

Papers and inspiration



Соколова Ирина

Никишина Ирина

Тихомиров Даниил

А. А. Бонч-Осмоловская

Даниил Скоринкин

Overview

1. Методы изучения прямой речи:
 1. Неколичественные
 2. Количественные
2. Извлечение имен персонажей
3. Разметка

1.1

Неколичественные методы

Типы чужой речи

1. Прямая
2. Косвенная
3. Несобственно-прямая

Больше
типов!

1. Неразвёрнутая прямая
2. Замещённая прямая
3. Пересказанная
4. Несобственно-авторская

Больше
подтипов!

Несобственно-прямая:

1. Косвенно-прямая
2. Изображённая
3. Абсорбированная
4. Цитатная
5. Малоформатная
6. Внутренний монолог
7. Поток сознания

Всех
интересует
несобственно
-прямая

- Контаминация прямой и косвенной речи
- Хронологически возникает позднее
- Нет вводящего глагола
- Формально входит в речь автора, но представляет собой чужую речь

Речь автора

- Информация о речевом акте, его цели, состоянии персонажа, сопутствующих действиях
- Конкретизирует и дополняет речь героя, может прямо ей противоречить
- Проявление авторской индивидуальности

Речь персонажа

- Контраст с авторской речью
- Односоставные предложения
- Краткость
- Междометия
- Эмоциональные эпитеты
- Стиль
- Грамматика

Речь персонажа

Они шли по мирной Мариинско-Благовещенской улице и разговаривали первый раз. Чудесным образом разговор этот был почти безглагольным, состоял из одних перечислений имен и вздохов, выдохов и междометий... Толстой? Да! Крейцера соната? Нет, Анна Каренина! О, да! Достоевский? Конечно! «Бесы»! Нет, «Преступление и наказание»! Ибсен! Гамсун! Виктория! Голод! Ницше! Вчера! Далькроз? Кто? Не знаю! Рахманинов! Ах, Рахманинов! Бетховен! Конечно! Дебюсси? А Глиер? Великолепно! Чехов? Дымов? Короленко! Кто? И я! Но «Капитанская дочка»! Какое счастье! Боже! Невероятно! Никогда ничего подобного! Еврейское? Шолом Алейхем? Да, в соседнем доме! Нет, Блок, Блок! Надсон? Гиппиус! Никогда! Совсем, совсем не знаю! О, это надо, надо! История античности! Да, греки, греки!

Речь персонажа

Стилистически окрашенный порядок слов:

- Порядок главных членов предложения
- Порядок частей сказуемого
- Порядок компонентов словосочетания
- Дистантное расположение определения и определяемого

Речь персонажа

- Обычно-то я человек молчаливый, а сейчас прорвало **будто**.
- Видно, **запропал Эрастушка**, сдал его тот огрызок ушастый в полицию...
- **Большущая** наметилась **операция**...
- Что, съел, **лапша итальянская**?

Кто говорит?

Автор

- Однородные члены
- Придаточные причины
- Расширенные метафоры
- Сравнения

Обломов

- Лексические повторы
- Эпитеты
- Фразеологизмы
- Вопрос-ответ

Новые явления

- Размывание границ между речью автора и персонажей
- Необычное графическое оформление

Новые явления

*Где-то там сейчас вкалывал его отец, погребенный
в сплетении труб и рычагов, перепачканный неф-
тью, за самой грязной работой. Потому что
(для вас с матерью из шкуры вон лезу хоть бы
спасибо сказал)*

*чистую и интересную эти халявишки подгребли
под себя. Но Лешка все исправит. Очень скоро эти
гады превратятся в обычных людей – только, в
отличие от честных буровиков вроде папы,
(что ты знаешь о любви козявка)*

Новые явления

Я смело вхожу. Смело. Я вхожу. Добрый вечер. Смело. Вечер добрый, радость моя, сделайте мне удовольствие. Я достаю патефон со шкафа. Довоенный патефон, купленный тогда-то и там-то. Кем-то.

Новые явления

*И после мучительного думанья внезапное озарение:
а он для нас такой большой, что она его просто не видит.
Она думает – дом, или – гром... Он думает – просто блоха.*

А кругом говорили, что «ах, какая красота!»

Статьи

- А. М. Ломов *Чужая речь в письменном тексте*
- Чой Чжи Ен Способы передачи чужой речи и тип художественного повествования (на материале рассказа А. П. Чехова «Скрипка Ротшильда»)
- О. А. Кострова Персонажно-авторская речь как способ моделирования чужого сознания
- А. А. Вердеш Прямая речь в художественных текстах неклассической парадигмы: новые явления
- И. А. Вороновская Порядок слов в авторской речи Б. Акунина
- Е. А. Казанкова Аспекты лингвистической характеристики авторского сопровождения прямой речи
- О. В. Шаталова Автор и персонаж: синтаксическая репрезентация в несобственно-прямой речи
- О. В. Арзямова Авторское оформление прямой речи как средство выражения идиостилены в новейшем русском художественном дискурсе
- Е. А. Покровская Чужая речь и диалог в потоке сознания (на материале русской литературы XX века)
- А. Н. Васильева Художественная речь
- Л. П. Позняк Особенности речевой характеристики главного персонажа в романе С. Шелдона «Ничто не вечно»

1.2

Количественные методы

Dialogism in the Novel: A Computational Model of the Dialogic Nature of Narration and Quotations

Muzny, G.,
Algee-Hewitt, M.,
Jurafsky, D.

- Подход вдохновлен работами М.М. Бахтина;
- Прямая речь «растворена» в тексте с разной степенью концентрации;
- Можно отделить прямую речь от не-прямой;
- И различать авторский стиль;
- Основан на POS-тэгах.

Muzny, G.,
Algee-Hewitt, M.,
Jurafsky, D.

Table 3 Rules for the QuoteAnnotator that help it extract quotes across a wide range of formats

#	Case	Rule
1	Quote delineated by regular ASCII quote marks (")	Ending quote must be followed by white space or punctuation
2	Quote delineated by single quote marks (')	Starting quote must be preceded by white space or punctuation and ending quote must be followed by white space or punctuation
3	Quote delineated by smart quotation marks ("")	Ending quote must pair with beginning quote and level of embeddedness must match
4	Smart quotes face the wrong direction ("")	Attempt extraction with quotes converted to ASCII
5	Problematic apostrophes	Attempt extraction disallowing single-quote delineated quotes

Muzny, G.,
Algee-Hewitt, M.,
Jurafsky, D.

Table 5 The eight factors that reliably distinguish narration from dialogue produced by using MDA on the dialogue in our corpus

Factor	Direction of dialogism (δ)	Correlated features
1	+	Interjections 1st-person pronouns 2nd-person pronouns Non-3rd-person singular present tense verbs Modals Bare verbs
2	+	Existential there 'it' pronouns 3rd-person singular present tense verbs
3	—	Gerunds
4	—	Coordinating conjunctions Prepositions and subordinating conjunctions
5	—	Adjectives
6	—	Determiners Nouns
7	—	Past tense verbs
8	—	3rd-person pronouns

Muzny, G.,
Algee-Hewitt, M.,
Jurafsky, D.

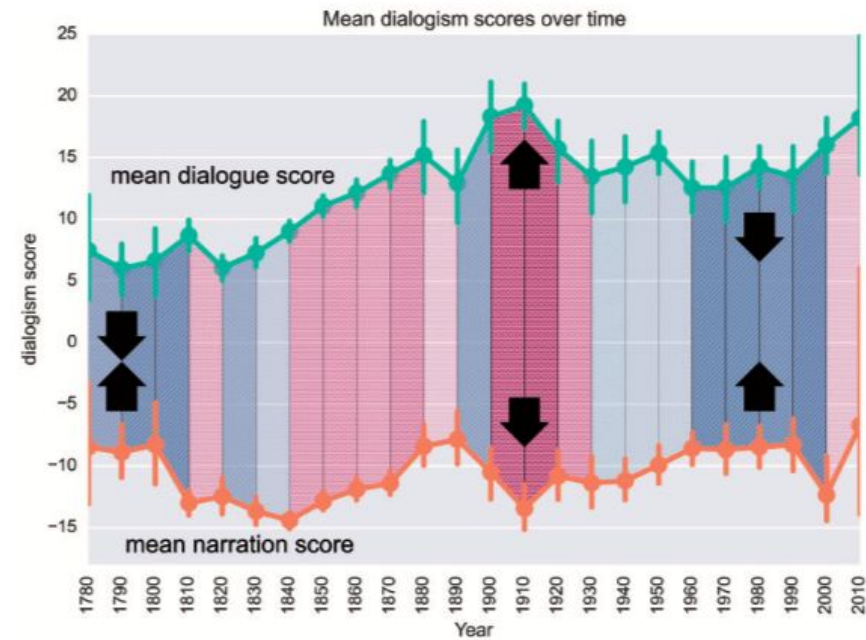


Fig. 5 Dialogism scores for narration and dialogue over time, bucketed per decade, rounded down, for every novel in our corpus. Periods where the difference in dialogism scores is unexpectedly high (shaded with circles) or low (shaded with diagonal lines) are highlighted with shading and arrows. Confidence intervals are shown at 95% confidence

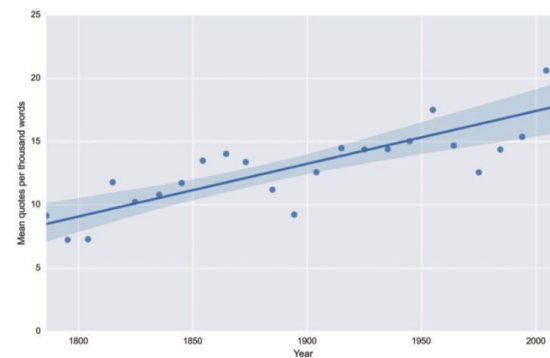


Fig. 2 Number of quotes per thousand words from 1782 to 2011, bucketed per decade, rounded down, so all novels from 1800 to 1809 contribute to the 1800 data point. This approach is used in the all bucketed figures. The line of best fit shows an upward-sloping trend with an r^2 value of 0.65

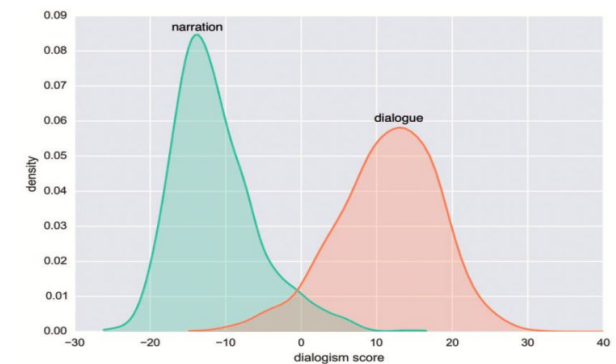


Fig. 3 Dialogism scores for narration and dialogue for all novels in our corpus

Using Models of Lexical Style to Quantify Free Indirect Discourse in Modernist Fiction

Brooke, J.,
Hammond, A.,
Hirst, G.

- Проблема: несобственно-прямая речь;
- Задача: научиться размечать модернистские тексты так же, как это делают студенты-литературоведы
- Лексический анализ:
 - Объективные;
 - Абстрактные;
 - Литературные;
 - Неформальные;
 - Конкретные;
 - Субъективные;
- ~900 слов в базовом словаре.

Brooke, J.,
Hammond, A.,
Hirst, G.

<said who="#lily" discourse="free"
aloud="false">She looked at the steps; they
were empty; she looked at her canvas; it was
blurred. With a sudden intensity, as if she saw
it clear for a second, she drew a line there,
in the centre</said>. It was done; it was fin-
ished. <said who="#lily" discourse="direct"
aloud="false">Yes</said>, she thought,
laying down her brush in extreme fatigue,
<said who="#lily" discourse="direct"
aloud="false">I have had my vision</said>.

Brooke, J.,
Hammond, A.,
Hirst, G.

Table 2. Stylistic profiles for various types of discourse in *To the Lighthouse* and 'The Dead'. 'Narration' refers to passages identified as the narrator's words; 'FID' to passages of free indirect discourse; 'thought' to passages of silent direct discourse; 'speech' to passages of direct discourse spoken aloud

Text	Discourse	Type count	Styles					
			Objective	Abstract	Literary	Colloquial	Concrete	Subjective
<i>To The Lighthouse</i>	Narration	765	0.00	0.00	0.00	0.00	0.00	0.00
	FID	2,916	0.08	0.17	0.02	-0.02	-0.15	0.02
	Thought	212	-0.15	0.21	0.07	0.30	-0.20	0.08
	Speech	172	-0.32	0.14	0.06	0.49	-0.20	0.11
'The Dead'	Narration	1,325	-0.01	0.02	0.08	0.04	0.00	0.09
	FID	400	-0.13	0.19	0.10	0.19	-0.15	0.11
	Thought	57	-0.43	0.18	0.12	0.74	-0.30	0.22
	Speech	651	-0.11	0.23	0.06	0.27	-0.19	0.16

Table 3. Stylistic profiles for the FID of various characters in *To the Lighthouse* and 'The Dead'. 'Other' refers to all characters in 'The Dead' other than Gabriel considered together

Text	Character	Type count	Styles					
			Objective	Abstract	Literary	Colloquial	Concrete	Subjective
<i>To The Lighthouse</i>	Mrs Ramsay	805	0.07	0.24	0.00	0.03	-0.22	0.03
	Mr Ramsay	70	0.09	0.58	0.27	0.01	-0.49	0.00
	William Bankes	248	-0.01	0.19	0.03	0.14	-0.17	0.08
	Lily Briscoe	1,485	0.06	0.17	0.03	-0.02	-0.15	0.01
	James Ramsay	540	-0.06	0.03	0.08	0.06	-0.03	0.02
	Cam Ramsay	381	-0.10	0.04	0.04	0.10	-0.06	0.00
	Charles Tansley	138	-0.07	0.21	-0.07	0.22	-0.23	0.05
	Gabriel	358	-0.12	0.21	0.12	0.17	-0.17	0.10
'The Dead'	Other	85	-0.30	0.06	0.02	0.44	-0.07	0.14

Other topics

- 'An Ocean Where Each Kind...': Statistical Analysis and some Major Determinants of Literary Style (Burrows, J.F.)
- Digital Literary Studies Corpus Approaches to Poetry, Prose, and Drama (Hoover, D.L., Culpeper, J., O'Halloran K.)
- Burrowing into Translation: Character Idiolects in Henryk Sienkiewicz's Trilogy and its Two English Translations (Rybicki J.)

The Dialogic Turn and the Performance of Gender: The English Canon 1782-2011

Table 1: Positive and negative contributors to dialogism score

+	-
Present tense verbs, bare verbs, modals, 1 st /2 nd /'it' pronouns, Wh-pronouns, interjections, existential there, adverbs	past tense verbs, 3 rd person pronouns, gerunds, particles, nominalizations, determiners, Wh-determiners, prepositions and subordinating conjunctions, adjectives

Muzny, G.,
Algee-Hewitt, M.,
Jurafsky, D.

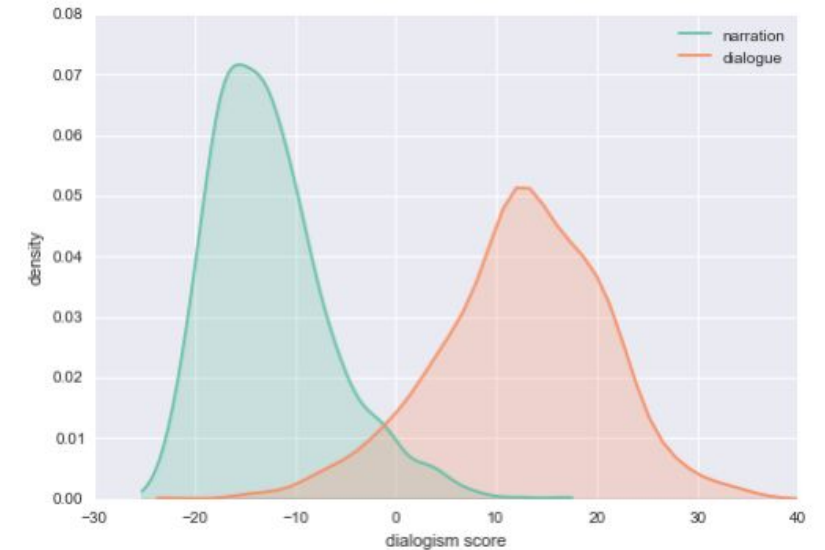
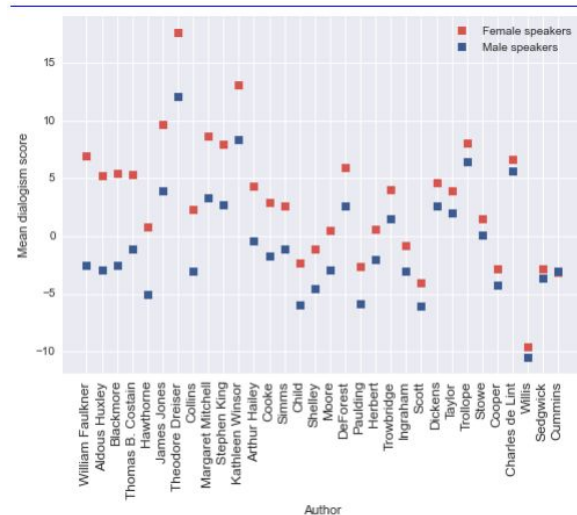


Figure 4: dialogism scores for narration and dialogue

3

Выделение персонажей

Identification of Speakers in Novels

Category	Example
Implicit speaker	<i>"Don't keep coughing so, Kitty, for heaven's sake!"</i>
Explicit speaker	<i>"I do not cough for my own amusement," replied Kitty.</i>
Anaphoric speaker	<i>"Kitty has no discretion in her coughs," said her father.</i>

	IS	AS	ES	Total
INDIVIDUAL	52.5	67.3	100.0	67.3
NEIGHBORS	63.1	76.4	100.0	74.8
ORACLE	74.2	69.1	99.1	80.1

Table 6: Speaker identification accuracy (in %) on Austen's *Emma* by the type of utterance.

Features	Novelty
Distance to Utterance	No
Speaker Appearance Count	No
Speaker Name in Utterance	No
Unsupervised Actor-Topic Model	Yes
Vocative Speaker Name	Yes
Neighboring Utterances	Yes
Gender Matching	Yes
Presence Matching	Yes

Feature	Example
start of utterance	<i>"Kitty ...</i>
before period	<i>... Jane.</i>
between commas	<i>..., Elizabeth, ...</i>
between comma & period	<i>..., Mrs. Hurst.</i>
before exclamation mark	<i>... Mrs. Bennet!</i>
before question mark	<i>... Lizzy?...</i>
vocative phrase	<i>Dear ...</i>
after vocative phrase	<i>Oh! Lydia ...</i>
2nd person pronoun	<i>... you ...</i>

Table 3: Features for the vocative identification.

Hua He
D. Barbosa
G. Kondrak

Mr. Bennet, his coachman, and the Archbishop walk into a bar but only one of them gets recognized:

On The Difficulty of Detecting Characters in Literary Texts

*Hardik Vala,
David Jurgens,
Andrew Piper,
Derek Ruths*

- 8 stages
- Social graph
- Reasons to be merged
- Speech verbs
- NER

	Precision	Recall	F1
Sherlock Holmes Stories	0.5910	0.5335	0.5608
Pride and Prejudice	0.7635	0.6879	0.7237
The Moonstone	0.3943	0.4613	0.4294

Table 2: Accuracy of proposed system without stage 7.

4

Разметка

TEI разметка

- A TEI Schema for the Representation of Computer-mediated Communication (*Beweißwenger, M., Ermakova, M. A. Geyken, A., I. Lemnitzer, I. Storrer, A.*)
- The Text Encoding Initiative: Its History, Goals, and Future Development (*Nancy M. Ide*)
- The Text Encoding Initiative and the Study of Literature (*Cummings, J.*)

*Hammond, A.
Brooke, J.*

GutenTag: A User-Friendly, Open-Access, Open-Source System for Reproducible Large-Scale Computational Literary Analysis

- Тексты Gutenberg Project;
- Метаданные: жанр, автор, издание
- Токенизация, лемматизация
- POS
- Структурные части текста, диалоги
- Распознавание персонажей
- TEI
- Встроенные инструменты анализа

The screenshot displays the 'DEFINE SUBCORPUS 1' interface of the GutenTag system. It features several filter sections:

- GENRE:** Includes checkboxes for 'Prose fiction', 'Poetry', 'Drama', 'Prose non-fiction', 'Periodicals', and 'All'.
- AUTHOR:** Contains fields for 'Author Name', 'Author Birth' (with 'from' and 'to' date pickers), 'Author Death' (with 'from' and 'to' date pickers), 'Author Gender*' (with radio buttons for 'Either', 'Just male', and 'Just female'), and 'Author Nationality*' (with a dropdown menu).
- TEXT:** Includes fields for 'Title of the text', 'Language*' (with a dropdown menu), 'Date of Publication*' (with 'from' and 'to' date pickers), 'Country of Publication*' (with a dropdown menu), 'Library of Congress Classification*' (with a dropdown menu), 'Library of Congress Subject*' (with a dropdown menu), and 'Collections (anthologies, etc.):*' (with radio buttons for 'All texts', 'Exclude collections', and 'Only collections').
- WITHIN-TEXT:** A section with a plus icon and a dropdown menu.
- LEXICAL FILTER:** A section with a plus icon and a dropdown menu.

Спасибо за
внимание!

