9. Shchukin, A. N. Moskovkin, L. V. (2010), *Reader on the methods of teaching Russian as a foreign language* [*Khrestomatiya po metodike prepodavaniya russkogo yazyka kak inostrannogo*], Russkiy yazyk, Moscow, pp. 132–133.

Hanna M. TRUBA,
Candidate of Philological Sciences, Associate Professor, Department of the Applied Linguistics, Odessa I. I. Mechnikov National University; 24/26 Francuzskiy blvd., Odessa, 65058, Ukraine; tel.: +38 063 2366706;
e-mail: 3182009060@ukr.net; ORCID ID: 0000–0001–9944–0476

SUGESTOLOGY AS ONE OF THE METHODS OF STUDYING THE UKRAINIAN LANGUAGE AS A FOREIGN: REGARDING THE APPLICATION OF G. K. LOZANOV'S SUGESTOLOGICAL METHOD

Summary. The article actualizes the problem of studying the Ukrainian languages a foreign language and the ways to optimize this process. This study of methods of teaching foreign languages allows linguists, evaluating the advantages and disadvantages of each methods, to use the most convenient and effective background for the formation of a grammatical and lexical basis for the creation of educational materials. G. K. Lozanov's suggestive method allows optimizing the learning process in all aspects. The main thesis of which is a relaxed atmosphere during class, which contributes to the natural way of assimilating the material without much effort and coercion. It is the absence of psychological pressure on the student in the "teacher-student" format that free sup additional memory reserves. The ***purpose*** of the article is not only to illuminate the suggestive method of studying the Ukrainian language as a foreign, but also to try to implement it according to the needs of the course. The ***object*** of the article is the studying of a suggestive approach, and the ***subject*** of research is the introduction of this method into the learning process. The ***relevance of the topic*** lies in the fact that the process of improving the methods of studying the Ukrainian language as a foreign language never stops, especially since it always requires more effective and more modern, the creation of more effective teaching materials. Among these search methods in the study of this topic, one can single out an analysis of vocabulary definitions, a descriptive method, an interpretation method using observation and generalization techniques. The ***practical value*** of the work lies in the fact that its results can be used in the course of lecturing and conducting seminars on the courses "Methods of Studying Foreign Languages", "Ukrainian as a Foreign Language", as well as in the practical study of Ukrainian as a foreign language.

Key words: Ukrainian as a foreign language, ignorantly didactic methods, G. K. Lozanov's suggestological method.

Статтю отримано 28.03.2021 р.

DOI: 10.18524/2307–4558.2021.35.237789

UDC 81'322.2/.3'271.14/.16'367

Olena POZHARYTSKA,
candidate of Philological Sciences (PhD), Associate Professor at the Chair of English Grammar,
Odessa Mechnykov National University; 2 Dvoryanska street, Odessa, Ukraine, 65082; tel. +380509632062;
e-mail: grammarlena@onu.edu.ua; ORCID ID: 0000–0003–4820–8129

Kyrylo TROITSKYI,
undergraduate student at the faculty of Romance-Germanic Philology, Odessa Mechnykov National University; 2 Dvoryanska street, Odessa, Ukraine, 65082; tel. + 380986567113; e-mail: troickiykirill@gmail.com;
ORCID ID: 0000–0002–3395–2724

DIGITAL TECHNOLOGIES FOR GRAMMATICAL ERROR CORRECTION: DEEP LEARNING METHODS & SYNTACTIC N-GRAMS

Summary. The ***object*** of this article is automated grammatical error detection as a field of linguistics. The ***subject*** of the article is the variety of methods and techniques used in grammatical error detection along with their applications and evaluation. The article considers the most productive methods used in the field of grammatical error detection and correction in computational linguistics. The ***purpose*** of the article is to review major rule-based and deep learning methods used in the area, evaluate and compare them. The ***methods*** of research used in this article are data analysis, description of abstract computational models and observation of their performance. The ***article offers and defines*** a model based on syntactic n-grams, describes the ways of its implementation and the necessary pre-processing steps for the model to work. The particular error types that the model is capable of detecting are noun-verb agreement errors, preposition errors, noun number errors and some article error types. Also, the article analyses a recent model based on the transformer architecture — GECToR (Grammatical Error Correction: Tag, Not Rewrite). This deep learning model is aimed at detecting and correcting much more complicated errors, including those that rely on extralinguistic realia. Additionally, it is very useful because in contrast to other models that just replace incorrect tokens without explanations, GECToR assigns labels that can be further interpreted for educational purposes. Also, ***conclusions*** were made about the advantages and disadvantages of the described models that were discovered after their practical implementation.

During the evaluation of the aforementioned models based on the BEA 2019 shared task, the following **results** were achieved: the model based on syntactic n-grams obtained the $F_{0.5}$ measure of 7,6 %, and the GECToR model's $F_{0.5}$ score was evaluated as 66,7 %. These ***results*** give an almost nine-fold increase in performance of deep learning methods, such as GECToR compared to rule-based methods, such as syntactic n-grams.

Key words: syntactic n-grams, computational linguistics, grammatical error correction, transformer, rule-based methods, deep learning methods.

Problem statement. Over the last few years, computers have become an essential part of scientific research not only in traditionally quantitative fields like mathematics, physics and genetics, but also in humanities such as language studies, which prompted the appearance of an entirely different field of linguistics — natural language processing (NLP). One of the areas of NLP research is grammatical error correction (GEC) that tries to introduce and improve methods of automatically finding errors in texts, which may prove useful both in detecting native speakers' typos and in helping language learners deal with linguistic challenges they face.

The topicality of this work and general research in the field of GEC is motivated by the growing interest towards automated methods in education as well as by the newness of the area. Though such famous researchers as Martin Chodorow, Claudia Leacock, Zheng Yuan, Kostiantyn Omelianchuk and others have studied the problems of Grammatical Error Correction, not much attention has been given to the comparison between various generations and types of methods used in GEC.

The object of this research is automated grammatical error correction in the English text.

The subject of the paper is a variety of computer techniques used in grammatical error correction in the English text.

The objective of this work is a contrastive evaluation of the two major rule-based and deep learning approaches used as instruments for grammatical error correction in the English text.

The immediate tasks of the research are to give a comparatie survey of the existing computer programmes used for eliminating grammatical errors in English texts; to give a practical analysis of computer models chosen; to find out most effective techniques of eliminating grammatical mistake and outline perspectives for the future study of computer orientated methods of education.

The methods of research used in this paper combine general philosophical methods used in linguistic analysis with specific lingual approaches to the object of investigation. The basic research platform of this paper is anthropocentrism, which presupposes studying language phenomena through the prism of human-beinh. In our case, this approach is realized indirectly by means of computer techniques created by man and used to analyse texts written by humans. The domineering method of investigation used in this paper is data analysis as understood in NLP. The main results of the paper were obtained by using quantitative and qualitative methods manifesting comparative validity of the computer models studied.

Presentation of the main material. Grammatical error correction methods can be roughly divided into two types depending on the principles elaborated in the computer model. These are rule-based methods and machine learning methods. Machine learning methods are often considered a separate field differing from statistic non-machine learning methods. They presuppose the use of either artificial neural networks or statistical machine translation [13, pp. 19–22].

Rule-based systems rely on grammatical rules of a certain language introduced by man. They are based on the fixed rules of syntax, morphology, and traditional santics of a given language. Writing all the rules for languages is time consuming and laborious [10, p. 2], which makes rule-based methods inefficient or sometimes even impossible in cases of complicated grammatical structures or relying on semantic data of the text. Rule-based approaches handle those error types that can be described in a simple way. For example, a regular expression is sufficient for identifying a mistake which occurs when a modal is (incorrectly) followed by a conjugated verb [6, p. 74].

Machine learning approaches take advantage of the corpora annotated by humans "to train" a certain language model which accumulates the algorithms of general lingual structures traditional for the given language [10, p. 2]. Speaking about grammatical error correction, machine translation methods can be used to translate the potentially wrong sentences into the correct ones.

A module in Python programming language was created in order to compare and contrast the capacity and efficiency of both machine learning and rule-based methods, mentioned above. The elaborated module used SpaCy library for text pre-processing [4]. Thus, the pre-processing stages in our paper rest on dependency parsing mostly. Dependency parsing is the process of describing a sentence in terms of lexemes represented in it and binary grammatical relations between those lexemes [5, p. 280]. For example, in the sentence *"I prefer the morning flight through Denver"* the SpaCy parser managed to correctly identify that the word *"I"* is the subject of the sentence and the word *"flight"* is the direct object. Other pre-processing techniques used in the paper are word and sentence tokenization, named entity recognition, part-of-speech tagging and word embedding. Still, we believe that the SpaCy parser efficiency in automated grammatical error correction could be improved with the help of rule-based branch of methods as those will grant a sufficient theoretical basis for the purpose in view.

To test the efficiency of the rule-based branch of methods, syntactic n-grams described by Sidorov et al. [11] were implemented in the computer model suggested in this paper. Syntactic n-grams allow finding

a proper position of a lexeme depending on the four basic syntactic modes traditionally singled out in the English language: the structures of predication, complementation, modification and coordination [8, pp. 201–206]. Hence, the computer model takes into consideration not a lineal organisation of the sentence, but is based on the interdependence of the lexemes used in it. For example, in the aforementioned sentence *"I prefer the morning flight through Denver"* one of the syntactic n-grams would be "I prefer flight" as these are located next to each other according to the dependency parse of the sentence. In other words, our approach is a computer refraction of the Tesnière tree used in the immediate constituents analysis [7].

As our investigation has revealed, the following error types could be detected with the help of syntactic n-grams:

Preposition errors: "*There's a growing need of engineers in ...*", where "for" should have been used instead of "of". In order to find these errors, a large corpus of text was parsed in search of the following prepositional pattern:

<head word> + <preposition> + <pos tag of the dependent word>

Here, the head word of some token is a word which has a dependency connection pointing to this token. A dependent word, then, is a word which has a dependency connection pointing to it from some other token. The pattern uses the part-of-speech tag of the dependent word in order to generalise better.

As stated above, we used a large body of text to extract these patterns. Namely, it was a small subset of modern literature (19 books), and a corpus of twitter text. This kind of model was described by Grigori Sidorov et al. [11, p. 4]. However, given the poor results of the original paper by Sidorov, our current model adds some new functionality. Firstly, it takes into account that the head word may be a named entity, in which case it replaces the text of the token with the NE tag. Then, in parallel with the described dataset, the model gathers another dataset in the following form:

<head word/NE tag> + <preposition> + <dependent word/NE tag>

This is done to account for the cases when there are several possible options following from the first dataset, in which situation the model will try to search for the exact pattern in the second dataset. If not found, the preposition with the biggest number of occurrences in the first dataset will be chosen.

The first dataset counts 211,971 entries, while the second dataset is made up by 601,266 entries.

Subject-verb agreement errors: *"The news are good"*, where "is" should be used instead of "are". To deal with this kind of problem we also have to use syntactic data because we cannot rely on the immediate valency of lexemes due to possible parenthetical clauses and adverbial modifiers. Syntactic n-grams allow finding dependency relations directly between the subject and the verb of the word sequence.

The basic rules used in the described model are strict and unambiguous:

1) If the subject is in plural and the verb's tag is "VBZ" (verb, 3rd person singular present), then change the verb so that it has the tag "VB" (verb, base form).

2) If the subject is in singular and the verb's tag is "VB", then it has to change to "VBZ".

The above is true in cases with one simple tense predicate only. If there is a modal or auxiliary verb within the compound predicate, the set of rules changes as given below:

1) The two basic rules apply to the modal/auxiliary.

2) If the sentence contains a modal verb or the auxiliary 'do'/'does', the dependent verb should have a "VB" tag.

3) If the sentence contains the auxiliary "have", the main verb should have a "VBN" tag (past participle) [11, p. 4].

This set of rules, however, relies on a few assumptions that might not be true in the real-world data. One such assumption is that the modal is chosen correctly. If this is not true, then the correction will probably also be erroneous.

Certain **article errors**: *"This is a good advice"*, where the article shouldn't be used. Some article error types can be detected by checking the countability of the noun the article refers to. Taking into consideration that some nouns can be both countable and uncountable (depending on the context), a dataset of 198 nouns with the most unambiguous meanings was created for the discussed model.

Apart from the methods that make use of syntactic n-grams, some other error types can be detected using the outlined rules:

The choice of "a/an": *"He is a honest person"*, where "an" should be used instead of "a". This error type can be detected by extracting the phonetic transcription of each word that has an indefinite article in front of it. The described model extracts phonetic transcriptions of words with the help of CMU Pronouncing Dictionary for US English, which provides transcriptions in the ARPABET form. After getting the phonetic transcription of the word following the indefinite article, the variant 'a' is chosen if the next sound is a consonant, and vice versa.

Noun number errors: *"The advices you gave to me are ...",* where an uncountable noun is pluralised. This can also be partially dealt with by using a dataset of uncountable nouns and checking that all plural nouns are countable.

The statistic model is compared with one of the leading machine learning models for grammatical error correction according to BEA 2019 shared task — GECToR (Grammatical Error Correction: Tag, Not Rewrite) [2]. A neural network architecture called "the Transformer" is the basis of GECToR. The

Transformer architecture was initially introduced in 2017 [12]. The special feature of transformers is "attention", which allows the model to pay more focus on relevant words and sequences, and neglect other, unimportant ones [12, pp. 1–2]. One of the language models GECToR uses for GEC is BERT (Bidirectional Encoder Representations from Transformers), which makes great use of the transformer architecture [3]. BERT-based systems try to predict probabilities of tokens in a sentence, considering that some tokens in the sentence are masked.

The way GECToR uses BERT is the following: it tries to predict grammatical correction tags in words instead of making the corrections themselves [9, p. 2]. This is very useful because in contrast to other models that just replace incorrect tokens without explanations, GECToR assigns labels that can be further interpreted for educational purposes.

The shared task used to contrast the two models is BEA-2019 [2] as it is the most comprehensive recent dataset for Grammatical Error Correction.

The metrics that are used to evaluate performance of the models are precision, recall and $F_{0,5}$ measure. The rule-based model obtained the following results based on the ABCN development dataset: precision — 18,2 %, recall — 2,3 %, $F_{0,5}$ measure — 7,6 %.

GECToR was also tested on the same dataset and obtained the following scores: precision 70,6 %, recall 54,8 %, $F_{0,5}$ score 66,7 %. So, as follows from the results of our research, GECToR model can perform with an almost nine-fold increase. The scores were calculated with the help of Error Annotation Toolkit (ERRANT) v.2.2.3 [1].

Conclusions. The main achievement of our paper is that it implements syntactic n-grams in automated grammatical error detection combining them with SpaCy, which is the most popular and efficient library for text parsing so far. However, while SpaCy only creates dependency trees, the module suggested here also extracts n-grams from them. Thus, it has been concluded that deep learning models, such as the Transformer, can turn out to be particularly useful in tasks related to grammatical error correction. It is especially evident when comparing them to rule-based methods that have obtained efficiency scores almost 9 times lower than those in the transformer-based GECToR model as our research has shown. This underlines the importance of research and development of Transformer-based models, as well as their vast potential for a large variety of applications, including Grammatical Error Correction. We believe the observations given here to be productive both for linguistic text analysis and foreign language teaching. The fact that deep learning methods gradually approach human-level performance should be taken into consideration when developing systems for language learning or any kind of error correction software.

References I

1. Bryant C., Felice M., Briscoe T. **Automatic annotation and evaluation of error types for grammatical error correction.** *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, vol. 1. Vancouver, Canada: Long Papers, 2017.

2. Bryant C., Felice M., Andersen Ø. E., and Briscoe T. The BEA-2019 Shared Task on Grammatical Error Correction. *Proceedings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications (BEA-2019)*. Florence, Italy: Association for Computational Linguistics, 2019, pp. 52–75.

3. Devlin J., Ming-Wei Chang, Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2019, 16 p.

4. Explosion. (2016). SpaCy: Industrial-Strength Natural Language Processing. [Online] Available: https://spacy.io/

5. Jurafsky D., James H. Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Draft of December 30, 2020, 623 p. [Online]. Available: https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf.

6. Leacock C., Gamon M., Brockett C. User Input and Interactions on Microsoft Research ESL Assistant. *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications*. Boulder, Colorado: Association for Computational Linguistics, 2009.

7. Tesnière L. Éléments de syntaxe structurale. Paris, France: Klincksieck, 1976, 674 p.

8. Морозова И. Б. Элементарная структура предикации как основа определения грамматического статуса предложения. Modern researches in philological sciences : collective monograph. Riga : Izdevnieciba "Baltija Publishing", 2020, pp. 200–217.

9. Omelianchuk K., Atrasevych V., Chernodub A. and Skurzhanskyi O. GECToR — Grammatical Error Correction: Tag, Not Rewrite, 2020, 8 p. [Online] Available: https://arxiv.org/abs/2005.12592

10. Rauf, S., Saeed, R., Khan, N. S., Habib, K., Gabrail, P. and Aftab, F. Automated Grammatical Error Correction: A Comprehensive Review. *NUST Journal of Engineering Sciences, vol. 10*, 2017.

11. Sidorov G., Gupta A., Tozer M., Català D., Catena A. and Fuentes S. Rule-based System for Automatic Grammar Correction Using Syntactic N-grams for English Language Learning (L2). *CoNLL Shared Task*, 2013, pp. 96–101.

12. Vaswani, A. et al. Attention is All You Need, 2017, 15 p. [Online]. Available: https://arxiv.org/abs/1706.03762

13. Zheng Y. Grammatical error correction in non-native English. Cambridge, United Kingdom: Cambridge University Press, 2017, 145 p.

References II

1. Bryant, C., Felice, M. and Briscoe, T. (2017). Automatic annotation and evaluation of error types for grammatical error correction. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, No. 1*
2. Bryant, C., Felice M., Andersen, Ø. E. and Briscoe, T. (2019). The BEA-2019 Shared Task on Grammatical Error Correction. *Proceedings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications,* pp. 52–75
3. Devlin, J., Ming-Wei, C., Lee K. and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. [Online]. Available at: https://arxiv.org/pdf/1810.04805.pdf
4. Explosion. (2016). *SpaCy: Industrial-Strength Natural Language Processing.* Retrieved from https://spacy.io/
5. Jurafsky, D. and Martin, J. H. (2020). Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. [Online]. Available at: https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf
6. Leacock, C., Gamon, M. and Brockett, C. (2009). User Input and Interactions on Microsoft Research ESL Assistant. *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications.*
7. Tesnière, L. (1976). Éléments de syntaxe structurale. Paris, France: Klincksieck.
8. Morozova, I. (2020). The elementary structure of predication as the basis for defining the grammatical status of a sentence ["Elementarnaya struktura predikatsii kak osnova opredeleniya grammaticheskogo statusa predlozheniya"]. *Modern researches in philological sciences : collective monograph*. Riga, Latvia: Baltija Publishing.
9. Omelianchuk, K., Atrasevych, V., Chernodub, A. and Skurzhanskyi O. (2020). GECToR — Grammatical Error Correction: Tag, Not Rewrite. [Online]. Available at: https://arxiv.org/abs/2005.12592
10. Rauf, S., Saeed, R., Khan, N. S., Habib, K., Gabrail, P. and Aftab, F. (2017). Automated Grammatical Error Correction: A Comprehensive Review. *NUST Journal of Engineering Sciences, vol. 10.*
11. Sidorov, G., Gupta, A., Tozer, M., Català, D., Catena, A. and Fuentes, S. (2013). Rule-based System for Automatic Grammar Correction Using Syntactic N-grams for English Language Learning (L2). *CoNLL Shared Task*, pp. 96–101
12. Vaswani, A. et al. (2017). Attention is All You Need. [Online]. Available at: https://arxiv.org/abs/1706.03762
13. Zheng, Y. (2017). Grammatical error correction in non-native English. Cambridge, United Kingdom: Cambridge University Press, 145 p.

ПОЖАРИЦЬКА Олена Олександрівна,
кандидат філологічних наук, доцент кафедри граматики англійської мови, Одеський національний університет ім. І. І. Мечникова, вул. Дворянська, 2, Одеса, Україна, 65082; тел. +380509632062;
e-mail: grammarlena@onu.edu.ua; ORCID ID: 0000–0003–4820–8129

ТРОЇЦЬКИЙ Кирило Володимирович,
студент бакалаврату на факультеті романо-германської філології, Одеський національний університет ім. І. І. Мечникова, вул. Дворянська, 2, Одеса, Україна, 65082; тел. +380986567113;
e-mail: troickiykirill@gmail.com; ORCID ID: 0000–0002–3395–2724

ВИКОРИСТАННЯ ЦИФРОВИХ ТЕХНОЛОГІЙ ДЛЯ ВИПРАВЛЕННЯ ГРАМАТИЧНИХ ПОМИЛОК: СИНТАКСИЧНІ N-ГРАМИ ТА МЕТОДИ ГЛИБИННОГО НАВЧАННЯ

Анотація. ***Об'єкт*** статті — автоматизоване виправлення граматичних помилок як галузь лінгвістики. ***Предмет*** статті — різноманітність методів та технологій, які використовуються у виправленні граматичних помилок, а також можливості їх використання та оцінка. У статті розглянуто найбільш продуктивні методи, що застосовуються у галузі виявлення та виправлення граматичних помилок в комп'ютерній лінгвістиці. ***Мета*** статті полягає у маніфестації ефективності застосування комп'ютерних програм задля виявлення граматичних помилок в англомовному тексті. Дослідницькі ***методи***, використані у статті: аналіз данних, опис абстрактних комп'ютерних моделей та спостереження над їх продуктивністю. У статті ***розглянуто*** комп'ютерну модель для виявлення та визначення граматичних помилок, засновану на синтаксичних n-грамах, дано її визначення, описано шляхи її реалізації та етапи попередньої обробки даних, необхідні для роботи моделі. Встановлено, що конкретними типами помилок, які залучена комп'ютерна модель може виявити, є помилки підмето-присудкового узгодження, помилки у виборі прийменника, числа іменників, а також деякі типи помилок, пов'язані з використанням артиклю. Також у статті проаналізовано іншу модель, засновану на архітектурі трансформера — GECToR (Grammatical Error Correction: Tag, Not Rewrite). Ця модель глибинного навчання спрямована на виявлення та виправлення набагато складніших помилок, у тому числі тих, що пов'язані з екстралінгвістичними реаліями. Крім того, вона є доволі корисною, оскільки, на відміну від інших моделей, які просто коригують неправильні слова без пояснень, GECToR призначає теги, які можна додатково інтерпретувати для навчальних цілей. У процесі аналізу зроблено ***висновок*** про переваги та недоліки розглянутих моделей та методів, що були виявлені після їх практичної реалізації.

Під час оцінки продуктивності вищезазначених моделей на основі спільного завдання BEA 2019 були отримані наступні ***результати***: модель, заснована на синтаксичних n-грамах, отримала показник $F_{0,5}$ 7,6 %, а оцінка $F_{0,5}$ моделі GECToR визначила її ефективність як 66,7 %. Отримані дані свідчать про майже дев'ятикратну перевагу ефективності методів глибинного навчання (типу GECToR) порівняно з методами, заснованими на правилах (типу методу синтаксичних n-грамів).

Ключові слова: синтаксичні n-грами, комп'ютерна лінгвістика, виправлення граматичних помилок, трансформер, системи, засновані на правилах, методи глибинного навчання.

Статтю отримано 08.05.2021 р.