did» here refers to the previous sentence, and as such, a more appropriate and simpler solution should have been «Я» or «Це зробив я». What makes this failure more surprising is the fact that recurrent neural networks were able to retain contextual information outside the bounds of a single sentence since early 2017. We believe this mistranslation may be due to an incorrectly trained target language neural network.

<table>
<thead>
<tr>
<th>Original text</th>
<th>Google Translate</th>
<th>Yandex.Translate</th>
<th>Microsoft Translator</th>
</tr>
</thead>
<tbody>
<tr>
<td>I do not know who invented the body bag. I do know who invented Kilgore Trout. I did.</td>
<td>Я не знаю, хто винайшов мішок для трупів. Я знаю, хто винайшов Кілгора форелі. Я робив.</td>
<td>Я не знаю, хто винайшов сумку для тіла. Я знаю, хто винайшов Кілгора. Я зробив.</td>
<td>Я не знаю, хто винайшов мішок для трупів. Я знаю, хто винайшов Кілгора. Я зробив.</td>
</tr>
</tbody>
</table>
Day weekend after that. But Dwayne spent a worse one.

In this example all of our systems were unable to transform the original sentence in such a way that would conform to grammatical rules of the target language. A more correct translation would be «Після того бідолашний Гаррі жахливо провів День ветеранів. Але Дуейн провів його ще гірше». The problem is once again in the lack of contextual transference between the first and the second sentence. Only Yandex tried to insert additional context into the second sentence, but it chose the wrong one, as there was no mention of night in the original text.

Here we once again observe that translation systems designed by primarily English-speaking multinational companies, such as Microsoft and Google, took a much more English-centric approach to the training of their NMT systems. Both of those systems exhibit typical issues, such as using a wrong grammatical gender, wrong tense (would in the case of a reminiscence should be translated by past tense in Ukrainian), small stylistic issues («Дай мені всі ваші гроші» where the first word implies familiarity towards a single recipient, but the second one implies either more recipients of the message or a degree of respectful address, which is not typical for a robbery) etc.

In this particular case, Google’s algorithm could not properly handle the tense of the sentence («був отруює») and case inflections («божевільному дилера») whereas Yandex’s hybrid approach and its predilection towards Slavic languages yielded almost perfect translation. It automatically picked up on the context that Holiday Inn was, in fact, a hotel, and specified this information for Ukrainian reader. One peculiar fact is that Yandex even transliterated word «Pontiac» and assigned it the inflection of the genitive case, which may suggest a more complete understanding of the concept and the context behind it, while two other systems opted to keep the brand in English. Microsoft’s algorithm once again mistranslated the protagonist’s last name as the service was never really designed to translate literary works.
From the outset, we can see a difference in translation of a person’s name. Google Translate opted to use the approach of transliteration. Yandex, on the other hand, decided on transcription. While both approaches may be valid, we believe the latter is the more accurate of the two. Microsoft Translator could not handle this task and settled somewhere in between the two other answers. What is more interesting is that the first two NMT systems are much more aware of cultural realia and famous works of literature as they are able to translate name «Ivanhoe» without issues. Microsoft’s business-oriented MT system apparently lacks this awareness which is actively affecting its ability.

This example was particularly interesting to analyze. All three systems managed to successfully translate a character’s name. However, only Google’s algorithm preserved the capitalization on the word «Chaos», which was correct as the author intentionally imbued it with significance, as if making it a deity from the Greek myth. Even more interesting is that phrase «outboard motor» was translated three different ways. Yandex was able to correctly specify the type of the motor, whereas Microsoft’s system went for generalization, hence the «човновий мотор».

Let us take a look at one final example:

«It's all like an ocean!» cried Dostoevski. I say it's all like cellophane.
Only Microsoft Translator had minor issues with translation of Dostoyevsky’s name. Barrng that, all three variants of translation are acceptable.

**Conclusions.** All in all, machine translation of bellettristic texts was handled exceptionally well by Yandex and was quite acceptable (barring numerous grammatical errors) on Google’s platform. The outlier in this case is Microsoft Translator, whose mistranslation of realia and same aforementioned mistakes make its output much less readable that its competitors.

Although the quality of machine translation is largely improved, it is still not usable for anything outside of facilitation of global communications. Considering that this area of research is closely related to computer speech recognition, an argument could be made that MT will find future uses as a tool for an on-the-fly interpreting for those traveling abroad for business without proper knowledge of the language. As the services of a highly-trained and qualified interpreter are costly, it would make sense for large corporations to invest upfront into machine translation to save costs later, which is exactly Google’s current trend with their system already having a beta version of said feature.

According to our research, the mistakes made by modern NMT systems mostly fall into the following categories: incorrect/inappropriate punctuation for the target language, where the translation may be coherent and adequate, but incorrectly used commas make it harder to parse; inflections, which usually come in the form of adjectival phrases (погана пес, зелений ліси etc.). The same is also true for verb conjugations.

The main problems we see arising from such translations arise from the fact that the systems depend on a large amount of high-quality data sets (i.e., corpora of texts for specific language pairs). The quality of these sets directly influences the quality of the output, which in our case is the quality of the target language text. It can be seen by comparing the average quality of translation between Google’s and Microsoft’s systems. The former one makes less mistakes on average and does not have as many issues in regards to identifying a contextual meaning of a polysemantic lexeme.

We believe that this issue can be fixed to a certain extent one of two ways: hiring professional translators and linguists to compile those parallel corpora or create a possibility for every person to contribute to this process even on a small scale. The first approach would be very time and labor consuming, but would ultimately provide us with a higher quality data set, which may lead to further improvements in MT. The second is already being deployed by all three major NMT systems but may lead slower progression due to lack of quality control and oversight.

A further analysis of this technology in the following years may yield useful information regarding their evolution and changes in behavior as the systems mature.

Moreover, a comparative study of the same inputs on the same systems in 3 to 5 years will depict qualitative differences and can give us a better insight as to the progression rate in the field.

For us, another potential prospect of this research lies in widening the subject area of texts chosen to reflect the variety of writing styles in use on the Internet right now. Inclusion of texts from confessional, business, and other styles may allow us to highlight more lacunae in the neural network models and to suggest further means of improvement.

**REFERENCES**


Анотація

Постановка проблеми. Сьогодні машинний переклад є одним із чинників повсякденної діяльності людини. Машинний переклад може значно покращити глобальні комунікації, пришвидшуючи процес перекладу, незважаючи на недосконалу якість вихідного тексту. Найчастіше результати онлайн-інструментів потребують постредагування й можуть ефективно використовуватися лише тими, хто певною мірою вже говорить мовою перекладу. Потреба в якісному перекладі зростає з кожним роком. Сьогодні пошук алгоритму для забезпечення такої якості перекладу є одним з найважливіших питань інформатики, кібернетики та лінгвістики, що доводить наукову новизну цієї роботи.

Мета статті – аналіз різних підходів до просування систем машинного перекладу, їхніх характеристик, ефективності та якості вихідного тексту на прикладах Google Translate, Microsoft Translator та Yandex. Для досягнення цієї мети було поставлено такі завдання: визначити найспроможніші алгоритми МП, що використовують сьогодні; порівняти відповідні результати перекладів, зроблених онлайн-перекладачами; проаналізувати типові стилістичні, лексичні та граматичні помилки, що постають у перекладі; виявити переваги та недоліки онлайн-перекладачів; надати рекомендації щодо вдосконалення машинного перекладу.

Методи дослідження. Для вирішення цих завдань було застосовано методи: описовий, порівняльний, метод аналізу, експерименту та лінгвістичної інтерпретації отриманих результатів.

Основні результати дослідження. Машинний переклад художніх текстів був надзвичайно добре опрацьований Яндексом і був цілком прийнятим (за винятком численних граматичних помилок) на платформі Google. Науковці вивели Microsoft Translator, неправильний переклад реалій і згадані помилки роблять його результат не таким зрозумілим, як після використання інших програм. Основні проблеми, які вбачаємо в таких перекладах, пов’язані з тим, що системи залежать від великої кількості високоякісних наборів даних (тобто корпусів текстів для певних мовних пар). Якість цих наборів безпосередньо впливає на якість виводу, у нашому випадку це якість тексту цільової мови. Це можна побачити, порівнюючи середню якість перекладу між системами Google і Microsoft. Перший у середньому робить менше помилок і не має стільки проблем щодо визначення контекстного значення полісемантичної лексеми.

Висновки і перспективи. Уважаємо, що аналізовану проблему певною мірою можна вирішити одним з двох способів: використати знання професійних перекладачів і лінгвістів для складання паралельних корпусів або створити можливість для кожної людини зробити внесок у цей процес навіть у невеликому масштабі. Перший підхід забирає багато часу та праці, але в кінцевому підсумку надає якісніший набір даних, що може привести до подальшого покращення якості перекладу. Другий підхід уже впроваджено всіма трьома основними системами наукового машинного перекладу, але може призвести до сповільнення
Abstract

Background. Today machine translation is one of types of human activity. Machine translation can greatly facilitate global communication, accelerating the translation process, despite the imperfect quality of the source text. Most often the results of online tools require post-editing and can only be effectively used by those who already speak the target language to some extent. The need for a competent translation is growing every year. Today, the search for an algorithm to deliver this quality of translation is one of the most important questions in computer science and linguistics, therefore informing the scientific relevance of this work.

The purpose of this paper is to analyze different approaches to the machine translation systems, their characteristics, efficacy and the quality of their output on the examples from Google Translate, Microsoft Translator and Yandex. To achieve this aim, the following tasks were set: to identify the most capable algorithms of MT in use today; to compare the results of translations made by online translators; to analyze typical stylistic, lexical and grammatical errors that appear in the translation; to identify the advantages and disadvantages of online translators; to provide recommendations for improving machine translation.

Methods. To solve these tasks, we use such methods over the course of this work: descriptive, comparative, analysis, experiment and the method of linguistic interpretation of the results obtained.

Results. Machine translation of belletristic texts was handled exceptionally well by Yandex and was quite acceptable (barring numerous grammatical errors) on Google’s platform. The outlier in this case is Microsoft Translator, whose mistranslation of realia and same aforementioned mistakes make its output much less readable that its competitors. The main problems we see arising from such translations arise from the fact that the systems depend on a large amount of high-quality data sets (i.e., corpora of texts for specific language pairs). The quality of these sets directly influences the quality of the output, which in our case is the quality of the target language text. It can be seen by comparing the average quality of translation between Google’s and Microsoft’s systems. The former one makes less mistakes on average and does not have as many issues in regards to identifying a contextual meaning of a polysemantic lexeme.

We believe that this issue can be fixed to a certain extent one of two ways: hiring professional translators and linguists to compile those parallel corpora or create a possibility for every person to contribute to this process even on a small scale. The first approach would be very time and labor consuming, but would ultimately provide us with a higher quality data set, which may lead to further improvements in MT. The second is already being deployed by all three major NMT systems but may lead slower progression due to lack of quality control and oversight. For us, another potential prospect of this research lies in widening the subject area of texts chosen to reflect the variety of writing styles in use on the Internet right now. Inclusion of texts from confessional, business, and other styles may allow us to highlight more lacunae in the neural network models and to suggest further means of improvement.

Keywords: machine translation, target language, source language, improvement, contextual meaning, communication.
PROPER NAMES PALETTE
IN THE NOVEL «CATHEDRAL» BY OLES HONCHAR

O. Panchenko
Doctor of Philology, Professor,
Head of the Department of Translation and Linguistic Training of Foreigners,
Oles Honchar Dnipro National University
epanchenko2017@gmail.com
orcid.org/0000-0003-2217-5613

P. Gorelova
Student of Translation and Linguistic Training of Foreigners,
Oles Honchar Dnipro National University
polina.gorelova23@gmail.com,
orcid.org/0000-0001-5538-4265

Introduction. The problem dealt with in this article is a relevant one as the role of a name have never stopped being great, the name define functions, appearance, meaning of a character in most books. Oles Honchar has decorated his novels with a number of names with obvious semantics and philosophical background. Oles Honchar’s novels are well-known in many countries and have been translated into numerous languages. This novel-warning is considered one of the greatest creations of the writer and is included into the list «1001 Books You Must Read Before You Die». The relevance of this article, thus, is preconditioned by the necessity to create the theoretical basis for the analysis of the novel content and to set the task of improving its translations into English and other languages. The onomastic problems have been discussed in a number of papers (Bondaletov, 1983; Yermolovych, 2001; Superanskaya, 1973; Jaleniauskiene. 2009 etc.). As far as the proper names in the books by Oles Honchar are concerned, we should mention the dissertation by V. Halych (Halych, 2004) and the article by D. Kudanov where some names (toponyms) are analysed (Kudanov, 2010).

The aim of the article is to analyze the system of proper names in «Cathedral» and their mutual linguistic relations. To achieve this aim it is necessary to fulfil the following tasks: to accumulate the names used in the whole novel; to classify the main types of onyms in the novel; to describe their functions and mutual relations.

Methods and methodology of investigation. The aim of the paper and its tasks define the choice of methods: descriptive which makes it possible to characterize the whole set of the onyms; classifying due to which it is possible to differentiate various types of onyms.

Results and discussion. To denote individual realia in a language there exist proper names, or onyms. Onomastics is engaged in their research. Although interest in proper names, attempts to study them, compiling dictionaries of anthroponyms and toponyms have existed for a long time, but as a scientific field onomastics stood out only in the 20th century (Superanskaya, 1973: 235–236). In 1959, the First Ukrainian Conference of Linguists took place, at which onomastics was defined as a separate scientific discipline. Since the mid-60's of the twentieth century the number of scientific works on onomastic issues published in Ukraine has sharply increased, and their specialization has deepened. From onomastics, which studies proper names, special disciplines that have a narrower subject of study – specific classes of proper names – have finally been singled out. In the second half of the twentieth century, a new direction of research of proper names on the material of the literary text was formed, which arises on the border of onomastics and stylistics, poetics, linguistics of the text, linguistic semantics, semiotics. The term «literary onomastics» was introduced to denote a special section of onomastics, closely related to the problems of literary criticism, genre theory, general trends in cultural life of the era, etc. Onomastics as a linguistic