Диалог в русской нарративной прозе XIX века в аспекте количественного литературоведения

Борис Орехов

Диалог и нарратив

Постановка проблемы

- Диалог — самая заметная и «маркированная» зона нарративного текста
- Вопросы, которые ставит количественное литературоведение:
- Меняется ли доля диалога в литературе со временем?
- Одинаково ли она распределена внутри одного текста?
- Два разных уровня анализа: корпус ↔ отдельный текст
- Наш фокус — внутритекстовая композиция

Временная динамика

- Sobchuk O. *The Evolution of Dialogues: A Quantitative Study of Russian Novels (1830–1900)* // Poetics Today. 2016. 37(1). P. 137–154
- Корпус: 400 русских романов 1830–1900 (Lib.ru)
- Метрика: коэффициент диалогической живости (по Б. Ярхо)
  - = число реплик / число слов в романе
- Основной вывод: за XIX век диалогичность выросла примерно вдвое (≈0.01 → ≈0.02)
- Рост нелинейный: резкий скачок в 1830–40-е, затем плато
- Отдельные высокодиалогичные романы уже в 1830-е («авангард эпохи»)



Sobchuk O. The Evolution of Dialogues: A Quantitative Study of Russian Novels (1830–1900)

Объяснение

- Диалог — устройство для создания «гэпов» (по М. Стернбергу)
  - каждая реплика открывает вопрос → следующая закрывает → открывается новый
- Следовательно, диалог повышает читательский интерес
- Рост доли диалога = успешная литературная эволюция (в дарвиновском смысле)
- Нелинейность объясняется насыщением: роман не может целиком превратиться в пьесу
- Ранние «выбросы» (1830-е) — заимствование из европейского исторического романа (В. Скотт, А. де Виньи)

Опровержение

- Орехов Б. *Открытые компьютерные инструменты для решения задач оцифровки и анализа русскоязычного текста в области Digital Humanities //* Цифровые гуманитарные исследования. 2025
- Инструмент: `direct-speech-extractor-ru` (автор модуля — Д. Есаян)
  - accuracy = 0.989, precision = 0.998, recall = 0.96, F = 0.979
- Повторное исследование по методике Собчука не подтвердило его выводов
- На протяжении всего XIX века доля прямой речи ≈ 30% — без выраженного скачка
- Вывод: сфера DH столкнулась с кризисом воспроизводимости (Pashler, Wagenmakers 2012)



Орехов Б. Открытые компьютерные инструменты для решения задач оцифровки и анализа русскоязычного текста в области Digital Humanities

Роль в композиции

Исследовательский вопрос и гипотеза

- Вопрос: меняется ли доля прямой речи от начала к концу произведения?
- Конкурирующие гипотезы:
  - H_1: монотонный тренд (растёт или падает)
  - H_2: U- или ∩-образный профиль (арка)
  - H_0: равномерное распределение (нет эффекта)
- Почему это интересно:
  - если профиль есть — это композиционная характеристика нарратива
  - она ортогональна временной динамике, которую изучали Собчук и Есаян

Материал

- Корпус нарративной прозы — 506 текстов
- Хронология: XIX век (уточнить по корпусу)
- Объём: медиана ≈ 7,8 тыс. слов на пятую часть (≈ 39 тыс. слов на текст)
- Каждый текст разбит на 5 равных по символам частей
- Итого: 2530 наблюдений (506 × 5)

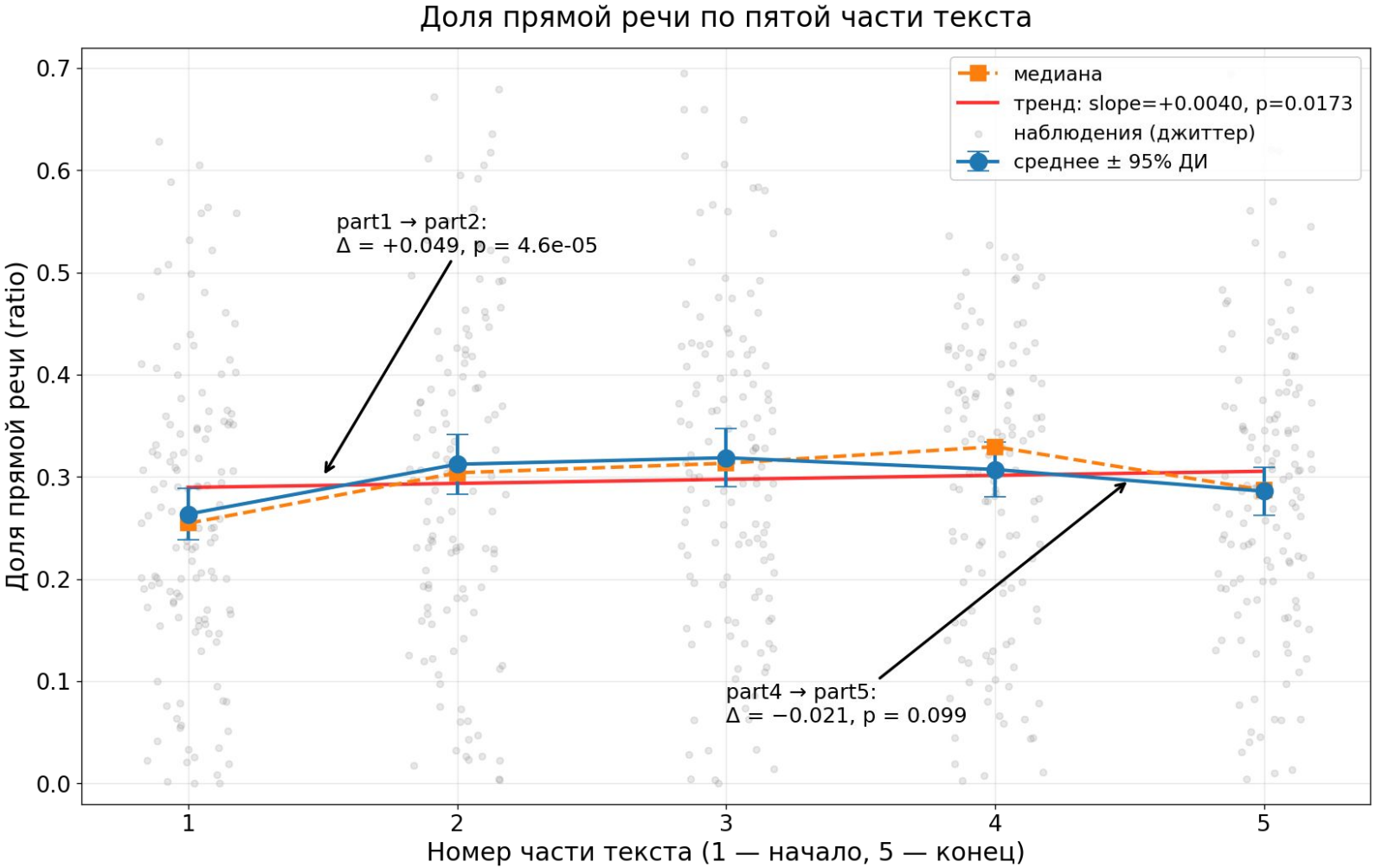
Метод

- Инструмент: `direct-speech-extractor-ru` v0.0.2 (Д. Есаян)
  - тот же, что использован в исследовании Д. Есаян
- Для каждой пятой части считаем:
  - `text_wordcount` — объём части
  - `direct_speech_wordcount` — объём прямой речи
  - `ratio = dswc / twc` — доля прямой речи
- Анализ:
  - описательные статистики + 95% ДИ (bootstrap для медианы)
  - линейная vs квадратичная регрессия (F-тест)
  - парные сравнения частей (Wilcoxon, Cohen's d, bootstrap по файлам)

Результат

Часть	mean	median	95% ДИ
1	0.257	0.253	[0.244; 0.271]
2	0.302	0.305	[0.287; 0.316]
3	0.296	0.296	[0.282; 0.310]
4	0.299	0.307	[0.285; 0.313]
5	0.286	0.290	[0.273; 0.299]

- Части 2–4 образуют «плато»
- Финал ниже плато, но выше начала


Доля прямой речи по пятой части текста
медиана
тренд: slope=+0.0040, p=0.0173
наблюдения (джиттер)
среднее ± 95% ДИ
part1 → part2:
Δ = +0.049, p = 4.6e-05
part4 → part5:
Δ = −0.021, p = 0.099
Доля прямой речи (ratio)
Номер части текста (1 — начало, 5 — конец)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
1
2
3
4
5

Результат

- Линейная модель: slope = +0.0055, r = +0.049, p = 0.014; R^2 = 0.002
  - формально значима, но описывает лишь 0.24% дисперсии и неверно передаёт форму
- Квадратичная модель: a = −0.0075, t = −3.98, p = $7 \cdot 10^{-5}$
  - F(1, 2527) = 15.84, p < 10^{-4} — квадратичный член значимо лучше линейного
- Знак `a < 0` ⇒ ∩-образный (inverted-U) профиль
- Bootstrap по файлам (2000 перевыборок целых текстов):
  - 95% ДИ для a = [−0.0097; −0.0055]
  - 100% перевыборок с a < 0 ⇒ робастно к кластеризации

Где локализован эффект

- Единственный сильный переход — 1→2: две трети текстов
- 2→3 и 3→4 — статистический шум (плато)
- 4→5 — слабый, но значимый спуск

Переход	Δ mean	Cohen's d	доля b > a	p
1→2	**+0.044**	**+0.372**	**67.8%**	**$< 10^{-10}$**
2→3	−0.006	−0.047	46.6%	0.21
3→4	+0.004	+0.031	52.2%	0.60
4→5	**−0.013**	**−0.115**	43.9%	**0.005**
1→5	+0.029	+0.224	60.9%	$5 \cdot 10^{-8}$

Интерпретация

Композиционная асимметрия

- Асимметричная ∩-арка с крутым восходящим и пологим нисходящим фронтом:
  - начало — экспозиция, ведётся преимущественно монологически
  - середина — «диалогическое плато», основное взаимодействие персонажей
  - финал — частичный возврат к авторскому резюмирующему повествованию
- Основной эффект локализован в начале текста, а не в конце
- Гипотеза: диалог — маркер «сценического» модуса, экспозиция его почти не использует

Величина эффекта и её трактовка

- Cohen's d для 1→2 = 0.37 — малый-средний эффект
- $R^2 < 0.01$ — эффект объясняет < 1% дисперсии
- Почему это не «слабый результат»:
  - межтекстовая дисперсия (SD ≈ 0.15–0.17) на порядок больше внутритекстового сдвига
  - но эффект систематический: 67.8% текстов сдвигаются в одну сторону
  - робастен к bootstrap по файлам (100% перевыборок сохраняют знак)
- Аналогия: в психологии такие эффекты считаются надёжными при высокой универсальности

Заключение

- Доля прямой речи в русском нарративе XIX века не описывается монотонным трендом
- Профиль значимо ∩-образный (a = −0.0075, bootstrap ДИ [−0.0097; −0.0055])
- Основной эффект — переход от первой части ко второй (Δ = +0.044, d = 0.37, 67.8% текстов)
- Середина — плато, финал — слабый спуск
- Гипотеза композиционной асимметрии: экспозиция монологична, диалог нарастает к сцене, финал возвращается к авторскому голосу