

# (Не)прямая речь



Руководители проекта:

Анастасия Бонч-Осмоловская  
Даниил Скоринкин

Тихомиров Даниил

Соколова Ирина

Шалганова Татьяна

Никишина Ирина

## О чем «речь»?

- Что мы хотим получить?  
*(Наши глобальные цели...)*
- ...И конкретные задачи
- Проблемы и возможности



# Наши глобальные цели

Корпус



Парсер



## Наши глобальные цели

- Научиться выделять различные виды чужой речи из художественных (и не только) текстов...
- и «действующих лиц» в каждом акте такой речи.
- Парсер, который будет уметь это делать самостоятельно...
- и создание размеченного корпуса с его помощью.



## Какая бывает речь?

- Прямая речь
- Диалог
- Косвенная речь
- Несобственно-прямая речь
- Замещённая прямая речь
- Неразвёрнутая прямая речь

# Элементы речевой ситуации

- Говорящий
- Адресат
- Слушающий
- Пропозиция
- Контекст (до, после)
- Глагол речи и его признаки

# Материал исследования

- Русский язык (английский язык?)
- Русская проза 19 и 20 века

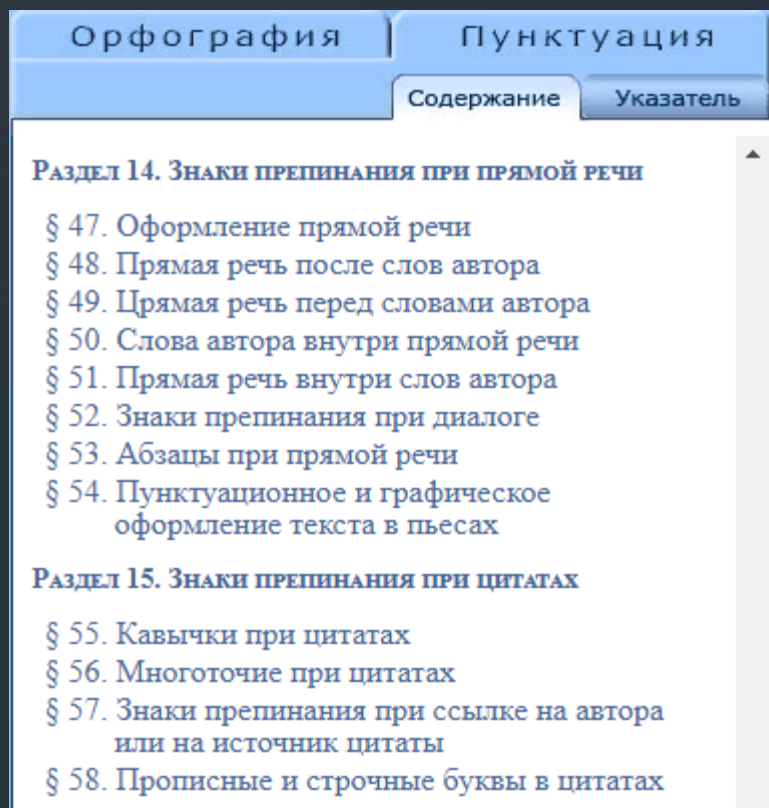
Корпус тьюториала “Стилометрия русской прозы”  
*О. Собчука и А. Шеля (Тартуский университет) на  
Второй Московско-тартуской школе по цифровым  
гуманитарным наукам*

# Наши конкретные задачи

1. Розенталь скрипт
2. Update разметки TEI (онтология признаков)
3. Update скрипта (тестовый файл с разными типами речи)
4. Соотношение онтологии и текстов.
5. Комбинации признаков VS типы прямой речи
6. ...



# Розенталь Script



- К..... сказал (.....): "К..... .....?!"
- К..... сказал:  
-- К..... !?...
- «К..... », — К.../ (.....) сказал .....
- -- К..... , -- К.../ (.....) сказал .....
- «К.....? — сказал ..... — К.....?!...».
- К... сказал: "К.....?!... ", -- к.../ (.....) сказал.....
- " К.....?! " -- " К.....". -- " К.....!" -- " К.....".
- " К.....?" — сказал к.....! " К.....", -- к.....

# Глаголы речи

- *(говорить, сказать, рассказать, заметить, обратиться, воскликнуть, закричать, прошептать, спросить, спрашивать, ответить, вставить, прервать, заговорить, начать, продолжать, подтвердить, пояснить, согласиться, поддакнуть, напомнить, посоветовать, упрекнуть, решить, подумать)*
- *(вопрос, ответ, слова, восклицание, голос, шёпот, звук, крик, мысль)*
- *(вспомнить, обрадоваться, огорчиться, удивиться, обидеться, возмутиться, ужаснуться и т. п.),*
- *улыбнуться, усмехнуться, рассмеяться, захохотать, вздохнуть, нахмуриться, подойти, подбежать, вскочить и т. п.). И те и другие глаголы допускают добавление к ним глаголов речи (обрадовался и сказал; удивился и спросил; улыбнулся и ответил; подбежал и воскликнул и т. д.)*

# TEI разметка

**<q>** (quoted) contains material which is distinguished from the surrounding text using quotation marks or a similar method, for any one of a variety of reasons including, but not limited to: direct speech or thought, technical terms or jargon, authorial distance, quotations from elsewhere, and passages that are mentioned but not used.

**<said>** (speech or thought) indicates passages thought or spoken aloud, whether explicitly indicated in the source or not, whether directly or indirectly reported, whether by real people or fictional characters.

**@direct** may be used to indicate whether the quoted matter is regarded as direct or indirect speech.

**@aloud** may be used to indicate whether the quoted matter is regarded as having been vocalized or signed.

**<quote>** (quotation) contains a phrase or passage attributed by the narrator or author to some agency external to the text.

**att.global.source** provides an attribute used by elements to point to an external source.

**@source** specifies the source from which some aspect of this element is drawn.

**<cit>** (cited quotation) contains a quotation from some other document, together with a bibliographic reference to its source. In a dictionary it may contain an example text with at least one occurrence of the word form, used in the sense being described, or a translation of the headword, or an example.

**<mentioned>** marks words or phrases mentioned, not used.

**<soCalled>** contains a word or phrase for which the author or narrator indicates a disclaiming of responsibility, for example by the use of scare quotes or italics.

- **<said></said>**
- **@type = "direct" / "indirect" / "free indirect"**
- **@who = "#person\_id"**
- **@corresp = "#person\_id"**
- ...

### Что понятно

- Пунктуационное оформление «традиционной» прямой речи;
- Более-менее — какие возможности TEI можно использовать

### Что непонятно

- Что является прямой речью, а что не является?
- Как фильтровать не-диалоги, похожие на диалоги?

# Траблы и возможности

- Диалогичность – не бинарный параметр, а градиентный
  - Muzny, Algee-Hewitt, Jurafsky (2017)
- Квантифицируемые лингвистические признаки, указывающие на неавторскую речь:
  - «Мягкие»: показатели модальности, наличие 1-го и 2-го лица, отсутствие причастных оборотов...
  - «Жесткие»: обращения, междометия...



# Траблы и возможности

- «Поток сознания»:
  - ...твоим на мотив из какой-то старой оперы пока иезуиты не обнаружили что он масон да он разгуливал с некоторыми из этих шинфейнеров или как там они себя называют и нес свой обычный вздор показал мне какого-то коротышку шеи нет почти и говорит это Гриффит восходящая звезда умен необыкновенно не знаю на вид что-то непохоже все что могу сказать...
  - Дж. Джойс, *Улисс*

# Траблы и возможности

- Письма:
  - Вам, наверно, приятно будет, сударыня, узнать об одной черте характера господина де Вальмона, весьма, как мне кажется, отличающейся от всех тех, какими вам рисовали его облик. Как тягостно иметь нелестное мнение о ком бы то ни было, как обидно находить одни лишь пороки у тех, кто, в сущности, обладает всеми качествами, необходимейшими для того, чтобы любить добродетель!...
  - Ш. де Лакло, *Опасные связи*

# Траблы и возможности

- Отклонения от традиционной пунктуации:
  - Да, дорогой автор, именно так: придти к нему домой, позвонить звучным велосипедным звонком у дверей – пусть он услышит и откроет. Кто там? Там-там, здесь живет Начальник такой-то? Здесь. Открывайте, пришли, чтобы спросить и получить правдивый ответ. Кто? Те Кто Пришли. Приходите завтра, сегодня уже поздно, мы с женой спим. Проснитесь, ибо наступила пора сказать правду. О ком, о чем? О ребятах вашей конторы. Почему ночью?
  - С. Соколов *Школа для дураков*

# Траблы и возможности

- Цитации:
  - Докладчик привёл слова М. В. Ломоносова: «России могущество будет прирастать Сибирью» — и этим закончил своё выступление.
  - Пример с [orfogrammka.ru](http://orfogrammka.ru)
- И много всего другого.

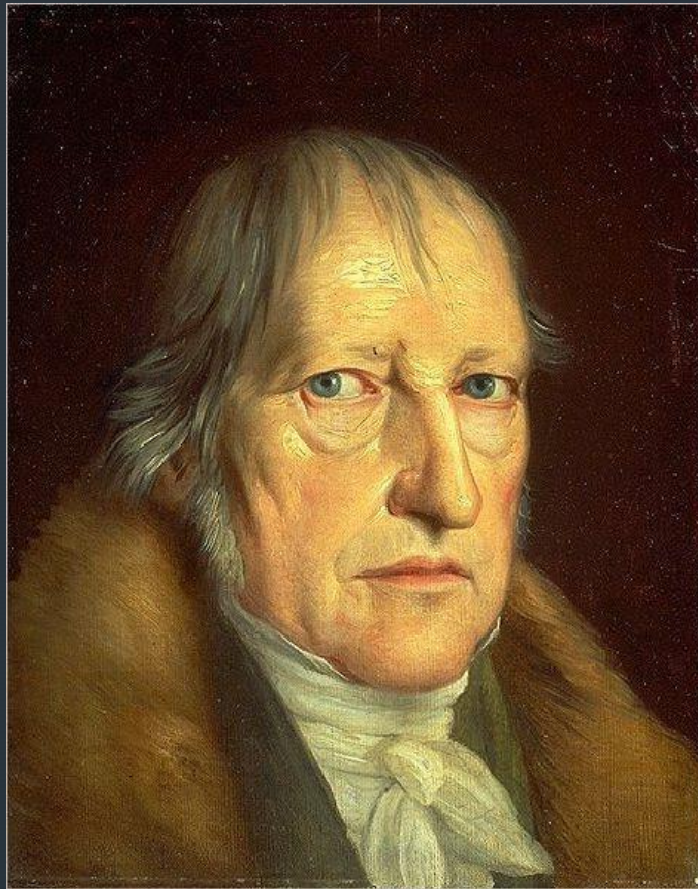
# Выводы:

- У нас есть план!
- Составить правила для выделения пунктуационно или лексически обозначенной прямой речи. TEI разметка. Принципы выделения признаков.
- Понять, как вписать в нашу систему сложные пограничные случаи и как отсеивать лишнее.
- Собрать разметчик и подготовить тестовый материал
- Протестировать и поправить (x100)
- Разметить корпус
- PROFIT!





«Речь — удивительно сильное средство,  
но нужно иметь много ума, чтобы  
пользоваться»



*Гегель Георг Вильгельм Фридрих*