

## Об исследовании

### Какова область исследования?

Исследование посвящено верификации параметров в русской традиции квантитативного определения авторства текста.

### Что мы исследуем?

Различные параметры текста, которые, предположительно, помогают определить авторство.

### Почему это исследование важно?

Существует необходимость знать настоящего автора определенных текстов.

### Кто занимался изучением этой проблемы?

Г. Хьетсо, М. И. Шапир, В. П. и Т. Г. Фоменко и И. В. Пешков и многие другие филологи и лингвисты.

### Что можно сказать об уже проведенных исследованиях?

Параметры не были проверены на большом количестве авторов и произведений.

Именно поэтому нам необходимо проверить, можем ли мы получить достоверную информацию при применении параметров, использованных исследователями.

## Изученные параметры

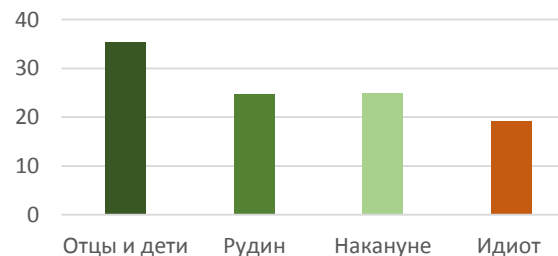
- Длина предложения (Хьетсо 1984, Фоменко 1984);
- Длина слова (Хьетсо 1984, Фоменко 1984);
- Количество служебных слов (Фоменко 1984);
- Распределение частей речи (Хьетсо 1984);
- Позиция частей речи в предложении (Хьетсо 1984);
- Сочетание частей речи (Хьетсо 1984);
- Коэффициент словарного состава (Хьетсо 1984);
- Лексический спектр текста (Хьетсо 1984);
- Словарный профиль текстов (Хьетсо 1984);
- Концентрация словаря (Хьетсо 1984);
- Коэффициент типичности знаков (Хьетсо 1984);
- N-граммы (Пешков 2000, Орехов, Пешков 2013).

## Методика

- Чтение исследований, проведенных в этой области
- Изучение всех параметров и методов их нахождения
- Сбор текстов с известным авторством (15 авторов, 60 текстов)
- Написание компьютерной программы, которая считает показатели по всем параметрам для каждого текста
- Написание компьютерной программы, которая проверяет работоспособность параметров
- Проверка наличия трех типов ошибок:
  - Показатели по разным авторам имеют близкие значения
  - Показатели по текстам одного автора имеют большой разброс
  - Показатели по разным авторам пересекаются
 ИЛИ проверка предусмотренная исследователями, применившими этот параметр
- Выявление вероятности ошибки для каждого параметра
- Определение рабочих и нерабочих параметров

## Пример

Из графика для параметра распределение предложения по числу слов видно, что значения по текстам одного автора имеют большой разброс, что недопустимо при определении авторства текста.



## Результаты

Ни один из параметров, рассмотренных в текущей работе, не может дать абсолютной точной информации об авторстве некоторого текста.

- ✓ Наименьший процент ошибки допускается при определении авторства текста при помощи параметра **коэффициент словарного состава**. Только в 5.85 процентах случаев этот параметр определяет авторство текста неверно.
- ✓ Разница между лексическими спектрами, распределение частей речи и позиция частей речи в предложении могут определять авторство текстов примерно в 50 процентах случаев, потому что иногда эти параметры показывают слишком близкие средние значения по разным авторам.



- ✓ Значения по параметрам, связанные с сочетаниями слов, не могут определить автора текста, потому что, помимо того, что они показывают слишком близкие значения по различным авторам, эти параметры имеют большое различие по одному автору, что непременно ведет к пересечению значений по разным авторам.
- ✓ Трехступенчатая проверка параметров показала, что все остальные параметры для разных авторов имеют пересекающиеся значения (зеленым отмечен диапазон одного автора, оранжевым – диапазон другого автора).



Исходя из этого, все рассмотренные параметры не способны определить автора текста с высокой точностью.

## Библиография

- Орехов, Пешков 2013 – Б.В. Орехов, И.В. Пешков. «На подступах к геному стиля Шекспира.» Новый филологический вестник (2013): 107-131.  
 Пешков 2000 – Пешков, И.В. «Делу – венец, или еще раз об авторстве М.Бахтина в «спорных текстах.» Бахтин, М.М. Фрейдизм. ФИЛ. МФЯ. Статьи. Москва: Лабиринт, 2000. 602-625.  
 Фоменко 1984 – В.П. Фоменко, Т.Г. Фоменко. «Авторский инвариант русских литературных текстов.» Методы количественного анализа текстов нарративных источников. Москва: Академия Наук СССР, 1984. 86-109.  
 Хьетсо и др. 1984 – Г. Хьетсо, С. Густавсон, Б. Бекман, С. Гил. Кто написал "Тихий Дон"? Нью-Джерси: Humanities Press, 1984.  
 Шапир 1997 – Шапир, М.И. «Феномен Батенькова и проблема мистификации.» Philologica 4 (1997): 85-139.