

**Всероссийский конкурс исследовательских и проектных работ школьников
«Высший пилотаж»**

**Сравнение лингвистических особенностей в онлайн-общении учеников и
учителей**

Исследовательская работа

Направление «*Лингвистика*»

Автор: Золотаревский Павел Витальевич

Учащийся 11 класса, ОАНО «Новая Школа», Москва

2024 г.

Содержание

1 Введение	3
1.1 Цель исследования	3
1.2 Гипотеза	3
2 Метод анализа	3
2.1 Источники данных	3
2.2 Критерии анализа	3
2.2.1 Средняя длина сообщений	3
2.2.2 Эмоциональный окрас	3
2.2.3 Популярность слов	3
2.2.4 Изменение популярности слов со временем	4
2.2.5 Пунктуационные и орфографические ошибки	4
2.2.6 Использование эмодзи и стикеров	4
2.3 Инструменты анализа	4
3 Анализ	4
3.1 Длина сообщений	4
3.2 Эмоциональный окрас	5
3.3 Популярность слов	5
3.4 Временной анализ	6
3.5 Пунктуация и орфографические ошибки	7
3.6 Эмодзи и стикеры	8
4 Выводы	8
Библиография	8

1 Введение

В коммуникации между людьми разных возрастов часто возникает барьер, связанный с особенностями общения разных поколений. Из-за особенностей интернет-общения молодого поколения, взрослые, не привыкшие к этой культуре, могут воспринимать их как необразованных или неграмотных [1]. Это проявляется и во взаимодействиях учеников с их учителями. Данное исследование сравнивает лингвистические особенности в общении учеников и общении учителей. Понимание этих различий может стать шагом к преодолению непонимания и снятию коммуникационного барьера.

1.1 Цель исследования

Выявление различий в лингвистических особенностях в общении учеников и учителей в рамках одной тематической области.

1.2 Гипотеза

Общение учеников будет придерживаться менее формального стиля, чем общение учителей [2], даже в рамках чатов одинаковой направленности.

2 Метод анализа

2.1 Источники данных

В качестве корпуса для анализа были взяты групповые чаты для людей из соответствующей категории. Поскольку в тематических групповых чатах практически гарантирована однородность выборки (большинство участников относится к анализируемой группе), они являются самым точным из пригодных для анализа источников.

Для учителей были взяты специализированные чаты для педагогов [3, 4], а для учеников были выбраны чаты про олимпиады и поступление в университеты [5, 6, 7].

2.2 Критерии анализа

Для подробной оценки лингвистических особенностей, выполнен анализ сообщений по шести разным критериям:

2.2.1 Средняя длина сообщений

При переписках в сети среди молодого поколения, длина сообщений обычно не превышает пары десятков слов [8]. При этом учителя, привыкшие к более формальным методам общения, таким как электронные письма, скорее всего, будут чаще писать длинные сообщения.

2.2.2 Эмоциональный окрас

И ученики, и учителя, могут выражать в переписке недовольство какими-то событиями или образовательными инструментами [9], но из-за более формального контекста чатов учителей, в них ожидается меньшее количество сообщений с открыто выраженным негативным эмоциональным окрасом.

2.2.3 Популярность слов

И ученики, и учителя, обсуждают темы, связанные со школой. Однако они обсуждают школу с разных позиций, и наиболее используемые с

2.2.4 Изменение популярности слов со временем

В онлайн-речи молодого поколения обильно используется интернет-сленг [2], который имеет свойство быстро появляться и устаревать. В общении учеников ожидается большее количество слов, чья популярность зависит от временного промежутка.

Также ожидается, что некоторые темы для обсуждения имеют цикличную популярность – например, ЕГЭ скорее всего обсуждают прямо перед экзаменом гораздо активнее, чем в самом начале учебного года.

2.2.5 Пунктуационные и орфографические ошибки

В онлайн-общении реже используются знаки препинания и чаще допускаются орфографические ошибки [2]. Эти особенности более распространены среди молодого поколения, особенно в сравнении с учителями, которые, скорее всего, уделяют большее внимание правильности речи.

2.2.6 Использование эмодзи и стикеров

Пользователи мессенджеров, вне зависимости от поколения, используют визуальные методы, такие как смайлики, для выражения своих эмоций. Предполагается, что эмодзи чаще используются учителями, тогда как школьники отдают предпочтение стикерам.

2.3 Инструменты анализа

Для анализа текста применён язык Python и следующие библиотеки:

- Polars для анализа данных [10]
- spaCy для обработки естественного языка [11]
- HuggingFace Transformers с моделью rubert-tiny2-russian-sentiment для анализа сентимента [12]
- NumPy для построения линейных регрессий [13]
- seaborn для визуализации [14]
- rypellchecker для проверки орфографии

3 Анализ

3.1 Длина сообщений

Для начала проанализированы общие лингвистические характеристики по каждой группе.

Анализ длин сообщений (Рис. 1) показал, что ученики, как правило, пишут более короткие сообщения ($Q_3 = 6$). Практически все сообщения длиннее 20 слов являются выбросами, тогда как длины сообщений учителей могут достигать и до 100 слов ($Q_3 = 23$). В обеих группах присутствует большое количество выбросов, но почти все из них – текст задач из учебников или экзаменов, которыми участники чата делились друг с другом, поэтому эти сообщения были исключены из анализа.

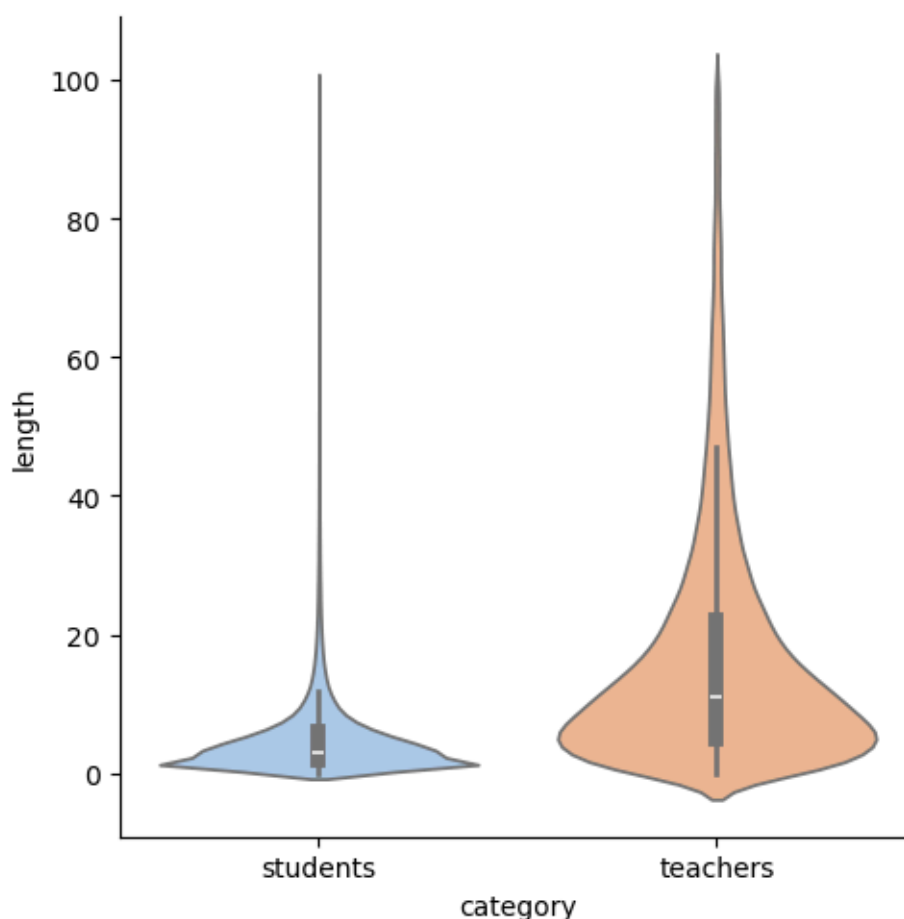


Рис. 1. Распределение длин сообщений (в словах)

3.2 Эмоциональный окрас

Большинство сообщений в обеих категориях имеют нейтральный эмоциональный окрас (83% для учеников и 88% для учителей). При этом у учеников в два раза чаще встречаются негативно настроенные сообщения: в чатах педагогов их доля составляет 4%, а в чатах учеников – почти 8%. Доля позитивно настроенных сообщений практически не различается (9% для учеников, 8% для учителей).

3.3 Популярность слов

Среди наиболее используемых¹ учениками слов² находится много слов, связанных с экзаменами (писать, балл, егэ). Учителя же чаще используют слова, напрямую связанные с их работой – ребёнок, класс, школа, учитель, работать (Таблица 1). Эти результаты являются ожидаемыми, учитывая тематику чата: экзамены и преподавание – главные универсальные темы, которые объединяют людей в чатах, посвященных обучению.

¹Из анализа исключены шумовые слова, такие как частицы и предлоги.

²Под словами в данном случае подразумеваются лексемы, а не конкретные формы слов. Например, за лексему «год» также считаются слова «года» и «лет».

Таблица 1 - Таблица 1. Наиболее часто используемые лексемы (по среднему кол-ву использований в месяц, округлено до целых значений)

Ученики		Учителя	
Слово	Исп./мес	Слово	Исп./мес
год	118	ребёнок	84
знать	117	класс	69
человек	96	школа	66
хороший	93	задача	65
писать	89	год	58
балл	88	учитель	54
задача	77	вопрос	45
егэ	74	знать	42
думать	72	работать	39
написать	71	задание	39

3.4 Временной анализ

Далее проведен анализ зависимости использования определенных слов от временного промежутка.

Поиск сленговых слов выполнялся при помощи линейной регрессии (OLS). Как и ожидалось, в чате учеников используются сленговые слова, которые набирают или теряют популярность в конкретный период (Рис. 2). Были найдены как утратившие популярность слова (например «кек», «тян», «блэт»), так и новый сленг, вступивший в обиход за последние пару лет («альтушка», «скуф», «база»³) В чатах учителей сленговых слов найдено не было.

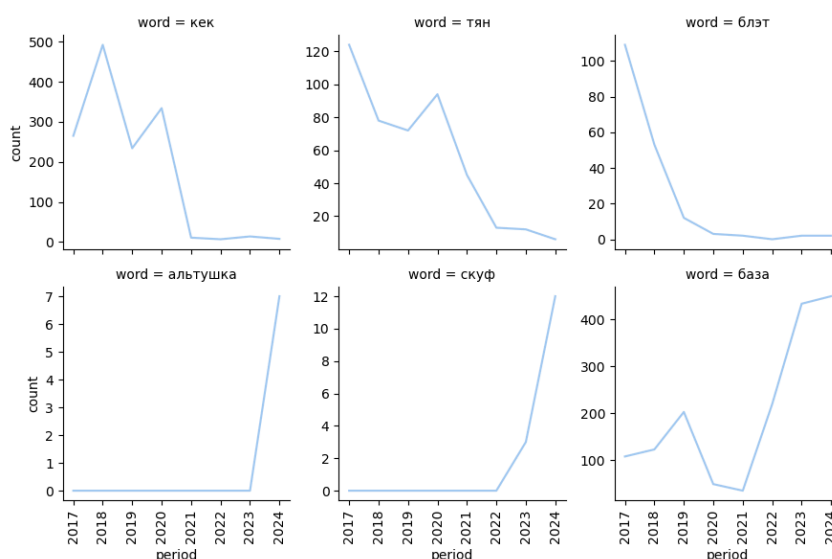


Рис. 2. Количество использований сленговых слов по годам

³В контексте чатов учеников, слово «база» используется также для обозначения базового уровня ЕГЭ и определенных баз данных. С этим связано использование этого слова до 2021 года, а также часть использований после 2021 года. Однако большая часть использований после 2021 года приходится именно на сленговое значение слова – выражение одобрения.

Слова, связанные с ЕГЭ, действительно активнее используются перед экзаменами. На Рис. 3 изображена популярность слова «ЕГЭ» в чатах учеников по месяцам. Можно заметить, что экзамен активнее всего обсуждают прямо перед ним, а к началу учебного года тема вновь теряет свою популярность. Точно такая же закономерность прослеживается и среди чатов учителей, а также со словами «балл», «пробник», «экзамен», «задача».

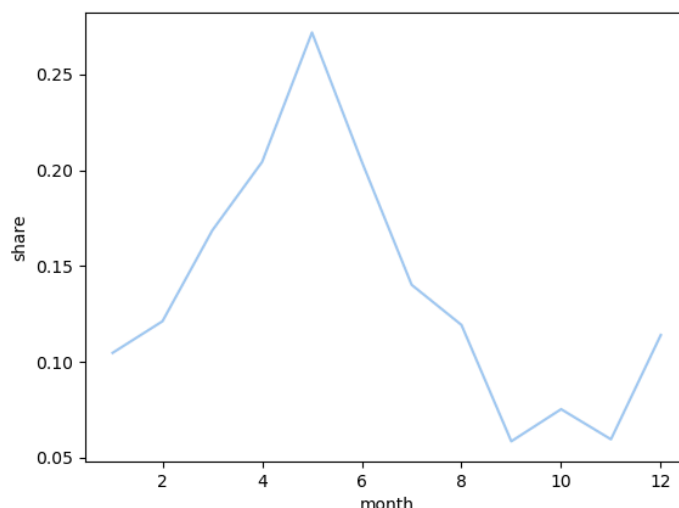


Рис. 3. Процент слова «ЕГЭ» от всех слов, использованных в каждый месяц

3.5 Пунктуация и орфографические ошибки

Учителя используют знаки препинания гораздо чаще, чем ученики. Среди текстов сообщений учителей, 3.6% всех символов являются пунктуацией, а среди сообщений учеников пунктуация составляет всего 2.6% символов. Таблица 2 показывает значения для конкретных пунктуационных знаков. Единственным знаком, который ученики используют чаще, чем учителя, оказался вопросительный знак.

Таблица 2 - Таблица 2. Процент знаков препинания среди всех символов

Знак	% у учителей	% у учеников
,	1.35	0.94
.	0.94	0.32
:	0.05	0.02
!	0.1	0.08
?	0.27	0.51
"	0.14	0.04

В обеих категориях практически все слова, не найденные в словаре, были аббревиатурами. Самым популярным сокращением в обеих группах было ЕГЭ. При этом остальные сокращения различались и соответствовали категории. Так в чатах учителей чаще всего использовались аббревиатуры МЦКО, МЭШ, ППЭ, а ученики чаще писали сокращения ИТМО, Физтех, ВШЭ, СПбГУ и другие названия университетов. Ученики также часто пользовались ненормативной лексикой, а также сокращенными формами

слов, такими как «мб» (2171 использования), «чел» (1820) или «кста» (1092). При этом конкретных опечаток⁴, встречающихся более чем в редких единичных случаях, обнаружено не было.

3.6 Эмодзи и стикеры

Изначальная гипотеза подтвердилась. Сообщения учителей в среднем содержат по 0.08 эмодзи, а сообщения учеников в среднем содержат всего 0.03. С стикерами закономерность противоположная - 3.2% сообщений учеников были стикерами, а среди сообщений учителей стикеры составляли всего 0.1%.

4 Выводы

По сравнению с педагогами, ученики придерживаются менее формального стиля в общении. Они могут использовать мат или сленг, писать короткие сообщения или заменять их стикерами, пренебрегать знаками препинания и более активно выражать свои негативные эмоции.

При этом, несмотря на неофициальность общения, в чатах учеников все еще присутствует большая доля организованности. Обсуждения, так же как и среди учителей, ведутся на значимые темы, соответствующие тематике чата и важные для участников: экзамены, поступление в университет, оценки. Даже несмотря на опускание пунктуации, свойственное для интернет-общения молодого поколения [2], ученики заботятся о понятности речи и не допускают большого количества орфографических ошибок.

Важно учесть ограничения данного исследования. Основные ограничения – маленькое количество чатов учителей для анализа и отсутствие вычислительных мощностей для использования передовых методов анализа текста. Корректировка ограничений может повлиять на конкретные числовые значения, однако общий полученный результат имеет четкий механизм и скорее всего останется неизменным.

Библиография

1. Kemp, N. Texting versus txtng: reading and writing text messages, and links with other linguistic skills / N. Kemp // Writing Systems Research. – 2010. – Т. 2. – Texting versus txtng: reading and writing text messages, and links with other linguistic skills. – № 1. – С. 53-71
2. Исаева, А. Ю. Лингвистические особенности коммуникации поколения Y и поколения Z в социальной сети Instagram (на материале английского и немецкого языков) / А. Ю. Исаева, Я. А. Нестерова // Вестник ЮУрГУ. Серия «Лингвистика». – 2020. – Т. 17. – Лингвистические особенности коммуникации поколения Y и поколения Z в социальной сети Instagram (на материале английского и немецкого языков). – № 2. – С. 5-10
3. Чат учителей математики
4. @infMSK учителя информатики города Москвы
5. DANO Chat
6. Поступашки 2024

⁴Здесь не учитываются преднамеренные опечатки, такие как «чо».

7.Поступашки - Вакансии, Мемы, Задачи

8.Lyddy, F. An Analysis of Language in University Students' Text Messages* / F. Lyddy, F. Farina, J. Hanney [и др.] // Journal of Computer-Mediated Communication. – 2014. – Т. 19. – An Analysis of Language in University Students' Text Messages*. – № 3. – С. 546-561

9.Spatiotis, N. Sentiment Analysis of Teachers Using Social Information in Educational Platform Environments / N. Spatiotis, I. Perikos, I. Mporas, M. Paraskevas // International Journal on Artificial Intelligence Tools. – 2020. – Т. 29. – Sentiment Analysis of Teachers Using Social Information in Educational Platform Environments. – № 2. – С. 2040004

10.Vink, R. pola-rs/polars: Python Polars 1.16.0

11.Honnibal, M. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing / M. Honnibal, I. Montani To appear. – 2017.

12.Wolf, T. Transformers: State-of-the-Art Natural Language Processing / T. Wolf, L. Debut, V. Sanh [и др.] // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations / ред. Q. Liu, D. Schlangen. – Online : Association for Computational Linguistics, 2020.. – Transformers: State-of-the-Art Natural Language Processing. – С. 38-45

13.Harris, C. R. Array programming with NumPy / C. R. Harris, K. J. Millman, S. J. van der Walt [и др.] // Nature. – 2020. – Т. 585. – Array programming with NumPy. – С. 357-362

14.Waskom, M. mwaskom/seaborn: v0.8.1 (September 2017)