

Lessons of Secondary School Teachers: From Automatic Speech Analysis to the Markers of Effective Teaching Practices

Ivan Mamaev¹, Maria Khokhlova², Maria Dayter³

¹ Saint Petersburg State University, Saint Petersburg, Russia;
Baltic State Technical University "Voenmeh" named after D.F. Ustinov, Saint Petersburg, Russia
E-mail: i.mamaev@spbu.ru, mamaev_id@voenmeh.ru
ORCID: <https://orcid.org/0000-0003-3362-9131>

² Saint Petersburg State University, Saint Petersburg, Russia
E-mail: m.khokhlova@spbu.ru
ORCID: <https://orcid.org/0000-0001-9085-0284>

³ Saint Petersburg State University, Saint Petersburg, Russia
E-mail: mvdayter@gmail.com
ORCID: <https://orcid.org/0000-0003-3653-6774>

DOI: 10.26907/esd.19.1.03

EDN: HMRUWD

Submitted: 21 September 2023; Accepted: 18 December 2023

Abstract

The problem of pedagogical discourse as a speech behavior form is a cutting-edge linguistic area. Within its framework, it is necessary to identify some lexical and semantic components that form a certain rhetorical and pedagogical ideal. To date, such studies are carried out manually. This paper describes the automatic study of pedagogical discourse. As part of the experiment, statistically significant discourse markers and patterns are extracted from the corpus of teachers' speeches, such markers characterizing both general trends in teaching methods and idiosyncrasy characteristics of a particular teacher. The results of the marker analysis make it possible to form a preliminary list of speech patterns that beginner teachers can use.

Keywords: pedagogical discourse, corpus linguistics, discourse markers, collocations, NLP.

Уроки учителей в средней школе: от автоматического анализа речи к маркерам результативных учителей

Иван Д. Мамаев¹, Мария В. Хохлова², Мария В. Дайтер³

¹ Санкт-Петербургский государственный университет, Санкт-Петербург, Россия;
Балтийский государственный технический университет «Военмех» им. Д.Ф. Устинова,
Санкт-Петербург, Россия
E-mail: i.mamaev@spbu.ru, mamaev_id@voenmeh.ru
ORCID: <https://orcid.org/0000-0003-3362-9131>

² Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
E-mail: m.khokhlova@spbu.ru
ORCID: <https://orcid.org/0000-0001-9085-0284>

³ Санкт-Петербургский государственный университет, Санкт-Петербург, Россия
E-mail: mvdayter@gmail.com
ORCID: <https://orcid.org/0000-0003-3653-6774>

DOI: 10.26907/esd.19.1.03

EDN: HMRUWD

Дата поступления: 21 сентября 2023; Дата принятия в печать: 18 декабря 2023

Аннотация

Изучение особенностей педагогического дискурса как формы речевого поведения является активно развивающимся направлением в лингвистике и педагогике. В рамках этого направления рассматривается ряд лексических и семантических компонентов речи учителя, формирующих определенный риторический и педагогический образ, и такие исследования обычно проводятся вручную. В данной статье описываются результаты, полученные при автоматическом анализе педагогического дискурса. В ходе эксперимента из расшифровок речи учителей извлекались статистически значимые дискурсивные маркеры и паттерны, характерные для педагогического дискурса в целом, а также идиостилистические характеристики речи конкретного учителя. Результаты анализа автоматически извлеченных маркеров позволяют сформировать предварительный список речевых паттернов, которые можно рекомендовать использовать начинающим учителям.

Ключевые слова: педагогический дискурс, корпусная лингвистика, дискурсивные маркеры, коллокации, автоматическая обработка текста.

Introduction

Teaching practices have great importance because students' well-being and academic achievements strongly depend on how successive teaching practice of a particular teacher is. Despite this, there is a lack of studies of what a particular lesson consists of and how the teacher behaves. However, there are a plethora of methodological and normative regulations that are to be implemented by teachers (Sergomanov et al., 2023). There are several questions to be answered, such as *What behavioral patterns that are intrinsic for a teacher sustain a class as a community?* It should be also noted that being inevitably digitized (largely due to the COVID-19 pandemic), education goes through a process of massive transformation. The whole phenomenon of education, its space and organization are being redefined (Sergomanov & Bysik, 2022). There are four ways to define effective teaching:

- an importance of results for students' activity;

- individualization of goals, which having been chosen, planned and achieved, become results;
- personalization of education;
- optimization of resources (Semenov, 2020).

In this research, we follow the first definition of effective teaching. The effective teachers are said to share some common features: linguistic markers and patterns, metaphors and means to regulate the mental condition of students (Sergomanov et al., 2023). In the aforementioned study, teaching practices were explored by analyzing videotapes of lessons that were given by effective teachers in different Russian cities. There were 74 lessons conducted by 11 teachers. The linguistic, discourse and social levels were analyzed.

There are many tools that help measure students' academic progress and their social well-being (Qridi¹, Navigated Learning Collaborative², etc.), which means that there is a certain amount of data that can be studied using mathematical methods and machine learning.

Related works

The importance of using natural language processing (NLP) procedures for texts of different genres cannot be overemphasized. Its ability to process large volumes of data, identify patterns, and extract meaningful insights can help individuals and organizations make informed decisions that are more efficient, objective, and relevant. Extracting discourse units is not an exception. The simplest way to detect significant phrases, collocations or keywords is to apply a statistical approach. Association measures (Evert, 2004) do not lose their importance in relation to a task of collocation extraction. More elaborate statistical methods involve calculating frequencies using stop-words or delimiters, for example, RAKE that is used for keyword extraction (Rose et al., 2010).

The discourse analysis attracts the attention of researchers because it allows us to understand what strategies can be most successful, as well as to demonstrate how to build effective teaching. Wang and Han (2015) performed a quantitative analysis of the speech of teachers and students at high schools focusing on students' interaction which proved to be low. They calculated the ratio of teachers' questioning, as well as the number of open and closed questions. The paper concluded that students participate in the learning activities passively and professors should encourage them to take initiative.

Basically, papers dwell on English lessons. For instance, Lee (2020) focused on the classroom discourse using an ethnographic approach and a corpus-based linguistic method. The study described the use of the construction *you know* in 24 English lessons taught by four highly experienced teachers. The author studied the frequency of the *you know* phrase in teachers' speech, described its functions, and showed the specific usage of *you know* in the classroom speech compared to the other types of English-language discourse. Husna et al.'s study (2022) analyzed transcripts of teachers' speech to describe speech acts produced by EFL teachers. Assertive, affirmative, and informative speech acts were described as a tool of lesson organization: explanation of new material, student assessment and many other aspects. However, there are studies that examine other lessons. Sharpe (2008) examined in detail the excerpts of two History lessons and described the teacher's language strategies that lead to student skill development.

The research demonstrates that the analysis of pedagogical discourse using linguistic tools is fairly promising for describing teachers' speech both in general and from a

¹ <https://hundred.org/en/innovations/qridi-a-digital-platform-for-formative-assessment#location>

² <https://hundred.org/en/innovations/navigated-learning-collaborative-powered-by-gooru-navigator>

methodological perspective. Since there are currently no spoken corpora of Russian-speaking teachers, the aim of the study is to propose a procedure for automatic analysis of teacher speech transcripts and comment on the obtained results. The proposed methodology can be useful for studying pedagogical discourse.

Experimental Setup

Corpus description

For this study, we have collected a corpus of hand-annotated speeches from more than 40 teachers who were originally recorded during their classes. Below we present some linguistic, extralinguistic and statistical parameters of the dataset (cf. Table 1). Basic statistics were obtained using the Profiling-UD tool (Brunato et al., 2020), which allows extracting more than 130 morphological, syntactic and semantic characteristics of the text. Its distinctive feature is that it supports multiple languages (including Russian), as it is developed using the Universal Dependencies framework. We present some features of the corpus.

Table 1. *Linguistic and statistical parameters of the corpus*

<i>Extralinguistic parameters</i>	<i>Values</i>
Number of regions	3
Type of school	secondary
Number of subjects	9
Period of recording	spring 2022
<i>Linguistic and statistical parameters</i>	<i>Values</i>
Corpus size in tokens	88786
Number of teachers' sentences	11977
Tokens per sentence	7.4
Average noun distribution	15.04
Average verb distribution	12.89
Average adjective distribution	6.04
Average adverb distribution	8.26

The texts of the corpus were recorded in three different cities of the Russian Federation: Nizhny Novgorod, Lipetsk and Kostroma. The subjects were as follows: the Russian Language, Literature, the Foundations of the Spiritual and Moral Culture of the Peoples in Russia, Handicraft, Geography, History, Mathematics, Physics, and IT. Within the framework of the study, lessons of English as a foreign language were also recorded. It is worth mentioning that we will not take them into consideration, since, unlike monolingual lessons, code switching is typical for foreign language lessons. It is necessary to develop a different methodology for such texts.

Another pivotal parameter is lexical density. According to Johansson (2008), it is defined as a fractional index of meaningful lexical units in a text (nouns, verbs, adjectives, adverbs), ranging from 0 to 1. We consider this parameter as a criterion for determining the formality of pedagogical discourse. The final index is close to 0.6, which indicates the lexical richness of teaching remarks. Thus, we can assume that certain discourse markers and topic related lexical collocations will be detected when applying statistical metrics.

Processing Tools

The procedure for automatic extraction of discourse markers consists of several stages. First, the texts of the corpus were tokenized and lemmatized using the Stanza library (Qi et al., 2020) for the Python 3.7 programming language³. We chose this library because it showed good results in processing both structured and unstructured text data of various genres in Russian (Lagutina, 2022; Mamaev et al., 2023). Secondly, on the basis of the Russian National Corpus⁴ and a Frequency Dictionary of Russian (Lyashevskaya & Sharov, 2009), a list of stop-words was compiled to exclude lexical units that do not contain an important semantic component: prepositions, conjunctions, auxiliary words. Finally, using the Natural Language Toolkit (NLTK) module⁵, we automatically detected biterm collocations from the texts and ranked them by the index of non-randomness in accordance with the values of the log-likelihood and t-score metrics. The first metric is calculated as the likelihood function ratio corresponding to two hypotheses about the random and non-random nature of a certain biterm collocation (Bogoyavlenskaya & Palytchuk, 2022). Since the calculation procedure does not involve information about absolute frequencies of words, this metric is not sensitive to the size of the corpus. On the contrary, t-score depends on the size of the corpus, since the metric takes into account absolute frequencies of both the main word and its collocate. That is why we need to compare the collocations extracted by two opposite metrics in terms of analyzing teachers' speeches. Figure 1 illustrates a part of the text preprocessing script.

```
[ ] swlist = []
    f = open('swl.txt', 'r', encoding='utf-8')
    for i in f:
        i = i.replace('\n', '')
        swlist.append(i)

    f.close()

[ ] frame = []

    for txt in glob.glob('*.txt'):
        f = open(txt, encoding='utf-8').read()
        doc = nlp(f)

        sentence = []

        for sent in doc.sentences:
            for word in sent.words:
                if word.lemma.lower() not in swlist:
                    sentence.append(word.lemma.lower())

        frame.append(sentence)
```

Figure 1. Example of the text preprocessing script

The whole collection of textual data was divided into separate groups depending on a city and a teacher. The first 100 bigrams from the final lists were subject to primary meaningful analysis, classification, as well as interpretation.

³ <https://www.python.org/downloads/release/python-370/>

⁴ <https://ruscorpora.ru/>

⁵ <https://www.nltk.org/>

Results

The results of the study allow us to identify a number of pattern markers that teachers use during their classes. Below we present some results of the log-likelihood value. One of the most common markers among all the teachers, regardless of the city and subject, is the *обратить_внимание* (*pay_attention*) collocation. The main function of this marker is to take notice of a certain speech segment. This collocation turned out to be used in both perfect (*обратить_внимание*) and imperfect (*обращать_внимание*) types. Table 2 presents the main usage examples with the final log-likelihood value of the metric for this group. This marker is used not only in the imperative form, but also as a part of a modal construction *должен_обратить_внимание* (*should_pay_attention*).

Table 2. Examples of the *обратить_внимание* (*pay_attention*) collocation

Group	Example	Value
Nizhny Novgorod_History	Это Куликовская битва. Вы наверняка о ней слышали, а сегодня мы попытаемся что-то новое для себя узнать. Если мы изучаем битву, на что должны обратить внимание ? This is the Battle of Kulikovo. You must have heard about it, and today we are going to learn new facts. If we study the battle, what should we pay attention to?	76.15
Kostroma_Russian_Literature	Итак, часть речи, обратите на это, пожалуйста, внимание ... So, part of speech, pay attention to this, please...	19.72
Lipetsk_Geography	Обратите внимание на устье реки. Она впадает куда? Pay attention to the mouth of the river. Where does it flow into?	433.24

Another large type of discourse collocations contains an addressee function. We were able to describe several subtypes. One of the subtypes can be described by the *PROPER NOUN + IMPERATIVE* construction, which is aimed at a particular student. Other constructions denote a collective name of the addressee or contain a metonymic transfer (cf. Table 3). Examples below are anonymized; all the names are replaced with symbols like X or Y. These discourse markers have certain stylistic differences, indicating the degree of distancing of the teacher from the students. For instance, using generalized expressions such as *молодые люди* (*young people*) indicates the presence of a well-defined hierarchical structure, while the usage of students' personal names and their reduced forms indicates blurring of the boundaries among teachers and students. As a result, it denotes close interaction within classes.

Table 3. Examples of collocations with an addressee function

Group	Example	Value
Nizhny Novgorod_Mathematics	X, пожалуйста, сядь как следует, вполоборота весь урок сидишь! X, please sit still, you sit half-turned the whole lesson!	17.30
Kostroma_The Foundations of the Spiritual and Moral Culture of the Peoples in Russia	Молодые люди , как приветствуют друг друга мужчины? Young people , how do men greet each other?	29.79
Lipetsk_Geography	X! Ты сегодня молодец! Третий ряд ! Y? X! You are doing well today! Third row ! Y?	42.04

Besides the common markers that characterize teachers as a social group, there are also idiosyncrastic markers that characterize the speech of an individual teacher. One of the teachers referred to himself using the third person instead of the first person (so-called *illeism*). The log-likelihood value is 170.39, it takes the second position for the analyzed group, second only to the collocation of *человек_искусство* (*man_art*), its log-likelihood value is 178.44. This discourse collocation may denote self-irony, as well as quite a superficial attitude towards oneself (Fisher, 2015). The name and the patronymic are anonymized with X and Y respectively (cf. Table 4).

Table 4. Examples of collocations with *illeism*

Original examples	English translations
Приготовились. X Y включает... Тишина!	Get ready. X Y turns on ... Silence!
Рассмотрите, пожалуйста, сколько уровней, сколько целей предлагает вам X Y в этом модуле. О чем эти цели?	Consider, please, how many levels, how many goals X Y offers you in this module. What are these goals about?
Смотрим, ребят, на презентацию. Сейчас у нас работа будет такая: внимательно слушаем X Y.	Guys, look at the presentation. Now our task will be as follows: we are listening carefully to X Y.

Finally, a separate type in the list of discourse collocations is represented by terminological units (cf. Table 5 and Table 6). In pedagogy, a terminological unit refers to a specific term or concept used within the field of education and teaching. Terminological units in pedagogy might include terms related to teaching methods, educational philosophies, assessment techniques, and various aspects of the teaching and learning process. The formation of terminological apparatus among students is a central component of the teachers' work. It helps standardize the language used in a particular subject. The list of specific contexts of their usage is quite large, however, when analyzing the selected terms and their contexts, we identified the following situational dominants:

- the selection of synonyms, related terms for analysis;
- the usage of specific examples that reveal the term;
- the usage of references to the term at the beginning and the end of a lesson.

Table 5. Examples of discourse terminological collocations that are used in Mathematics

Group	Example	Value
Lipetsk_Mathematics	Давайте вспомним общий принцип решения линейных уравнений ! Let's recall the general principle of solving linear equations !	15.99
Nizhny Novgorod_Mathematics	Ну нельзя одну двадцать четвертую в десятичную дробь перевести, ну никак нельзя. You can't convert one twenty-fourth into a decimal fraction , you can't.	93.80
Nizhny Novgorod_Mathematics	Мы сейчас вспомним... а на экзамене у вас есть справочные материалы, где вы можете взять формулу для суммы арифметической прогрессии ... We will recall now ... and at the exam you have reference materials where you can take the formula for the sum of an arithmetic progression ...	28.00

Table 6. Examples of discourse terminological collocations that are used in History

Group	Example	Value
Lipetsk_History	Итак, ребят, Рим поэтапно завоевывает всю Италию, и вот римская армия подходит к греческим колониям . Да, греческие колонии . So, guys, Rome is gradually conquering all of Italy, and now the Roman army is approaching the Greek colonies . Yes, Greek colonies .	42.41
Lipetsk_History	Ребята, но после третьего этапа у нас появилось такое выражение — это пиррова победа . Пиррова победа! То есть победа слишком дорогой ценой... Guys, but after the third stage, we had such an expression — this is a Pyrrhic victory . Pyrrhic victory! That is, this victory comes at a great cost.	27.81
Nizhny Novgorod_History	Вот, крестный ход — это... как бы вам это объяснить... Это когда какое-нибудь действие идет по поводу святых людей, либо праздник какой-то священный... Here, the procession is ... how can I explain it to you ... This is when some kind of action takes place on the occasion of holy people, or some kind of sacred holiday...	32.78

For the t-score metric, any collocation with a t-score of 2.00 or higher can be significant; i.e., the combination of the main word and its collocate is not just the result of chance (Nekrasova, 2009). Such collocations may tend to be frequency reproducible in both oral and written texts. However, it should be noted that this condition is sensitive to the language and corpus type. Therefore, it is necessary to make comparisons with the results for other metrics. The results of the second experiment show that there might be some common collocations if we use both metrics, and these collocations can even have the same rank in the frequency lists. For example, Table 7 shows that the first eight collocations occur in two lists.

Table 7. Examples of common collocations used in the Kostroma_Russian_Literature group

Rank	Collocation	Log-likelihood value	T-score value
1	<i>точка_зрение (point_view)</i>	86.90	2.64
2	<i>часть_речь (part_speech)</i>	70.61	2.63
3	<i>римский_цифра (roman_numeral)</i>	66.32	2.44
4	<i>цифра_три (number_three)</i>	58.61	2.44
5	<i>безличный_глагол (impersonal_verb)</i>	54.74	2.18
6	<i>борис_васильев (boris_vasilyev)</i>	54.74	2.00
7	<i>двадцать_восьмой (twenty_eighth)</i>	52.81	2.00

The common collocations in one way or another correlate with the discourse groups we described previously, since both common discourse markers (*точка_зрение (point_view)*) and terms (*часть_речь (part_speech)*, *безличный_глагол (impersonal_verb)*) were detected in the t-score experiment.

Discussion

The obtained results are in line with the previous studies aimed at finding out some typical features of pedagogical discourse (Lee, 2020; Husna et al., 2022; Sharpe, 2008). At the same time, although we focused on the markers that are common for all teachers, we assume that any discourse should also include some individual features of each speaker. Thus, our approach to pedagogical discourse analysis can be described as both socio- and psycholinguistic. How to move from the linguistic analysis to a broader pedagogical analysis of teaching practices is a separate debatable issue.

While we provide an automatic approach of detecting teachers' markers that undeniably improve standard procedures of discourse analysis, it comes with certain limitations that cannot be ignored. Firstly, statistical methods of natural language processing leave aside possible polysemy of pedagogical markers, since their automatic frequency analysis is often carried out without taking contexts into account. For instance, the markers with the addressee function we have considered in Table 3 can be used by teachers as an attribute group. To improve the search for markers, one can add some filters (punctuation marks framing a possible marker, its position in a sentence, etc.). The usage of such a system of formal marker filters would make it possible to present a more complete classification of pedagogical markers, as, for example, it was done by Popescu-Belis & Zufferey (2011).

Secondly, the corpus texts are currently provided only with an orthographic representation of teachers' speech, and this representation raises at least two problems. The first one is that an orthographic representation without phonetic and intonation text annotations does not allow separating markers from each other, which will change their target functions depending on intonation structures. The second problem is related to the fact that the orthographic annotation does not provide information on how the discourse markers were pronounced by a certain speaker. Thus, voice assistants designed for teachers, if provided with the lists of discourse markers only in orthography, will probably lack some useful data for the assessment how close the speech of a certain teacher is to the speech patterns of effective teachers.

Conclusion

The article describes how computational linguistics can be used to identify the markers that characterize pedagogical discourse. We have presented a method that can be used to find out typical word collocations that teachers use in their lessons. These are not just words that are often found nearby, but rather linguistically connected word combinations that serve to solve certain communicative aims. Linguistic analysis of speech using automatic text processing methods makes it possible to analyze not only certain words, but also their grammatical parameters, which in turn can provide information about what grammatical means teachers use to solve the pedagogical issues.

The most interesting results we obtained from the considered material include the clear dominance of the collocation *обратить_внимание* (*pay_attention*) used to attract the attention of students during the lesson, as well as data on those terms that are most often found in the speech of teachers teaching different subjects.

We believe that the method for analyzing teacher practices can be used for a comprehensive interdisciplinary study of pedagogical discourse. For instance, we can conduct an experimental study on the perception of teachers' speech by students to test the assumptions made during the interpretation of the results.

Below we provide specific examples of using the described algorithm and its practical application.

1) Educational studies: identifying characteristics of effective teachers' speech that distinguish it from the speech of their less effective colleagues:

- testing the hypothesis that more effective teachers use the collocation *обратимь_внимание* (*pay_attention*) more frequently to capture students' attention compared to less effective ones;

- investigating the specific usage of imperative forms in teachers' speech. As noted in (Sergomanov et al. 2023), such forms can be considered markers of control. Therefore, it is crucial to determine how often, in what function, and concerning whom (individual students, several students, the whole class) more and less effective teachers employ imperatives;

- comparing the terminology introduced by different teachers in the same subject, as well as checking whether the sets of typical terms match the content of textbooks and scientific manuals on the subject. Additionally, we can assess whether frequent mention of a term in class enhances student understanding, although this may require the development of new tests or experimental procedures.

2) Automatic teacher assistant: If we have access to quantitative data on the speech of more effective teachers, we can use them as benchmarks for other teachers seeking to align their speech with successful teaching practices. This could involve an automatic teacher assistant that analyzes a specific teacher's speech and compares it to established standards.

3) Based on the analysis described above, it is possible to develop recommendations for enhancing teacher speech, both in a general context (e.g., for teachers in a specific subject) and on a personalized level. This would complement the recommendations currently available in educational literature, which primarily rely on general concepts, authors' introspection, and fragmentary observations (Klimova & Kaurova, 2018; Zuyeva, 2009).

Acknowledgments

The paper is a part of the project "The Role of Linguistic Features of Speech in Effective Teaching Practices: A Corpus-based and Psycholinguistic Study", supported by St. Petersburg State University (project ID 101747352).

References

- Bogoyavlenskaya, Yu. V., & Palytchuk, D. A. (2022). Association measures for establishing the force of semantic-syntagmatic connection of elements of word combination. *Humanitarian studies. History and Philology*, 5, 69–78. <https://doi.org/10.24412/2713-0231-2022-5-69-78>
- Brunato, D., Cimino, A., Dell'Orletta, F., Venturi, G., & Montemagni, S. (2020). Profiling-UD: a tool for linguistic profiling of texts. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 7145–7151). European Language Resources Association.
- Evert, S. (2004). *The Statistics of Word Cooccurrences: Word Pairs and Collocations* [Doctoral dissertation, University of Stuttgart]. Computational Approaches to Collocations <https://www.stephanie-evert.de/PUB/Evert2004phd.pdf>
- Fisher, T. (2015). *The Psychological Case for Talking in the Third Person*. <https://www.mic.com/articles/111300/the-psychological-case-for-talking-in-the-third-person>
- Husna, A., Rahman, A., & Abduh A. (2022). Analysis of Speech Acts Used by EFL Teachers in Class Interaction: A Case Study at High School Bosowa Makassar. *Pinisi Journal of Art, Humanities, and Social Studies*, 2(5), 177–184.
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: A developmental perspective. *Working papers/Lund University, Department of Linguistics and Phonetics*, 53, 61–79.
- Klimova, A. V., & Kaurova, A. N. (2018). Teacher's Speech and Requirements to Speech. *Economic and Humanitarian Research of Regions*, 2, 20–25.

- Lagutina, K. (2022). Topical text classification of Russian news: a comparison of BERT and standard models. In *31st Conference of Open Innovations Association (FRUCT)* (pp. 160–166). <https://doi.org/10.23919/FRUCT54823.2022.9770920>
- Lee, J. J. (2020). Spoken Classroom Discourse. In E. Friginal & J. A. Hardy (Eds.), *The Routledge Handbook of Corpus Approaches to Discourse Analysis* (pp. 82–97). Routledge. <https://doi.org/10.4324/9780429259982-6>
- Lyashevskaya, O., & Sharov, S. (2009). *A Frequency Dictionary of Russian*. <http://dict.ruslang.ru/freq.php>
- Mamaev, I. D., Mamaeva, A. A., & Axenova, D. A. (2023). The Semantic Shifts of the Topical Structure in the Corpus of Lentach News Posts. In V. Malykh & A. Filchenkov (Eds.), *Artificial Intelligence and Natural Language. AINL 2022* (pp. 27–39). Communications in Computer and Information Science, 1731. Springer. https://doi.org/10.1007/978-3-031-23372-2_3
- Nekrasova, O. (2017). Factors influencing the process of English linguistic collocations. *Bulletin of Moscow Region State University. Series: Linguistics*, 5, 120–132 <https://doi.org/10.18384/2310-712X-2017-5-120-132>
- Popescu-Belis, A., & Zufferey, S. (2011). Automatic identification of discourse markers in dialogues: An in-depth study of Like and Well. *Computer Speech & Language*, 25(3), 499–518. <https://doi.org/10.1016/j.csl.2010.12.001>
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 101–108. <https://doi.org/10.18653/v1/2020.acl-demos.14>
- Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic Keyword Extraction from Individual Documents. In M. W. Berry & J. Kogan (Eds.), *Text Mining: Theory and Applications* (pp. 1–20). John Wiley & Sons.
- Semenov, A. (2020). Resultative education of an extended personality in a transparent world on a digital platform. *Herzen Readings: Psychological Research in Education*, 3, 590–596.
- Sergomanov, P., & Bysik N. (2022). Teaching practices: research and their platformization in the digital era. *Educational politics*, 1(89), 54–65.
- Sergomanov, P., Maltsev, M., Bysik, N., Beketov, V., & Baiburin, R. (2023). Sociology of the Lesson: Discourse Organization of Successful Teaching Practices. *Educational Studies Moscow*, 1, 191–218. <https://doi.org/10.17323/1814-9545-2023-1-191-218>
- Sharpe, T. (2008). How can teacher talk support learning? *Linguistics and Education*, 19(2), 132–148.
- Wang, W., & Han, M. (2015). Quantitative analysis of the speech of the teachers and students in high school English classroom-based on information technology-based interaction analysis system. *Theory and Practice in Language Studies*, 5(10), 2107–2111. <http://dx.doi.org/10.17507/tpls.0510.18>
- Zuyeva, D. A. (2009). Mathematical Speech Standards of a Teacher: Basic Qualities and Conditions for Their Development. *Izvestia: Herzen University Journal of Humanities & Sciences*, 112, 134–139.