

# Towards the Issue of Creating Linguistic Pedagogical Profiles

Ivan D. Mamaev\*

The Department of Mathematical Linguistics  
Saint Petersburg State University  
Saint Petersburg, Russian Federation  
The Department of Theoretical and Applied Linguistics  
Baltic State Technical University "VOENMEH"  
Saint Petersburg, Russian Federation  
i.mamaev@spbu.ru

## Abstract

The study explores the development of linguistic profiles for evaluating teacher effectiveness in pedagogical discourse. It emphasizes the role of language in educational practices and how linguistic features can distinguish more effective teachers from less effective ones. The research utilizes a corpus of transcripts of Russian secondary school lessons that are analyzed using statistical and machine learning techniques. Key linguistic metrics, such as noun-verb correlation, sentence complexity, and lexical density, are used to create profiles of teaching performance. The study finds that more effective teachers demonstrate a balanced linguistic way of delivering information, employing varied syntactic structures and diverse vocabulary. In contrast, less effective teachers tend to exhibit repetitive and rigid language patterns, which may impede student comprehension. While the study findings provide insights into the relationship between language use and teaching success, the authors acknowledge limitations in the dataset size and suggest future research to expand on the current findings. The preliminary profiles can serve as diagnostic tools to help in developing better pedagogical strategies for trainee teachers in adopting language use that enhances learning outcomes.

## CCS Concepts

• Machine Learning; • Document Management and Text Processing;

## Keywords

Computational Linguistics, Pedagogical Corpus, Profiling, Correlation

## ACM Reference Format:

Ivan D. Mamaev. 2024. Towards the Issue of Creating Linguistic Pedagogical Profiles. In *2024 8th International Conference on Natural Language Processing and Information Retrieval (NLPPIR 2024)*, December 13–15, 2024, Okayama,

\*Corresponding author. mamaev\_id@voenmeh.ru

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

NLPPIR 2024, Okayama, Japan

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 979-8-4007-1738-3/2024/12  
<https://doi.org/10.1145/3711542.3711584>

Japan. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3711542.3711584>

## 1 Introduction

Social communities have long been central to sociological and anthropological research, offering insights into how individuals interact, share values, and form collective identities. These communities, whether defined by geography, shared interests, or cultural practices, shape human behavior, social norms, and provide a sense of belonging. Learning communities, in particular, are crucial due to their impact on personal and social development. These communities, both formal (e.g., schools, universities) and informal (e.g., student networks), are defined by shared goals and practices that support knowledge and skill development [1].

Pedagogical practices within learning communities are key to shaping educational experiences, with communication between teachers and students being essential for knowledge transfer. Language plays a pivotal role in this process, facilitating understanding, critical thinking, and relationship-building [2]. Features like morphology, syntax, and discourse patterns influence how effective the teaching process is. The use of inclusive language fosters belonging, while thoughtful instructional language enhances engagement. Statistical methods, as shown in recent papers, are useful in analyzing linguistic patterns in teaching, helping identify what makes certain teachers more effective. This study aims to develop linguistic profiles of effective and less effective teachers in Russian classrooms, using statistical and machine learning techniques to potentially identify markers of lesson success.

## 2 Related Work

Statistics is significant in linguistic studies as it provides rigorous methods for analyzing linguistic data, enabling researchers to identify patterns, and draw empirically grounded conclusions [3]. It facilitates the quantification of different language phenomena, such as parts of speech, syntactic constructions and stylistic devices. For instance, Baburchenkova [4] delves into the individual style of William Blake through a quantitative analysis of his early lyrics. The main idea of the paper is to uncover the hidden patterns of concept compatibility within Blake's metaphors by applying a correlation analysis, focusing specifically on the concepts of "Creature" and "Psychological Sphere." The research employs stylometric methods to statistically analyze the frequency and co-occurrence of these concepts within metaphorical structures. The results show that these central concepts are both highly active and productive

within Blake’s metaphorical framework, with specific tendencies highlighted in how these concepts are combined with others in his poetry.

Silnitsky [5] investigates the English verb system through a differential correlation analysis. The method involves clustering verb features based on their differential correlations, which effectively distinguishes between motivating and limiting functions of verb features. The results demonstrate the complex interrelations within the verb system, where certain features act as dominant factors influencing the behavior of others, thus contributing to the dynamic structure of the English verb system.

Coloma [6] also involves a quantitative linguistic analysis, though its specific details are not fully clear from the provided excerpts. However, it involves the use of correlation analysis to examine the relationships between different linguistic measures. The paper, while less detailed in the provided description, likely follows a similar quantitative approach as in [4] but in a broader linguistic context.

Wahyuni et al. [7] investigate the correlation between students’ mastery of parts of speech and their writing achievement among 11th-grade students at MAN 4 AGAM. Using a correlational research design, the authors employed tests to assess parts of speech mastery and documentation to evaluate writing performance. The data analysis, conducted via Pearson Product Moment Correlation using SPSS, revealed a positive low correlation value of 0.376, indicating a statistically significant relationship between the two variables. This suggests that while parts of speech mastery does influence writing achievement, the effect is not substantial.

In [8], the link between mastery of sentence patterns and translation ability among third-semester students at Ahmad Dahlan University is studied. The study employs a quantitative correlational approach with a sample size of 122 students, using sentence pattern and translation tests to measure performance. Pearson’s Product Moment correlation was used for statistical analysis. The results show a moderate positive correlation ( $r = 0.464$ ) between sentence pattern mastery and translation ability. Students’ sentence pattern mastery had an average score of 29.83 (out of 45), with translation abilities averaging 64.50 (out of 100), both classified as moderate.

Köhn et al. [9] explore the relationship between syntax and prosody in spoken German texts. This paper leverages a corpus of 31,803 sentences from the Spoken Wikipedia project, focusing on how syntactic functions, such as subjects and objects, affect prosodic features like pitch, duration, and loudness. The research uses linear mixed effects models to analyze these relationships. Notably, the study found that subject words are spoken with a higher pitch (about 0.2 semitones higher) and longer duration (6% longer), while object words are associated with lower loudness and shorter duration.

Kettunen [10] investigates whether the Type-Token Ratio (TTR) can indicate the morphological complexity of languages. The study analyzes 21 languages using TTR and a Morphological Complexity Index (MCI). Results show that TTR does not consistently correlate with MCI, as languages with high morphological complexity, such as Finnish and Turkish, do not always have high TTR values. Instead, TTR is influenced more by text length and lexical diversity

rather than morphological complexity. The study concludes that TTR alone is insufficient for measuring linguistic complexity.

Litvinova [11] discusses the use of computational linguistics for diagnosing personality traits from text through linguistic analysis. The paper highlights the development of a specialized text corpus called “Personality,” which includes metatags about authors (gender, psychological traits). The study identifies correlations between grammatical features of texts and authors’ characteristics, such as gender and psychological profiles, using natural language processing and statistical modeling. This research has significant applications in forensic linguistics and psychological profiling.

This study builds on foundational research such as [12], which emphasizes that teacher effectiveness is not solely dependent on linguistic features but also involves cognitive and motivational factors. While linguistic features are important, they must be contextualized within broader pedagogical practices. For example, Hattie’s meta-analysis identified several non-linguistic traits, such as clarity of instruction and the teacher’s ability to manage classroom dynamics, that contribute to effective teaching.

Furthermore, Hathairat and Maleerat [13] demonstrate the importance of corpus-based vocabulary analysis in the educational context, showing how lexical variety can enhance students’ learning. Similarly, Abdul et al. [14] emphasize the need for advanced corpus-based tools for text processing, particularly in multilingual settings. These tools could be adapted to enhance future analyses in this study by improving the identification of key linguistic markers across languages and contexts.

### 3 Pedagogical Corpus and Experimental Statistical Methodology

The study uses automatic transcripts of secondary school lessons in the Russian Federation, generated with the Whisper program [15]. All the recordings were subsequently subject to corrections. Recorded in spring 2023, the corpus includes 33 texts from more effective teachers and 22 from less effective ones. Effective teachers work in diverse, non-selective schools and achieve above-average State Final Certification results (the 9<sup>th</sup> Grade). Foreign language lessons were excluded. The statistics in Table 1.

One potential limitation noted in the comparison between more effective and less effective teachers involves the representation of different grades and subjects. While the study attempts to maintain diversity across grades 5–9, future work should consider controlling for subject-specific language patterns. The research [16] suggests that the linguistic demands of different subjects can vary significantly. For instance, history lessons may naturally include more abstract language compared to biology, which often involves concrete terminologies and procedural language. Expanding the corpus to include more lessons from varied subjects will help address this issue. The current study serves as an exploratory investigation rather than a conclusive one.

Linguistic features of the lessons were extracted and quantified using the online Linguistic Profiling Tool [17]. It provides a range of linguistic metrics, such as lexical density, syntactic complexity, and sentence structure, which might be key for identifying distinguishing features between the lessons of more effective and less effective teachers. For each text (lesson), the following linguistic measures

**Table 1: General Corpus Statistics**

Feature	Value for the More/Less Effective Teachers' Subcorpus
Subcorpus size in tokens	93,841/50,721
Grade 5 texts	13/8
Grade 6 texts	7/5
Grade 7 texts	6/4
Grade 8 texts	7/4
Grade 9 texts	0/1
Subjects (for all subcorpora)	Russian, Literature, Biology, History, Geography, Social Studies

were calculated: number of nouns, number of verbs, average length of dependent links, pre- and post-subordinate clauses, sentence length, arity (argument structure complexity), average length of prepositional chains, lexical density, type-token ratio, number of determiners. Once the linguistic values were collected, the distribution of each value samples were analyzed to determine whether they followed a normal or abnormal distribution (Smirnov-Kolmogorov test of normality). Next, intertextual correlations between different pairs of linguistic features were calculated. The aim was to explore how syntactic, lexical, and structural elements relate to one another across the different lessons. The combination of such correlations are *linguistic pedagogical profiles*, i.e. a set of linguistic correlates that characterize lesson features. This is a modified definition of *linguistic profiles* described in [18, 19]. While factors such as social status and student-teacher relationships undoubtedly influence educational outcomes, they fall outside the scope of this particular research, which is designed to quantify and analyze linguistic patterns alone. Additionally, the absence of data on social dynamics makes it impossible to factor them into this analysis. By isolating linguistic features, this study provides an essential first step toward understanding how language shapes pedagogical effectiveness, which can later be supplemented with broader contextual factors in future research. The following correlations were computed:

- (1) Nouns-Verbs.
- (2) Dependent Links-Average Sentence Length.
- (3) Pre-Subordinate Clause-Average Sentence Length.
- (4) Post-Subordinate Clause-Average Sentence Length.
- (5) Average Arity-Average Sentence Length.
- (6) Average Prepositional Chain Length-Average Sentence Length.
- (7) Lexical Density-Type Token Ratio (TTR).
- (8) Number of Determiners-Average Sentence Length.

The choice of these correlations is based on theoretical background provided in [20–24]. For instance, if there are few nouns used in the *nouns-verbs* correlation, it indicates a tendency towards using a nominative structure in the construction of pedagogical discourse or the predominance of predicate structures. The value of the *lexical density-type token ratio* correlation investigates the relationship between lexical density (vocabulary richness) and the type-token ratio (variation in vocabulary usage). Finally, as for the *average arity-average sentence length* correlation, verbs with higher arity typically require more explicit arguments or complements. A transitive verb (arity of 2) may lead to longer sentences because it

necessitates a subject and a direct object. In contrast, intransitive verbs (arity of 1) usually require fewer components.

Correlations were calculated using Spearman's rank-order method, as the datasets deviate from normal distribution. The analysis was performed with the Correlations Widget in Orange Data Mining [25], and p-values were computed to assess statistical significance ( $\alpha = 0.05$ ). Scatter plots, generated using the Scatter Plot Widget in Orange, visualized the strength and direction of the correlations.

## 4 Results and Discussion

Table 2 shows the resultant linguistic profiles for both subcorpora presented in the form of correlation values.

The linguistic correlations in Table 2 show key metrics for comparing more and less effective teachers. These values highlight general language trends in teaching but are not formal markers of effectiveness across all subjects or grades. Notably, correlations like *dependent links-average sentence length*, *average arity-average sentence length*, *number of determiners-average sentence length* are nearly identical and cannot be used to determine teacher effectiveness.

### 4.1 Nouns-Verbs

Figure 1 shows correlations between noun and verb usage for two teacher groups: more effective (BP) and less effective (HP). Both groups exhibit negative correlations, with the BP group showing a weaker negative correlation ( $r = -0.202$ ) compared to HP ( $r = -0.255$ ). The BP group's data points are more spread out, suggesting a flexible teaching style, while the HP group shows a clearer negative pattern, indicating an imbalance in noun/verb usage that may affect instructional clarity. Teaching practices and linguistic interpretations are detailed in Table 3.

Table 3 provides concrete examples of noun-verb usage from both more and less effective teachers across different subjects and grades. These examples illustrate typical linguistic patterns but are not meant to imply a strict correspondence between specific subjects or grades and teacher effectiveness. Rather, the table demonstrates how less effective teachers tend to show rigid or unbalanced language patterns. This distinction holds across subjects, focusing on language use rather than the content of specific disciplines.

### 4.2 Lexical Density-Type Token Ratio

The correlation between lexical density and TTR is weaker for more effective teachers (Figure 2), suggesting they use a broader range

**Table 2: Pedagogical Linguistic Profiles**

Linguistic Correlations	More Effective Teachers	Less Effective Teachers
Nouns-Verbs	-0.202	-0.255
Dependent Links-Average Sentence Length	0.908	0.88
Pre-Subordinate Clause-Average Sentence Length	0.076	-0.011
Post-Subordinate Clause-Average Sentence Length	-0.076	0.011
Average Arity-Average Sentence Length	0.789	0.78
Average Prepositional Chain Length-Average Sentence Length	0.131	-0.401
Lexical Density-Type Token Ratio	0.185	0.267
Number of Determiners-Average Sentence Length	0.341	0.3

**Table 3: Examples for the Nouns-Verbs correlation**

Russian Example	English Translation	Linguistic Interpretation
<b>More Effective Teachers</b>		
Pravil’no, <i>zapisyvaj</i> , uzhe <i>zapisyvav</i> ’ nado. Uzhe vse <i>nachertili</i> , vrode kak. (History, the 5 <sup>th</sup> Grade)	That’s right, <i>write</i> it <i>down</i> , it’s time to <i>write</i> it <i>down</i> . We <i>have</i> already <i>drawn</i> everything, it seems (History, the 5 <sup>th</sup> Grade).	No nouns are used. Teaching recommendation
Immunitet. <i>Davajte, davajte, davajte</i> , shire <i>dumajte, dumajte, dumajte, davajte, ne stesnyajtes</i> ’, vot kakie u vas associacii prihodiat – <i>govorite</i> . (Biology, the 8 <sup>th</sup> Grade)	Immunity. <i>Come on, come on, come on, think</i> outside the box, <i>think, think, come on, don’t be shy</i> , what associations <i>come</i> to you – <i>tell</i> me (Biology, the 8 <sup>th</sup> Grade).	Very few nouns are used. Lexical repetition of verbs denotes a very active call for action
<b>Less Effective Teachers</b>		
<i>Kupechestvo</i> postepenno prevrashchaetsya v <i>burzhuaziyu</i> . Zapisyvaem. <i>Kupechestvo</i> postepenno prevrashchaetsya v <i>burzhuaziyu</i> . Vsyo, to est’ vot u nas, pozhalujsta, s vami rabochij <i>klass, burzhuaziya, predpriyatiya</i> . (Social Studies, the 8 <sup>th</sup> Grade)	The <i>merchant class</i> is gradually turning into <i>the bourgeoisie</i> . Let’s write it down. The <i>merchant class</i> is gradually turning into <i>the bourgeoisie</i> . That’s it, that is, here we have, please, <i>the working class, the bourgeoisie, the enterprises</i> (Social Studies, the 8 <sup>th</sup> Grade).	Parallel syntactic constructions. Repetition of noun terms without a description of what contribution they make to a given fragment of the lesson indicates uncertainty in the pedagogical speech.
Vpolne veroyatno. To est’, smotrite. Ya vam chasto govoryu, chto, devochki, chto ya vam govoryu? Kogda vy <i>vra</i> . . . <i>vyrastite</i> , vy <i>b</i> . . . <i>obrashchajte</i> vnimanie na. . . na chto? (Literature, the 6 <sup>th</sup> Grade)	Probably. I mean, look. I often tell you that, girls, what do I tell you? When you <i>gr</i> . . . <i>grow</i> up, you <i>p</i> . . . <i>pay</i> attention to. . . to what? (Literature, the 6 <sup>th</sup> Grade)	Hesitation on verbs indicate the vague formulation of pedagogical thoughts.

of unique words without significantly increasing lexical density. Subjects like literature and biology show varied lexical strategies, indicating adaptability. In contrast, the correlation is stronger for less effective teachers, with higher lexical density tied to greater use of unique lemmata. The distribution is more compact, with points clustering at moderate-to-low lexical density, suggesting that less effective teachers may prioritize vocabulary diversity at the expense of clarity. Examples are shown in Table 4.

### 4.3 Average Prepositional Chain Length-Average Sentence Length

Figure 3 presents two scatter plots for correlation. More effective teachers show a slight positive correlation ( $r = 0.131$ ). In the lower left corner of the upper plot, we observe a concentration of points

with both short prepositional chains and shorter sentences. However, moving toward the upper right, we see points with longer chains and longer sentences, which reinforces the trend. The overall situation reflects better command of complex sentence construction and suggests more fluid, sophisticated speech patterns in the classroom. On the other side, there is a stronger negative correlation ( $r = -0.401$ ) for less effective teachers. A more notable cluster of points in the upper left part of the bottom plot suggests a frequent usage of longer sentences with shorter prepositional chains. As prepositional complexity increases (rightward movement), we see fewer points with long sentences, confirming the negative correlation. It is suggested that teachers’ speech becomes more fragmented or less cohesive when using complex syntactic structures, which

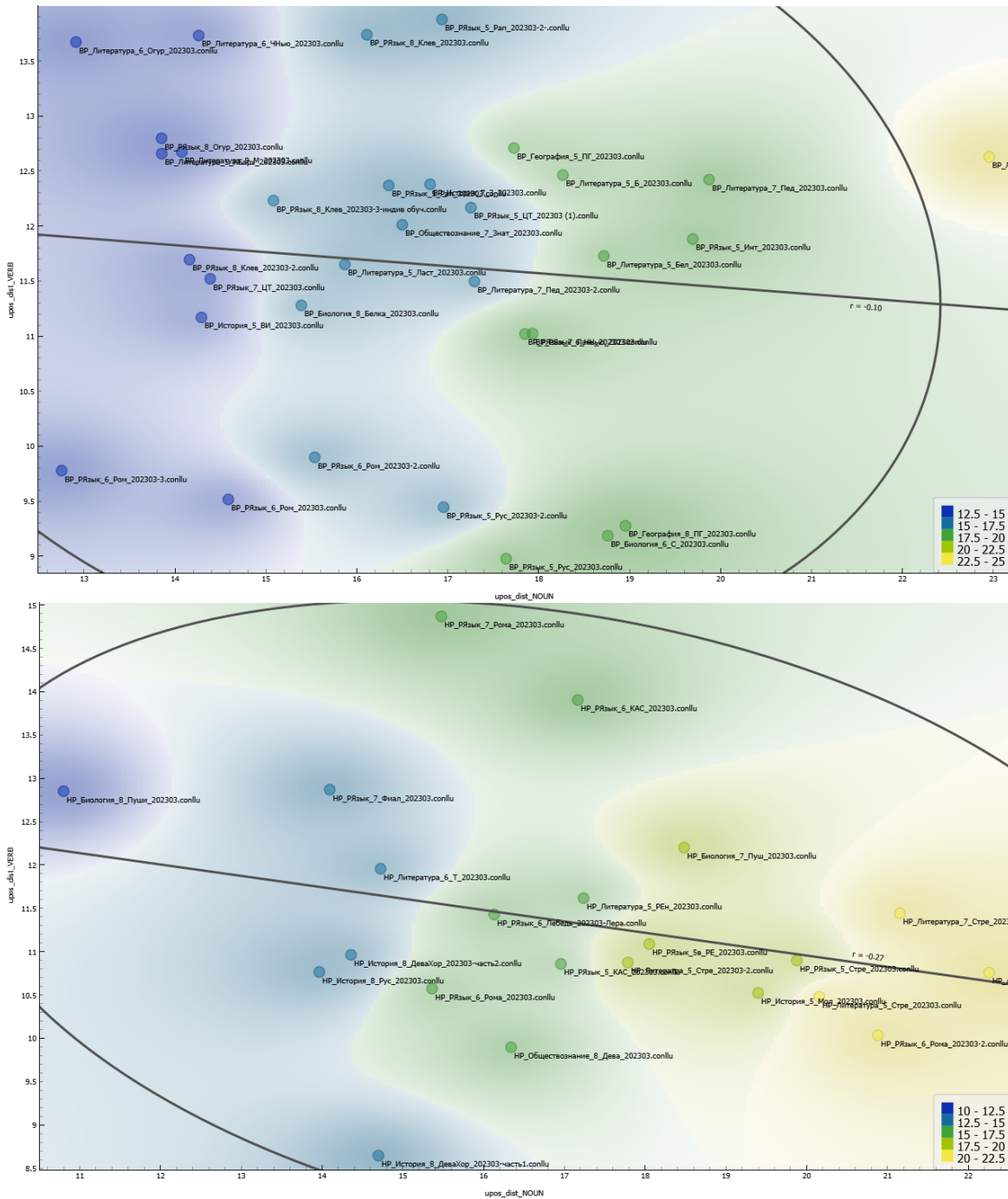


Figure 1: The visualization of the *nouns-verbs* correlation for more effective (above) and less effective (below) teachers

might impact the clarity of their communication. Some examples are presented in Table 5.

#### 4.4 Pre-Subordinate Clause-Average Sentence Length

As for less effective teachers, there is a more concentrated clustering of data points around shorter sentence lengths and lower pre-subordinate clause usage. Despite all the values being almost

Table 4: Examples for the Lexical Density-Type Token Ratio correlation

Russian Example	English Translation	Linguistic Interpretation
More Effective Teachers		
Potomu chto korni, rastushchie ot <i>stblej</i> ili ot <i>list'ev</i> , nazyvayutsya <i>pridatochnymi</i> . Esli ya ego posazhu v otidel'nyj gorshochek, u menya vyrastet <i>rasteniya</i> ? (Biology, the 6 <sup>th</sup> Grade)	Because <i>roots</i> that grow from <i>stems</i> or <i>leaves</i> are called <i>adventitious</i> . If I plant it in a separate pot, will I grow a <i>plant</i> ? (Biology, the 6 <sup>th</sup> Grade)	The usage of five terms per 24 words. There are no repetitions of terms.
Less Effective Teachers		
<i>Kryl'ya</i> – eto <i>perednie konechnosti</i> , kotorye prevratilis' v <i>kryl'ya</i> , da, to est' esli my <i>zemnovodnyh</i> budem sravnivat' s <i>pticami</i> , tam dve <i>lapki</i> dve pary <i>lapok</i> vernee. (Biology, the 7 <sup>th</sup> Grade)	<i>Wings</i> are the <i>forelimbs</i> that have turned into <i>wings</i> , yes, that is, if we compare <i>amphibians</i> with <i>birds</i> , there are two <i>feet</i> , it is better to say that they have two pairs of <i>feet</i> . (Biology, the 7 <sup>th</sup> Grade)	The usage of seven terms per 25 words. Repeating terms inside definitions might confuse students.

Table 5: Examples for the Average Prepositional Chain Length-Average Sentence Length correlation

Russian Example	English Translation	Linguistic Interpretation
More Effective Teachers		
Eto proizvedenie bylo napechatano v <i>samom nachale</i> tvorcheskoy deyatel'nosti Maksima Gor'kogo, kogda on eshcho ne byl znamenitym pisatelem. No poyavilos' ono v <i>pechati</i> , vyshlo v <i>svet v «Samskoy gazete» v tsysyacha vosem'sot devyanosto pyatom godu</i> , kogda on byl uzhe yarkim i izvestnym pisatelem. (Literature, the 7 <sup>th</sup> Grade)	This work was published at the very beginning of Maxim Gorky's creative activity, when he was not yet a famous writer. But it appeared in print, came out (a prepositional construction in the Russian language) in "Samara newspaper" in one thousand eight hundred and ninety-five, when he was already a bright and famous writer. (Literature, the 7 <sup>th</sup> Grade)	The usage of extended prepositional constructions in compound and complex sentences.
Less Effective Teachers		
Sosloviya, molodcy, pravil'no, sosloviya, da? To est', eto vseгда <i>u nas</i> bylo. Takzhe <i>u nas</i> , poluchaetsya, chto byli? Potomstvennye i lichnye dvoryane, da? Pochyotnye grazhdane, kupcy. U kupcov po... vspominaem, vspominaem, ch... kupcy kak delyatsya? <i>Na kakie kategorii? U kupcov</i> kakie kategorii? Shest'desyat vos'maya stranichka, podsmatrivaem, podskazyvaem sebe. (Social Studies, the 8 <sup>th</sup> Grade)	Estates, well done, right, estates, right? That is, <i>we (a prepositional construction in the Russian language)</i> have always had this. What did <i>we (a prepositional construction in the Russian language)</i> turn out to have? Hereditary and personal nobles, right? Honorary citizens, merchants. Merchants have... we remember, we remember, how are merchants divided? <i>Into what categories?</i> What categories do merchants have? Page sixty-eight, look, prompt yourselves. (Social Studies, the 8 <sup>th</sup> Grade)	The usage of prepositional constructions in short simple sentences, most of the constructions have the head (a preposition) and a dependent (a noun), they are not extended.

equal to zero, less effective teachers tend to use simple and straightforward sentences. Turning to the *post-subordinate clause-average sentence length* correlation, it is necessary to mention that the results are the same, but with the opposite sign. Table 6 presents some examples (proper nouns of students are anonymized with X).

More effective teachers use complex sentence structures, fostering deeper engagement, while less effective teachers simplify language, often due to inexperience or lack of confidence [26]. The analysis shows that effective teachers employ a wider variety of syntactic structures and maintain a balanced noun-verb ratio,

enhancing lesson dynamism and accessibility. In contrast, less effective teachers exhibit rigid patterns, overusing nouns or verbs, leading to monotonous discourse. Effective teachers also use more sophisticated sentence structures and prepositional chains, adjusting language complexity to students' needs, which aligns with research on pedagogical success and effective communication [12].

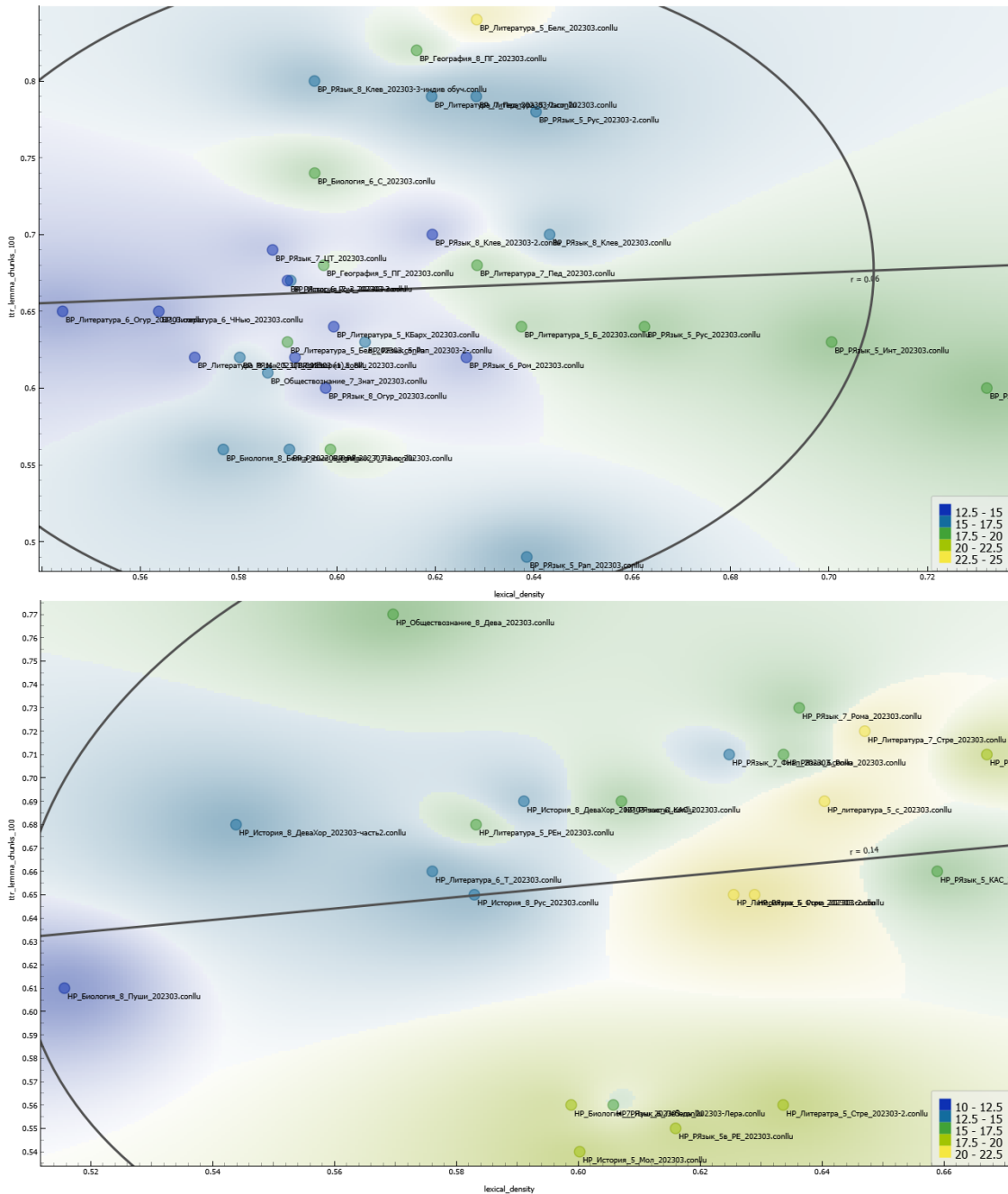
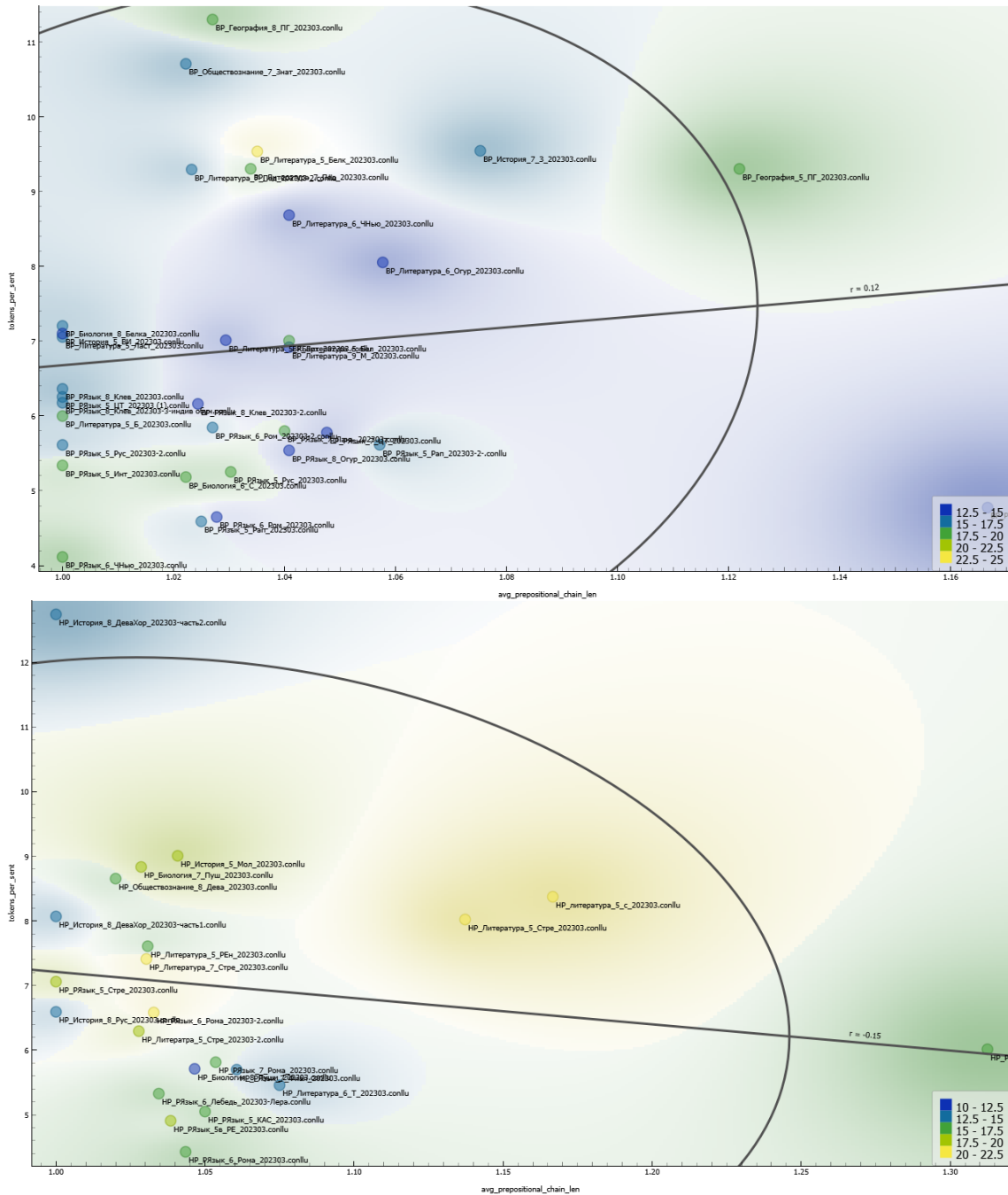


Figure 2: The visualization of the *lexical density-type token ratio* correlation for more effective (above) and less effective (below) teachers

### 5 Conclusion and future work

This research continues a series of studies on effective and ineffective pedagogical practices [26–28]. In this study, we have successfully developed linguistic pedagogical profiles for more effective and less effective teachers using statistical and machine learning methods. The analysis demonstrated that key linguistic features – such as noun-verb usage, sentence complexity, and lexical density

– correlate with teaching effectiveness. More effective teachers were found to employ a balanced and flexible linguistic approach, characterized by varied syntactic structures and a broader range of lexical items. In contrast, less effective teachers exhibited more rigid and repetitive patterns, which may hinder student engagement and comprehension. These findings provide an important



**Figure 3: The visualization of the average prepositional chain length-average sentence length correlation for more effective (above) and less effective (below) teachers**

step towards understanding how linguistic factors contribute to successful pedagogy.

The current study has limitations. While the corpus size is sufficient for initial analysis, it could be expanded to include a broader range of subjects and contexts. The identified correlations highlight

trends but don't fully capture the complex interaction between linguistic features and teaching outcomes. Future research could use larger datasets and qualitative methods, such as student feedback, to explore these dynamics more deeply.

Looking ahead, there are several promising research directions. These include applying linguistic profiles in teacher training as



**Table 6: Examples for the Pre-Subordinate Clause-Average Sentence Length correlation**

Russian Example	English Translation	Linguistic Interpretation
<b>More Effective Teachers</b>		
<i>Esli by vas poprosili v kraskah izobrazit' etu kartinu</i> , kakie by cveta vy ispol'zovali, X? (Literature, the 5 <sup>th</sup> Grade)	<i>If you were asked to paint this picture</i> , what colors would you use, X? (Literature, the 5 <sup>th</sup> Grade)	Using a conditional sentence with a pre-subordinate clause makes students pay attention to the topic of the lesson.
Da, <i>kto vse sdelal, poka zhdyom ot drugih</i> , mozhetе i chislo napisat'. (Russian, the 5 <sup>th</sup> Grade)	Yes, those, who did everything, can write the date while we are waiting for the others (both clauses are pre-subordinate in the Russian language). (Russian, the 5 <sup>th</sup> Grade)	The use of subordinate clauses allows the teacher to parallelize the actions of students in the classroom.
<b>Less Effective Teachers</b>		
Znachit, i ona v tsysyacha sem'sot sorok chetvyortom godu privozit emu iz Germanii, mozno skazat', vypisyvaet, vot... zvuchit eto dostatochno grubo, da? (History, the 8 <sup>th</sup> Grade)	So, in 1744, she brings it to him from Germany, you could say, orders it, well... it sounds pretty crude, right? (History, the 8 <sup>th</sup> Grade)	No pre-subordinate clauses are used. Paraphrasing the sentence makes it difficult to follow the main idea of the lesson.
Vot sejchas my pokazываем. Tak. Horosho. My s vami eshchyo posmotrim migraciyu ptic segodnya, nebol'shoj interaktiv. Na primere polyarnoj pticy, kotorye u nas obitayut v Arktike. (Biology, the 7 <sup>th</sup> Grade)	We are showing it now. So. Great. We will also look at bird migration today, it will be a small interactive part of the lesson. Using the example of polar birds that live in the Arctic. (Biology, the 7 <sup>th</sup> Grade)	No pre-subordinate clauses are used. Simplifying compound sentences through the use of parcellation.

diagnostic tools, expanding the study to include cross-linguistic and cross-cultural comparisons, and incorporating AI-powered tools for real-time linguistic analysis to help teachers adjust their language and improve lesson effectiveness.

## Acknowledgments

The author acknowledges Saint Petersburg State University for a research project 103923108 and SberObrazovanie LLC for a research No. 230712-107-YUL.

## References

- [1] Martin-Kniep, G. O. 2007. *Communities That Learn, Lead, and Last: Building and Sustaining Educational Expertise*. John Wiley & Sons.
- [2] Walton, J., Priest, N., and Paradies, Y. 2013. Identifying and Developing Effective Approaches to Foster Intercultural Understanding in Schools. *Intercultural Education*, Vol. 24, 181–194. <http://dx.doi.org/10.1080/14675986.2013.793036>.
- [3] Rasinger, S. M. 2013. *Quantitative Research in Linguistics: An Introduction*. A&C Black.
- [4] Baburchenkova, I. O. 2016. Compatibility of Units for the Conceptual Sphere in the Early Works of W. Blake: Correlation Analysis [Sochetaemost' edinic konceptosfery v rannem tvorchestve U. Blejka: korrelyatsionnyj analiz]. *Quantitative Linguistics*, Vol. 4, 31–43.
- [5] Silnitsky, G. G. 2017. Differential Correlation Analysis of Multilevel Verb Features in English. *Proceedings of Smolensk State University*, Vol. 4, 197–210.
- [6] Coloma, G. 2022. Correlation between Linguistic Measures: An Extended Analysis. *Studies in Linguistics and Literature*, Vol. 6, 109–132. <http://dx.doi.org/10.22158/sll.v6n4p109>.
- [7] Wahyuni, C. and Safitri, L. 2022. The Correlation Between Parts of Speech Mastery and Writing Achievement. *International Journal of Research on English Teaching and Applied Linguistics*, Vol. 2, 37–44. <http://dx.doi.org/10.30863/ijretal.v2i2.2453>.
- [8] Sari, D. M. 2022. The Correlation Between Sentence Patterns Mastery and Translating Ability. *Zenodo*. Vol. 8, №16, 163–170.
- [9] Köhn, A., Baumann, T., and Dörfler, O. 2018. An Empirical Analysis of the Correlation of Syntax and Prosody. <https://arxiv.org/abs/1806.05900>.
- [10] Kettunen, K. 2014. Can Type-Token Ratio be Used to Show Morphological Complexity of Languages? *Journal of Quantitative Linguistics* 21, 223–245. <http://dx.doi.org/10.1080/09296174.2014.911506>.
- [11] Litvinova, T. A. 2015. Resources of Computational Linguistics for the Task of Author Profiling: a Corpus-Based Study. *Proceedings of Voronezh State University. Series: Linguistics and Intercultural Communication*, Vol. 3, 37–41.
- [12] Hattie, J. 2003. Teachers make a difference: What is the research evidence? *Paper presented at the Australian Council for Educational Research Annual Conference on Building Teacher Quality*. Melbourne, Australia. Retrieved from [http://research.acer.edu.au/research\\_conference\\_2003/4/](http://research.acer.edu.au/research_conference_2003/4/).
- [13] Ketmaneechairat, H., and Maliyaem, M. 2023. Corpus-Based Vocabulary List for Thai Language. *Journal of Advances in Information Technology*, Vol. 14, №2, 319–327, 2023.
- [14] Jabbar, A., Illahi, M., Iqbal, S., Khan, A. R., ElHakim, N., and Saba, T. 2023. PWMStem: A Corpus-Based Suffix Identification and Stripping Algorithm for Multi-lingual Stemming. *Journal of Advances in Information Technology*, Vol. 14, №4, 863–875. <http://dx.doi.org/10.12720/jait.14.4.863-875>.
- [15] Radford, A., Kim J. W., Xu T., Brockman G., McLeavey C., and Sutskever, I. 2023. Robust Speech Recognition via Large-Scale Weak Supervision. *Proceedings of the 40th International Conference on Machine Learning (ICML'23)*. Vol. 202, Article 1182, 28492–28518.
- [16] Redish, E. F., and Kuo, E. 2015. Language of physics, language of math: Disciplinary culture and dynamic epistemology. *Science & Education*, Vol. 24, 561–590.
- [17] Brunato, D., Cimino, A., Dell'Orletta F., Venturi, G., and Montemagni, S. 2020. Profiling-UD: a Tool for Linguistic Profiling of Texts. *Proceedings of the Twelfth Language Resources and Evaluation Conference*, Marseille, France. European Language Resources Association, 7145–7151.
- [18] Mamaev, I. D. 2024. Cluster Analysis of Linguistic Profiles of Hidden Communities. *Philology. Theory and Practice*, Vol. 17, №5, 1739–1747. <http://dx.doi.org/10.30853/phil20240250>.
- [19] Mamaev, I. D. 2024. Linguistic Profiles of Hidden Communities: A Morphosyntactic Aspect. *Philology. Theory and Practice*, Vol. 17, №4, 1155–1162. <http://dx.doi.org/10.30853/phil20240168>.
- [20] Akimova, G. N., Vyatkina, S. V., Kazakov, V. P., Rudnev, D. V. 2013. *Syntax of the modern Russian language*. Faculty of Philology, St. Petersburg State University.
- [21] Konyushkevich, M. I. 2013. Transformation of Prepositional-Case Syntaxeme into a Predicative Unit: Correlation of Preposition and Connection Indicator of a Complex Sentence. *Linguistic Studies*, Vol. 26, 93–99.
- [22] Orekhov, B.V. 2022. Metric and lexical Diversity in the Poems of A.A. Voznesensky. *Proceedings of the V.V. Vinogradov Russian Language Institute*, 50–58. <http://dx.doi.org/10.31912/pvrl-2022.3.4>.
- [23] Potebnyia, A. A. 1958. *Notes on Russian Grammar [Iz zapisok po russkoj grammatike]*. Uchpedgiz Publisher, Vol. 1–2.
- [24] Reinhart, T. and Siloni, T. 2005. The Lexicon-Syntax Parameter: Reflexivization and Other Arity Operations. *Linguistic Inquiry* 36, 389–436. <http://dx.doi.org/10.1215/00141801-2005-003>.

- 1162/0024389054396881.
- [25] Demšar, J., Zupan, B., Leban, G., and Curk, T. 2004. Orange: From Experimental Machine Learning to Interactive Data Mining. *Lecture Notes in Computer Science*, 537–539. [http://dx.doi.org/10.1007/978-3-540-30116-5\\_58](http://dx.doi.org/10.1007/978-3-540-30116-5_58).
- [26] Owen, D. 1996. Do Teachers Modify Their Speech According to the Proficiency of their Students. *English Language Teacher Education and Development*, Vol. 2, №. 1, 31-51.
- [27] Alibali, M. V., Nathan, M. J., Wolfgram, M. S., Breckinridge, R. C., Jacobs, S. A., Martinez, C. J. and Knuth, E. J. 2014. How Teachers Link Ideas in Mathematics Instruction Using Speech and Gesture: A Corpus Analysis. *Cognition and Instruction*. Vol. 32, №1, 65-100.
- [28] Ford, C. E. 1984. The Influence of Speech Variety on Teachers' Evaluation of Students with Comparable Academic Ability. *TESOL Quarterly*. Vol. 18, №1, 25-40.