

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное
учреждение высшего образования «Национальный исследовательский
университет ИТМО»

Факультет инфокоммуникационных технологий

Математическая лингвистика

Практическая работа №2

Выполнили:

студент группы К34422

Малаев Степан Геннадьевич

Проверил:

доцент практики, КТН

Болгова Екатерина Владимировна

Санкт-Петербург

2025

Цель работы. Ознакомиться с разделом количественной лингвистики и базовыми статистическими понятиями, научиться исследовать авторские инварианты.

Ход работы.

1. Выбор поэтов и сбор стихотворений

Были выбраны следующие авторы:

- Серебряный век
 - Александр Введенский (1904-1941)
 - Николай Заболоцкий (1903-1958)
- Позднесоветский период
 - Николай Рубцов (1936-1971)
 - Сергей Гандлевский (1952-н.в.)

Стихотворения:

- Александр Введенский
 - Мне жалко что я не зверь
 - Элегия
 - Все
 - Умный Петя
 - Лошадка
 - Когда я вырасту большой...
 - Значенье моря
 - Колыбельная
 - Гость на коне
- Николай Заболоцкий
 - Можжевельный куст
 - Не позволяй душе лениться
 - Некрасивая девочка
 - О красоте человеческих лиц
 - Признание
 - Портрет

- В этой роще березовой...
- Я не ищу гармонии в природе
- Я воспитан природой суровой...
- Николай Рубцов
 - Звезда полей
 - Тихая моя родина
 - В горнице моей светло
 - Родная деревня
 - Привет, Россия
 - Березы
 - Первый снег
 - Про зайца
 - По вечерам
- Сергей Гандлевский
 - В чёрном теле...
 - До колючих седин...
 - Есть в растительной...
 - Здесь реки кричат...
 - Как ангел...
 - Когда я жил...
 - Мне нравится смотреть...
 - Мы знаем приближение...
 - Осенний снег...

2. Подготовка и предварительная обработка данных

Ввиду того, что поэмы имеют предсказуемую структуру было принято решение унифицировать лишь форматирование строк и строф.

3. Количественный лингвистический анализ

Соберем следующие метрики:

- Строки
 - Общее количество
 - Среднее количество на строф
- Строфы
 - Общее количество
- Предложения
 - Общее количество
- Части речи
 - Общее количество
 - Среднее количество каждой
 - Среднее количество каждой на строку
 - Среднее количество каждой на предложение
- Слова
 - Среднее количество на строф
 - Среднее количество на строку
 - Среднее количество на предложение

4. Анализ метрик и тропов

Рассмотрим полученные метрики:

- Александр Введенский
 - Количество строф: 2.22
 - Количество строк: 78.44
 - Количество предложений: 17.33
 - Среднее количество строк в строфе: 52.56

- Среднее количество слов в строке: 3.80
- Среднее количество слов в предложении: 68.71
- Количество частей речи
 - Союз: 27.89
 - Наречие: 16.00
 - Причастие: 3.80
 - Инфинитив: 4.11
 - Сравнительная частица: 3.00
 - Существительное: 66.89
 - Предлог: 26.00
 - Местоимение: 28.44
 - Междометие: 2.00
 - Прилагательное: 20.00
 - Краткое прилагательное: 2.00
 - Глагол: 39.33
 - Краткое причастие: 1.00
 - Числительное: 1.75
 - Частица: 12.78
 - Деепричастие: 5.20
 - Предикатив: 1.75
- Николай Заболоцкий
 - Количество строф: 3.89
 - Количество строк: 25.89
 - Количество предложений: 7.89
 - Среднее количество строк в строфе: 10.50
 - Среднее количество слов в строке: 4.72
 - Среднее количество слов в предложении: 15.32
 - Количество частей речи
 - Наречие: 4.11
 - Инфинитив: 2.43

- Союз: 10.67
- Причастие: 2.00
- Сравнительная частица: 2.00
- Существительное: 21.11
- Предлог: 12.44
- Местоимение: 6.67
- Прилагательное: 15.22
- Краткое прилагательное: 1.80
- Глагол: 9.67
- Краткое причастие: 1.50
- Числительное: 2.00
- Частица: 4.44
- Деепричастие: 2.50
- Предикатив: 1.00

- Николай Рубцов

- Количество строк: 3.89
- Количество строк: 18.11
- Количество предложений: 7.33
- Среднее количество строк в строфе: 6.56
- Среднее количество слов в строке: 4.90
- Среднее количество слов в предложении: 13.33
- Количество частей речи
 - Союз: 8.25
 - Наречие: 4.67
 - Причастие: 2.25
 - Инфинитив: 3.50
 - Сравнительная частица: 1.33
 - Существительное: 14.78
 - Предлог: 12.44
 - Местоимение: 3.56

- Прилагательное: 9.56
- Краткое причастие: 1.00
- Глагол: 8.78
- Краткое прилагательное: 1.00
- Числительное: 1.00
- Частица: 2.89
- Деепричастие: 1.00
- Сергей Гандлевский
 - Количество строф: 1.89
 - Количество строк: 19.00
 - Количество предложений: 7.44
 - Среднее количество строк в строфе: 14.11
 - Среднее количество слов в строке: 4.96
 - Среднее количество слов в предложении: 20.62
 - Количество частей речи
 - Союз: 10.00
 - Наречие: 4.33
 - Инфинитив: 2.25
 - Причастие: 1.00
 - Существительное: 18.33
 - Предлог: 10.44
 - Местоимение: 5.88
 - Прилагательное: 8.22
 - Краткое прилагательное: 1.00
 - Глагол: 8.56
 - Числительное: 1.00
 - Частица: 4.50
 - Деепричастие: 1.00
 - Предикатив: 1.00

5. Сравнительный и статистический анализ

Александр Введенский характеризуется высоким абсолютным количеством существительных 66.89, глаголов 39.33 и союзов 27.89. Он демонстрирует протяжённость и сложность формы, что соответствует его авангардным экспериментам в поэтическом языке.

Николай Заболоцкий демонстрирует умеренные показатели, где доминируют существительные 21.11, союзы 10.67 и прилагательные 15.22. Такие показатели соответствуют лирической направленности поэзии, где экономия слов и точность выразительных средств играют важную роль.

Николай Рубцов показывает несколько более низкие абсолютные значения, особенно по показателям существительных 14.78 и глаголов 8.78, также характеризуются лаконичностью и экономией выразительных средств, что отражает лирическую направленность.

Сергей Гандлевский характеризуется средними значениями, где заметно выделяются существительные 18.33, союзы 10.00 и глаголы 8.56. Более высокая средняя длина предложений 20.62 слов и компактная структура стихотворений указывают на современный синтаксический стиль с акцентом на эмоциональную насыщенность, ритмику и гармонию.

6. Визуализация результатов

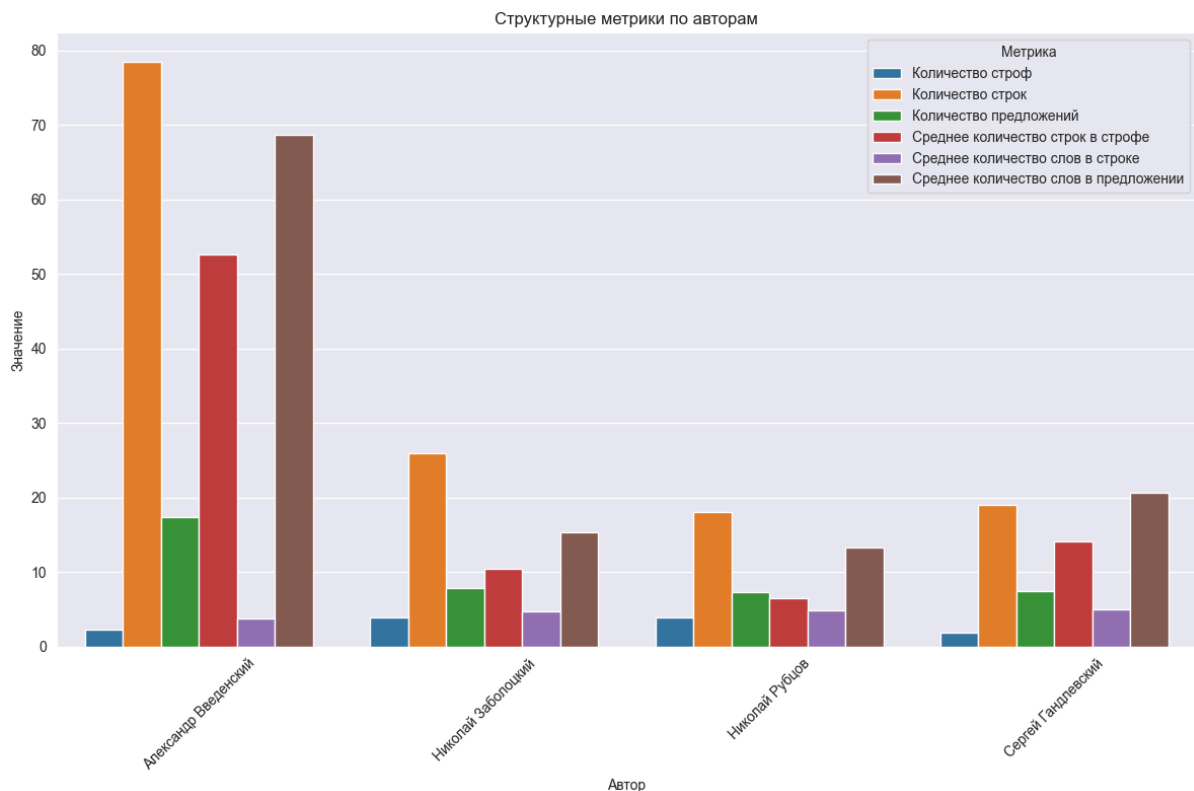


Рисунок 1 – График структурных метрик по авторам

Александр Введенский демонстрирует наибольшее среднее количество строк около 78 при относительно малом числе строф, что приводит к крайне высокой средней длине строф, около 52 строк на строфу.

Одновременно наблюдается заметно увеличенное среднее число слов в предложении, указывающее на усложнённую синтаксическую организацию текстов и склонность к развернутым конструкциям.

Николай Заболоцкий и Николай Рубцов располагаются ближе друг к другу по структурным показателям. У обоих около 4 строф и от 18 до 26 строк в среднем. Различия заключаются в несколько более высокой плотности лексических единиц у Рубцова, в среднем около 4.9 слов на строку, при более сжатой композиции. У Заболоцкого же наблюдается большее среднее число строк в строфе и более длинные предложения, что может показывать склонность к развёрнутой фразе и более детализированному описанию.

Сергей Гандлевский занимает промежуточную позицию по числу строф и строк, в среднем около 2 строф и 19 строк, но отличается достаточно высокой средней длиной предложения, свыше 20 слов. Это может указывать на более современную тенденцию к сложному синтаксису при относительно компактном построении строф.

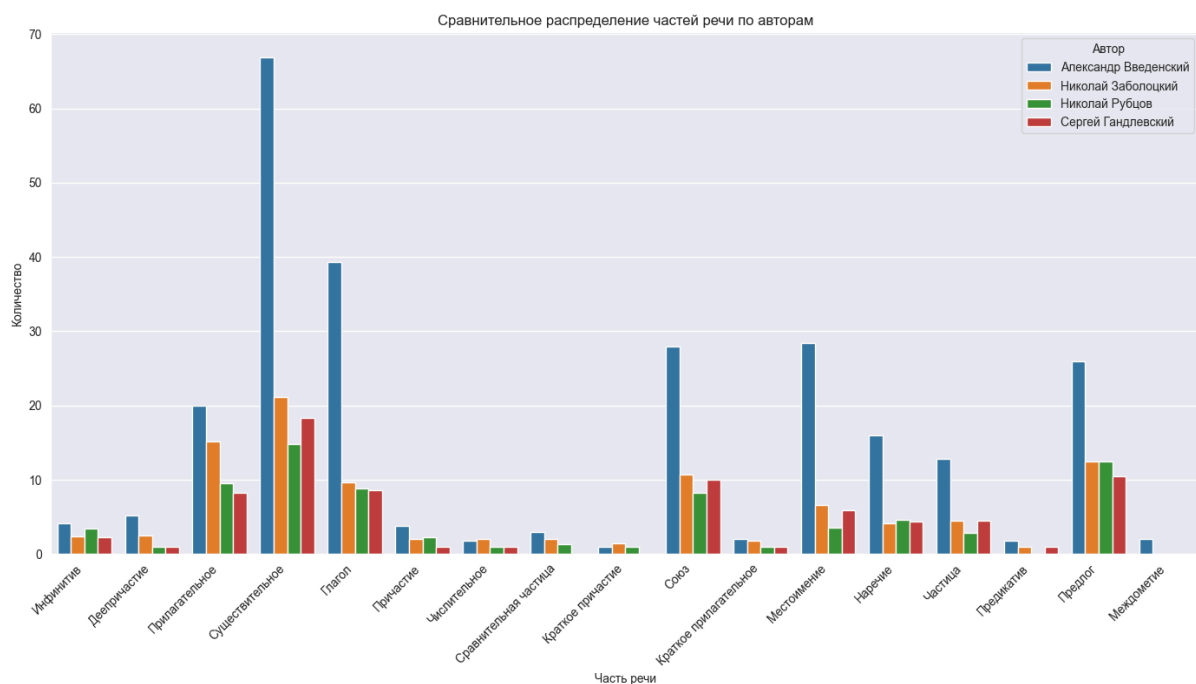


Рисунок 2 – График распределение частей речи по авторам

На диаграмме распределения частей речи наиболее выделяется Александр Введенский, у которого наблюдается значительный перевес существительных и глаголов, а также сравнительно высокие показатели союзов и местоимений. Это близко к экспериментальной, авангардной стилистики его поэзии, где синтаксические элементы и активное использование глаголов создают эффект “нагнетения”. Николай Заболоцкий и Николай Рубцов характеризуются более умеренными показателями, среди которых выделяются существительные,

предлоги и прилагательные. Подобная структура указывает на лирическую ориентацию текстов.

Сергей Гандлевский демонстрирует относительно сбалансированное распределение, при этом средняя длина предложений у него высока, что говорит о комбинировании различных частей речи в более протяжённых синтаксических конструкциях.

Вывод.

В ходе лабораторной работы была выполнена серия шагов по сбору, предварительной очистке и количественному анализу поэтических текстов четырёх авторов. Структурные метрики продемонстрировали различия в композиционном построении стихотворений, а анализ частей речи позволил выявить особенности лексического наполнения.

Таким образом, количественный лингвистический анализ подтвердил, что поэты разных эпох и направлений демонстрируют чётко выраженные стилистические и структурные различия

Ссылки.

Репозиторий с исходным кодом, использованный в данной лабораторной работе: [\[ссылка\]](#)