UDC 81.322.4
DOI https://doi.org/10.32782/2410-0927-2022-16-31

Olga SHEVCHENKO
MBA, Senior Lecturer at the Department of English for Humanities 3, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Peremohy Ave., Kyiv, Ukraine, 03056
ORCID: *0000-0001-6726-7269*

Olga OGURTSOVA
Ph.D. in philology, Associate Professor at the Department of English for Humanities 3, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Peremohy Ave., Kyiv, Ukraine, 03056
ORCID: *0000-0002-0828-3303*

To cite this article: Shevchenko, O., Ogurtsova, O. (2022). Basic Principles and Limitations of Neural Machine Translation. *Current Issue of Foreign Philology*, 16, 215–218, doi: https://doi.org/10.32782/2410-0927-2022-16-31

BASIC PRINCIPLES AND LIMITATIONS OF NEURAL MACHINE TRANSLATION

The article describes some problems of Neural Machine Translation (NMT) and the role of neural networks in the translation process. The mechanism of neural machine translation, its specific features, its differences from other machine translation systems and the system limitations are also analyzed. NMT systems use artificial neural networks that are trained on a large number of pairs of parallel sentences ('parallel corpora'). These networks can read a word or a sentence in the source language and translate them into a target language. However, word matching and breakdown into phrases is no longer needed. This seems to be the main difference between the NMT system and other machine translation systems, such as Rule-based or Statistical MT. In order to create a NMT system one must provide the availability of several million pairs of sentences translated by human translators All modern NMTsystems are equipped with encoder-decoder and 'attention' mechanisms. The unique role of the 'attention' mechanism is to predict subsequent words during the translation process. While focusing on one or more words of the original sentence, it adds this information to the encoded full text. This process is similar to the behavior of a human translator who first reads the entire sentence and then looks at individual source words and phrases already translated or yet to be translated. In spite of its advantages, like fluency, NMT systems have a number of drawbacks. The most frequent are adequacy errors, as well as omissions and additions of content. Transfer of semantic content from the source to the target language often produces mistranslations. The source phrases need to be very clear, coherent and void of ambiguity to prevent low quality output.

Key words: *Neural machine translation (NMT), AI networks, source language, target language, parallel corpora, encoder, decoder, attention mechanism, adequacy error.*

Ольга ШЕВЧЕНКО
МБА, старший викладач кафедри англійської мови гуманітарного спрямування № 3 Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», просп. Перемоги 37, м. Київ, Україна, 03056
ORCID: *0000-0001-6726-7269*

Ольга ОГУРЦОВА
кандидат філологічних наук, доцент кафедри англійської мови гуманітарного спрямування № 3 Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського», просп. Перемоги 37, м. Київ, Україна, 03056
ORCID: *0000-0002-0828-3303*

Бібліографічний опис статті: Шевченко, О., Огурцова, О. (2022). Basic Principles and Limitations of Neural Machine Translation. *Актуальні питання іноземної філології*, 16, 215–218, doi: https://doi.org/10.32782/2410-0927-2022-16-31

ОСНОВНІ ПРИНЦИПИ ТА ОБМЕЖЕННЯ НЕЙРОННОГО МАШИННОГО ПЕРЕКЛАДУ

У статті розглядається специфіка системи нейтронного машинного перекладу, що здійснюється за допомогою нейронних мереж. Розглядається механізм роботи нейтронного машинного перекладу, його відмінності від інших систем машинного перекладу, його складові частини, а також існуючі недоліки системи. Системи нейронного машинного перекладу використовують штучні нейронні мережі, які навчаються з використанням великої кількості паралельних пар речень («паралельний корпус»). Ці мережі здатні читати слова або речення з вихідної мови та перекладати їх на цільову мову. Проте, зіставлення слів та розбивка на фрази вже не потрібні. І це є основною відмінністю між системою NMT та іншими системами перекладу, такими як RbMT (машинний переклад на основі правил), або SMT (статистичний машинний переклад). Щоб створити систему NMT, знадобиться кілька мільйонів пар речень, попередньо перекладених перекладачами. Всі сучасні системи NMT обладнані механізмом «кодер-декодер» та механізмом «уваги». Особлива роль механізму «уваги». полягає в тому, щоб в процесі перекладу передбачати кожне наступне слово. Зосереджуючись на одному чи кількох словах вихідного речення, механізм «уваги» додає цю інформацію до закодованого повного тексту. Цей процес схожий на поведінку людини-перекладача, яка спочатку читає все речення, а потім дивиться на окремі вихідні перекладені або ще не перекладені слова і фрази. Незважаючи на такі переваги, як швидкість, система NMT має також ряд недоліків. Найчастіше спостерігаються помилки, які впливають на адекватність перекладу, а також мають місце пропуски та додавання змісту при перекладі Передача семантичного змісту з вихідної мови на мову перекладу часто призводить до неправильного тлумачення. Вихідні фрази мають бути дуже чіткими, зрозумілими та без двозначності, щоб уникнути неякісного перекладу.

Ключові слова: *нейронний машинний переклад (NMT), вихідна мова, цільова мова, паралельні корпуси, кодер, декодер, механізм.*

For decades, scientists have been trying to develop new translation methods to improve translation efficiency. The idea of using neural machine translation method appeared as far back as in the early 1950s. But it was not until 2013 that this method found its further development and practical application.

In 2013 N. Kalchbrenner and P. Blunsom from Oxford University proposed a new translation method which can be regarded as the birth of the Neural Machine Translation (NMT). In 2016, Google introduced the Neural Machine Translation System (GNMTS) to improve the performance of Google's translation service. Google started using NMT, which replaced Phrase-based Machine Translation system (a variety of Statistical Machine Translation (SMT)) previously used by the Google Translate service.

The Neural Machine Translation System (NMTS) uses a new approach which is different from previous methods of machine translation and is very similar to the work of the human brain. NMTS uses a deep learning technology as part of Machine Learning and Artificial Intelligence (AI). This technique allows the system to imitate human behavior through learning. The NMT system has clearly demonstrated its capability of self-learning which resulted in further improvement of the efficiency of neural machine translation.

In recent years a great number of researchers have been actively engaged in studying and developing Neural Machine Translation as a new approach to machine translation. This method was first proposed by N. Kalchbrenner and P. Blunsom (Kalchbrenner et al., 2013), I. Sutskever (Sutskever et al., 2014) and D. Bahdanau (Bahdanau et al., 2015). It differed from previously known popular MT methods such as Statistical Machine Translation (SMT) which is based on the analysis of existing translations from bilingual text corpora and Rule-based Machine Translation (RbMT) which is a matching translation system based on linguistic information about the source and target languages.

This paper investigates the main principles of the NMT and some of its existing limitations. All NMT systems comprise basically three main components: the encoder, the attention mechanism, and the decoder.(Cho K. et al., 2014). The decoder-encoder mechanism is an indispensable part of the system and is used to predict subsequent words with certain properties according to the sentence structure. The input sentence is first encoded into an abstract set of numbers and the sentence is analyzed in terms of grammar (context) and segments (words) contained therein. Then the decoder, having "read" the sentence to the end and simultaneously from left to right and from right to left, begins to decode and predict words. Each predicted word is used to predict the next word and so on. Finally, the target sentence is generated. During this process, the attention mechanism plays an important role as it helps the decoder to analyze different parts of the input sentences and helps the system to memorize long sentences. The following diagram (Fig. 1) shows the neural machine translation model:

In order to create an effective NMT system it is required to have a huge bilingual corpus that contains text in the source language with the equivalent translation in the target language (Shen G., 2011).

In the process of translation the source word is analyzed in terms of its morphological structure, grammatical category and grammatical role in a sentence. After that the system finds the same word representation in the target language.

NMT system is trained on large number of pairs of sentences and also it uses 'knowledge' obtained previously. Word matching and breakdown into phrases are no longer needed. Thus, for a more precise choice of a word for translation, the context of the entire original sentence is used, as well as the context with all previously predicted words. And this is more like how a human translation occurs. In the process of training of a neural system, there is a constant comparison of each predicted word with the 'correct' word, i. e. with the word that is used in a target training sentence.. If the predicted word does not coincide with it, then the parameters of the model are updated. This process is repeated multiple times and, as a result, the neural system goes through a multitude of pairs of sentences which are mixed in random order several times. The process stops when there is no further improvement in word prediction.

According to some estimates NMT outperforms other types of machine translation systems in fluency and accuracy. It also gives better results in terms of inflection and reordering (Bentivogli L. et al., 2016).

In spite of the obvious advantages NMT has also certain weaknesses.

The translation quality deteriorates in case the sentence segments are too long. NMT may become totally ineffective if a sentence comprises more than 20 words. Satisfactory results in solving this problem can be achieved by using a long-sentence preprocessing technique (Ha N. T. et al., 2019). The method uses the extraction of bilingual phrases and creation of bilingual corpus based on these phrases.

NMT system has difficulty with translating rare and low-frequency words. Also, its overall performance remains relatively slow and requires post-editing.

In many cases NMT remains domain-specific which means that the system is trained to translate texts from specific domains, e. g. business, e-commerce, legal, medical, etc. Besides the need to prepare large amounts of data for training, a significant computing power is required to train the system.

Neural machine translation aims at building a single neural network which can be trained to achieve the best translation results.

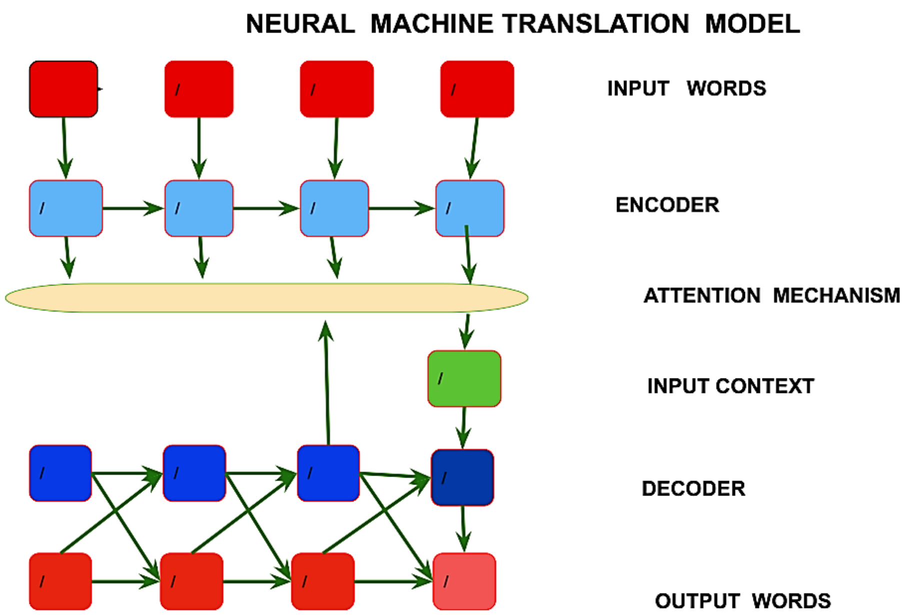

Fig. 1.

Table 1

NMT Adequacy Errors

№	Context	Adequacy error	Incorrect translation (by Google)	Correct translation
1	The kitten is lovely. I like it	Gender	Кошеня милий. Мені це подобається.	Кошеня миле. Воно мені подобається
2	The secret is out. I am in hot water	Idioms	Секрет розкритий. Я в гарячій воді	Секрет розкритий. В мене проблеми
3.	No. one failed to come	Negative pronouns	Ніхто не прийшов	Всі прийшли
4	He was dined and wined in the best houses of Paris	The Passive Voice	Його обідали і винували в найкращих будинках Парижа	Його пригощали обідом та вином в найкращих будинках Парижа
5	The bed was not slept in	The Passive Voice	Ліжко не спали	В ліжку не спали
6	Світає! Дощить!	Impersonal sentences	Dawns! To rain!	It dawns! It rains!
7	Цю статтю написав Петро Ворона	Proper names	This article was written by Peter the Crow	This article was written by Petro Vorona

An important prerequisite for starting the NMT process is the availability of a dataset (parallel corpora) in both source and target languages.

The neural machine translation can be of two types: the classical NMT and NMT with attention mechanism. Both of them were studied and the results showed that NMT with the attention mechanism has significantly outperformed the classical NMTin its performance (Tan Z. et al., 2020).

Nowadays NMT is widely used to increase fluency and accuracy in translation. However, adequacy still remains a major problem in NMT In spite of its obvious advantages over other types of translation systems, (Kong X. et al., 2019). NMT generates numerous adequacy errors during the translation process. It means that a great deal of post-editing work on the part of human translators is still required (Ustaszewski, 2019). There can be different types of adequacy errors made by the NMT system including omissions, additions or mistranslations. Some typical examples of adequacy errors produced by the NMT system are given in Table 1.

In conclusion it should be noted that NMT in spite of its growing popularity is still facing some problems which have to be overcome to make this method more effective and flawless. The major factors for future development of NMT are the growing volumes of big data and online content which requires an efficient and robust machine translation system able to cope with the emerging challenges.

REFERENCES:

1. Kalchbrenner, N., Blunsom, P. (2013). Recurrent Continuous Translation Models Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 1700–1709.

2. Sutskever, I., Vinyals, O., Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks, arxiv.org/pdf/1409.3215.pdf, 1–9.

3. Bahdanau, D., Cho, K., Bengio, Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate, arxiv.org/pdf/1409.0473.pdf

4. Ustaszewski, M. Exploring Adequacy Errors in Neural Machine Translation with the Help of Cross-Language Aligned Word Embeddings, Proceedings of the Second Workshop on Human-Informed Translation and Interpreting Technology (2019). (HiT-IT 2019), 122–128.

5. Shen, G. (2011). Corpus-based Approach to Translation Studies. Cross Cultural Communication. 6 (4), 181–187.

6. Bentivogli, L., Bisazza, A., Cettolo, M., Federico, M. 2016, Neural versus phrase-based machine translation quality: a case study. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP, 257–267.

7. Ha, N. T., Huyen, N. T. M. (2019). Long Sentence Preprocessing in Neural Machine Translation. 2019, IEEE-RIVF.

8. Tan, Z., Wang, S., Zonghan Yang, Z., Chen, G., Huang, X., Sun, M., Liu, Y. (2020). Neural Machine Translation: a Review of Methods, Resources and Tools. arxiv.org/pdf/2012.15515.pdf, 1–20.

9. Kong, X., Tu, Z., Shi, S., Hovy, E., Zhang, T. (2019). Neural Machine Translation with Adequacy-Oriented Learning. 33rd AAAI Conference on Artificial Intelligence (AAAI-19), 6618–6625.

10. Cho, K., Merrienboer, B., Bahdanau, D., Bengio, Y. (2014).On the Properties of Neural Machine Translation: Encoder-Decoder Approaches. arxiv.org/abs/1419.12/59