

Контрастивный
корпусный анализ
специфических слов
русских классиков
XIX века

Борис Орехов



Метод и материал

23.10.2026

Контрастивный корпусный анализ

Контрастивный подход

- Сравнение с другими авторами эпохи
- Какие слова и обороты часто встречаются у одного автора, но редко или никогда — у других авторов того же периода?
- Речь не о частотных единицах в обычном смысле
- Ищем неочевидные даже при медленном чтении «любимые» слова писателей

Материал

- Национальный корпус русского языка
- Авторские тексты:
 - И. С. Тургенев — 846 тыс. слов
 - Ф. М. Достоевский — 1 млн 960 тыс. слов
 - И. А. Гончаров — 950 тыс. слов
 - Н. С. Лесков — 1 млн 254 тыс. слов
- Фон: случайные тексты XIX века — 27 млн словоупотреблений
- Исключены драматические и стихотворные произведения

Методика

- Лемматизация с помощью Mystem
- Служебные части речи удалены
- Мера **TF-IDF** (term frequency — inverse document frequency):
 - Понижает рейтинг слов, частотных в любом тексте
 - Повышает рейтинг слов, редких у других авторов
- Комбинирование с абсолютной частотностью и числом текстов
- Аналогично — для триграмм (последовательностей из трёх слов)

Как работает TF-IDF

- TF (term frequency) — как часто слово встречается в тексте автора:
 - $TF = (\text{число вхождений слова}) / (\text{длина текста})$
 - Показывает «любимость» слова у данного автора
- IDF (inverse document frequency) — насколько слово редкое в других текстах:
 - $IDF = \log(\text{общее число текстов} / \text{число текстов со словом})$
 - Чем реже слово у других, тем выше IDF
- Итог: $TF-IDF = TF \times IDF$
 - Высокий балл — слово часто у автора и редко у остальных
 - Низкий балл — слово частотно везде (служебное, общее) или редко у автора
- Смысл: мера «отсеивает» общезыковую частотность и оставляет авторскую специфику

Результаты

Тургенев: специфические слова

- **полузавядший** — уникальное, сентиментальное:
 - «...от ее **полузавядших** лепестков веяло другим, еще более тонким запахом...»
- **предрассветный** — уникальное:
 - «В саду пел соловей свою последнюю, **предрассветную** песнь»
- **лоснистый** — не употребляет никто, кроме Тургенева:
 - «...вся ее фигура, от **лоснистых** волос до кончика ботинки, была так изящна...»
- **хохотанье** — других употреблений в XIX веке нет:
 - «...преследуемый звонким, но не злым **хохотаньем**, убежал к себе в комнату...»
- **мякенький** — три раза, суффикс -еньк-:
 - «Ты славный малый; но ты все-таки **мякенький**, либеральный барич...»
- **шапонька** — не прижившаяся форма:
 - «...а только уж одна Васина **шапонька** по воде плывет»

Тургенев: обороты и редкие слова

- **повиливать** — редкое, нравится больше, чем другим:
 - «...всё ходила за ним, **повиливая** хвостиком»
- **злюка** — редкое:
 - «Да, думалось мне, эта **злюка** – это новый тип...»
- **заробеть** — редкое:
 - «**Заробел** так, что пересказать невозможно»
- **улицезреть** — почти ни у кого:
 - «...свет, часть которого он **улицезрел** у нее в доме, отталкивал его...»
- **брюхач** — редкое:
 - «Не поеду я к этому **брюхачу**...»

Уникальные обороты: «Голос изменил ей», «С небольшой запинкой», «Начал он неверным голосом», «После небольшого молчанья», «Сквозь частую сетку», «Словно на цыпочках»

Достоевский: диминутивы и усилители

- **жертвочка** — диминутив, много раз:
 - «...он обнимал и придерживал меня с наслаждением, как **жертвочку**»
- **суффикс -еньк-**:
 - **вороватенький**: «...глазки такие вороватенькие, в халате засаленном...»
 - **горденький**: «Я, дескать, сам люблю горденьких»
 - **щекотливенький**: «...щекотливенькая она, а важная»
- **приставка пре-**:
 - **прекомический**: «Прекомические иногда случаи случаются в этом роде-с»
 - **предосадный**: «...в игорные залы я вошел с предосадным чувством»

Достоевский: уникальные слова и обороты

- всепримирение — характерное, «Дневник писателя», «Подросток»: «...высшая русская мысль есть **всепримирение** идей»
- порывчатость — уникальное, «Идиот» и др.: «...вся эта **порывчатость**, горячка и нетерпение – не что иное, как бессознательное отчаяние...»
- фартикультяпность — уникальное: «Остолоп! Никакой **фартикультяпности** нет...»
- фраппировать — «ошеломлять»: «Знал бы только, что это вас так **фраппирует**...»
- единнительный — уникальное, до него только Ломоносов: «...в наше, столь народное, столь **единнительное** время...»
- замертветь: «...целое, которое без него распадется или **замертвеет**»
- **Обороты** (уникальные): «Я говорю лишь про», «Похолодев от испуга», «Я вас давеча», «И вот, верите ли», «А почти как»

Достоевский и Кропоткин: лексическая улика

- **скорлупчатый** — «Идиот»:
 - «...оно коричневое и **скорлупчатое**, пресмыкающийся гад...»
- П. А. Кропоткин, «Поездка в Окинский караул» (за четыре года до романа):
 - «Тут графит имеет строение несколько **скорлупчатое**...»
- Больше в XIX веке нигде не встречается
- Имя Кропоткина в наследии Достоевского не упоминается
- Гипотеза: возможное знакомство Достоевского с трудами Кропоткина

Гончаров: жанровая специфика

- **марсофал** — морской термин, «Фрегат "Паллада"», больше ни у кого
- **оголландившийся** — уникальное, единственный раз в корпусе: «...населены голландцами, **оголландившимися** французами и отчасти англичанами»
- **расстановисто** — любимое, все романы:
 - «Поди сюда! – **расстановисто** и настойчиво произнес Обломов»
- **непривилегированный** — любимое, все романы:
 - «...при звуке **непривилегированного** поцелуя потрясутся своды...»
- **скакнуть** — все три романа + «Паллада»:
 - «...они **скакнули**, вытянулись, и тройка ринулась по дороге в лес»
- **пропекать** — больше других авторов:
 - «...на палубе **пропекает** насквозь тропическое солнце...»
- **притворщица, кадочка, архимедов рычаг**:
 - «О газетах потише – это Архимедов рычаг...»

Лесков: уникальные новообразования

- **избрехаться** — только Лесков:
 - «...избрехались, несли друг на друга всякую всячину...»
- **требоисправление** — попало в корпус только благодаря Лескову:
 - «...за обязательные **требоисправления**»
- **отпырхиваться** — ни у кого, кроме Лескова:
- «...только как-то **отпырхивался** и отдувался, напоминая то свинью, то лошадь»
- **благочинствовать** — уникальное:
 - «Я пока уже третий год **благочинствую**...»
- **авторствовать**, нотаточка — учтены в «Словаре новообразований Н. С. Лескова»:
 - «...манила его руку “занотовать” еще одну “нотаточку”...»
- **лобочек, дедуня** — непривычные диминутивы:
 - «...в этом мягком лобочке веселости и приветов больше»; «Спасибо тебе, бедный **дедуня**!»
- **Обороты**: «Не к масти», «Сквозь крепкий сон»

Выводы

- Дополнения к традиционной авторской лексикографии
- Специфические черты стиля на фоне современного языка
- Разнообразие специфических слов

Спасибо!

nevmenandr@gmail.com



Борис Орехов
НИУ ВШЭ, ИРЛИ
РАН, МГУ