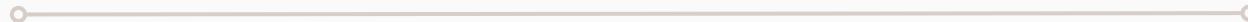




Проблемы атрибуции паломнических текстов: возможности стилометрии



НУГ «Паломничество в России во второй половине XIX –
начале XX вв. в контексте модернизации»

Стилометрия

Метод анализа текстов, которые переводит стилевое своеобразие текста в исчислимые параметры, позволяющие **измерять и количественно сравнивать стиль разных авторов**

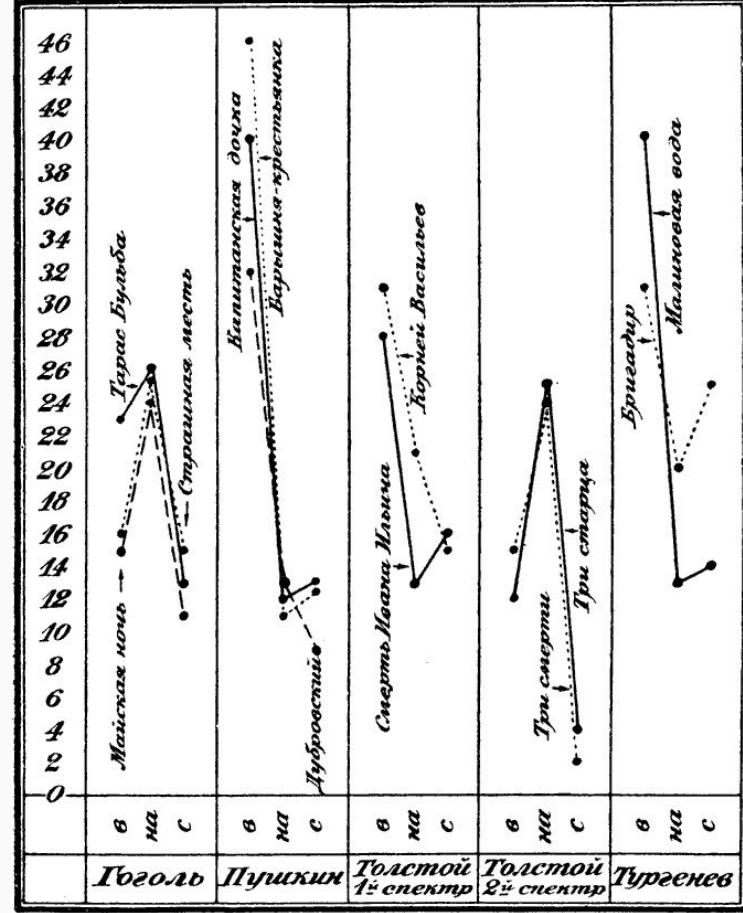
(Орехов Б.В. Компьютерный анализ текста // Цифровые гуманитарные исследования: монография / А. Б. Антопольский, А. А. Бонч-Осмоловская, Л. И. Бородкин [и др.]. Красноярск: Сиб. федер. ун-т, 2023. С. 120—157.)

Из истории стилометрии

Н.А. Морозов,
начало XX века



Джованни Морелли





Джон Берроуз, 1990-е - 2000-е гг.

Метод “Дельта Берроуза” – использование частотностей слов для сопоставления текстов между собой (определение Д. Скоринкина)

Каждый текст представляется как вектор частотностей

Посмотрим классический пример о Роулинг

Каждый текст – вектор частотностей:

	rowling_chamber	rowling_goblet	rowling_hallows	rowling_order	rowling_prince	rowling_prisoner	rowling_stone	tolkien_lord1	tolkien_lord2	tolkien_lord
the	4.414578	4.482777	4.696009	4.229664	4.127059	4.415193	4.338762	5.623838	5.248270	5.935850
and	2.344034	2.425887	2.472936	2.252543	2.171365	2.569458	2.418285	3.782166	3.537469	4.344676
to	2.352251	2.485908	2.243849	2.507990	2.493114	2.389742	2.331105	2.073566	2.027329	2.057737
of	1.876871	2.022443	1.869874	2.030131	1.970931	2.010880	1.923004	2.597432	2.556816	2.942046
a	2.001291	1.791232	1.448879	1.792354	1.849616	2.078881	1.981124	1.916350	1.837024	1.731658
was	1.481308	1.423278	1.125752	1.387096	1.394831	1.384302	1.488370	1.313315	1.084867	1.131274
I	0.881507	0.848643	0.525424	0.858532	1.247729	0.699437	0.798514	1.492428	1.246691	1.140148
in	1.167909	1.116910	0.995079	1.167758	1.101213	1.228871	1.175029	1.419435	1.359320	1.508368
he	1.162040	1.365344	1.374522	1.346763	1.519076	1.296872	1.277370	1.221231	1.262873	1.436641
said	1.414402	1.372129	0.972663	1.515012	1.434097	1.165728	1.001933	0.825384	0.952819	0.855478
you	0.962498	0.894572	0.963368	1.083249	1.176229	0.806295	0.972873	0.871987	0.954113	0.917581
that	0.726568	0.803758	0.904319	0.836637	1.124070	0.733437	0.726496	1.247059	1.101049	1.271011
it	0.991842	0.934238	1.009841	0.933054	0.945320	0.791723	1.004460	1.143746	1.037614	1.038108
his	1.305241	1.294885	1.227993	1.218847	1.269999	1.661162	1.129544	0.801801	0.919159	0.963430
had	0.788779	0.970772	0.950793	0.894641	0.965833	0.918010	0.880640	0.657499	0.559910	0.606303
on	0.679617	0.627871	0.543466	0.734458	0.539764	0.796581	0.780826	0.668168	0.591628	0.579684
at	0.758260	0.854906	0.631493	0.811668	0.739612	0.777152	0.734077	0.539026	0.536608	0.563418
her	0.326310	0.447286	0.514489	0.672229	0.483502	0.349718	0.219844	0.085346	0.081559	0.126436
with	0.550502	0.592902	0.537452	0.600396	0.616539	0.582864	0.506652	0.571592	0.639528	0.644751

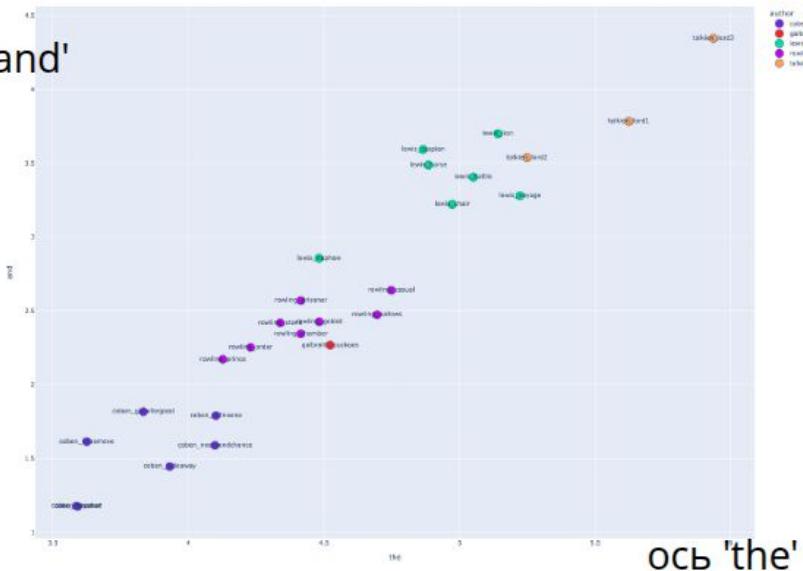
Закон Цифпа,
закон Хипса в
лингвистике

Использованы слайды
Д. Скоринкина

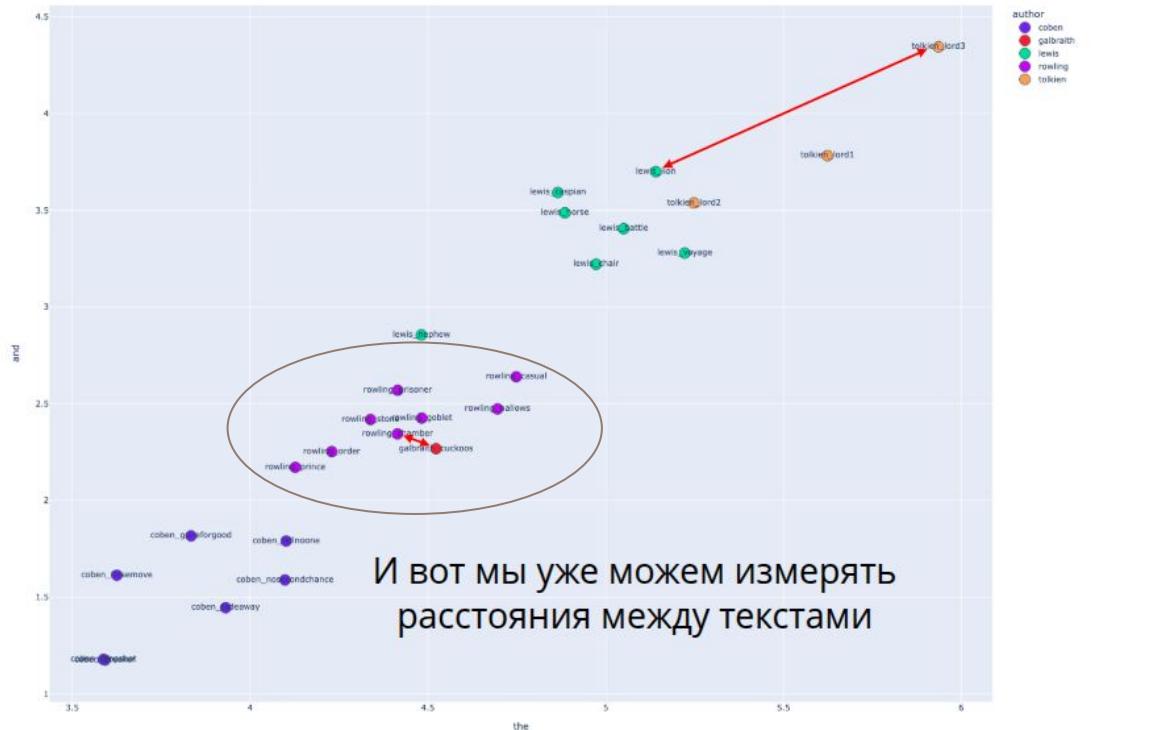
Посмотрим классический пример о

rowling_chamber	rowling_goblet	rowling_hallows	rowling_order	rowling_prince	rowling_prisoner	rowling_stone	tolkien_lord1	tolkien_lord2	tolkien_lord:	
the	4.414578	4.482777	4.696009	4.229664	4.127059	4.415193	4.338762	5.623838	5.248270	5.93585t
and	2.344034	2.425887	2.472936	2.252543	2.171365	2.569458	2.418285	3.782166	3.537469	4.34467t

Two-dimensional word-free space drama for HUJI-PUB school



Посмотрим классический пример о

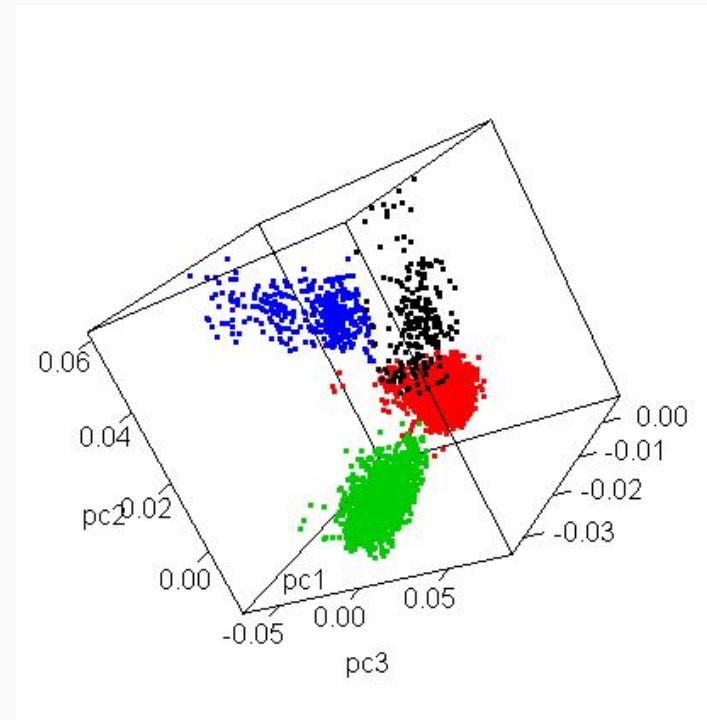


Роберт Гэлбрейт Джоан Роулинг

И вот мы уже можем измерять
расстояния между текстами

Посмотрим классический пример о Роулинг

Стилометрия делает то же самое, но не с 2 словами, а с топом 100-500 частотных слов



Какие задачи решает стилометрия

- Определения авторства
- Изменения стиля внутри творчества одного автора (стилохронология ("ранний/поздний стиль"), жанровые различия)
- Коллaborация над текстом (соавторство, перевод и влияние переводчика, влияние редактора)
- Некоторые задачи за пределами текстов

Перейдем к примерам в RStudio

```
install.packages("stylo")
library(stylo)
stylo()
```

