

# Использование марковских цепей при анализе вариативности в русских диалектных корпусах

Георгий Алексеевич Мороз, Светлана Сергеевна Земичева

Международная лаборатория языковой конвергенции (НИУ ВШЭ)

«Основные приложения математики», НИУ ВШЭ

презентация доступна по ссылке: [tinyurl.com/202aha9q](https://tinyurl.com/202aha9q)

23 марта 2023

# План доклада

Лингвистика: мифы и реальность

Обо мне

Корпусная лингвистика

Использование причастий и деепричастий в устной речи

- умеет читать на всех письменностях мира

- умеет читать на всех письменностях мира
- знает все языки на свете

- умеет читать на всех письменностях мира
- знает все языки на свете
- умеет распознавать каждый язык на слух

- умеет читать на всех письменностях мира
- знает все языки на свете
- умеет распознавать каждый язык на слух
- может рассказать о происхождении каждого слова

- умеет читать на всех письменностях мира
- знает все языки на свете
- умеет распознавать каждый язык на слух
- может рассказать о происхождении каждого слова
- пишет без ошибок и знает все правила орфографии

- умеет читать на всех письменностях мира
- знает все языки на свете
- умеет распознавать каждый язык на слух
- может рассказать о происхождении каждого слова
- пишет без ошибок и знает все правила орфографии
- не знает математики и программирования



- умеет читать на всех письменностях мира
- знает все языки на свете
- умеет распознавать каждый язык на слух
- может рассказать о происхождении каждого слова
- пишет без ошибок и знает все правила орфографии
- не знает математики и программирования
- все вышеперечисленное, конечно, неправда

# Лингвистика

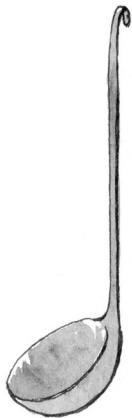
- прескриптивная

- прескриптивная
- вся остальная (дескриптивная)
  - каталогизация языкового разнообразия, описание языковых контактов
  - исследования грамматики языка
  - исследования распределения грамматических особенностей в языках мира
  - исследования когнитивных способностей человека и других животных, связанных с языком
  - исследования в области синтеза и распознавания речи и языка
  - создание компьютерных инструментов для решения самых разных задач
  - вспомогательные инструменты лингвистического исследования и документации
  - корпусная лингвистика
  - исследования в области NLP, языковых моделей и т. п.
  - симуляционные модели в лингвистике

# Прескриптивная vs. дескриптивная лингвистика

Назовите, пожалуйста, что изображено на картинке.

(рисунок Тани Пановой)



# Прескриптивная vs. дескриптивная лингвистика

Это часть опроса Ивана Левина:



# План доклада

Лингвистика: мифы и реальность

Обо мне

Корпусная лингвистика

Использование причастий и деепричастий в устной речи

- полевые исследования (29 поездок)
- фонетист, фонолог, количественный лингвист
- езжу на Кавказ
- преподаю статистику и R (язык программирования)
- написал несколько лингвистических пакетов для R
  - `lingtypology`
  - `phonfieldwork`
  - `lingglosses`

# План доклада

Лингвистика: мифы и реальность

Обо мне

Корпусная лингвистика

Использование причастий и деепричастий в устной речи

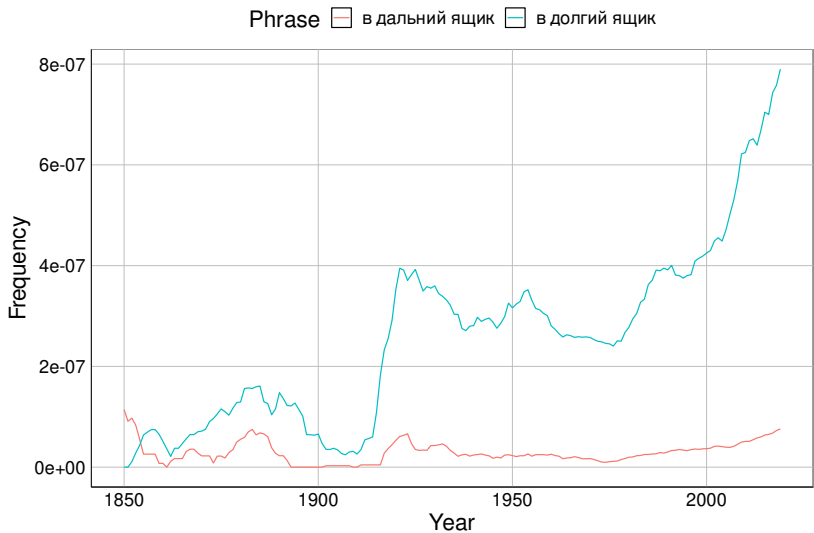


Корпусная лингвистика — это область лингвистики, которая занимается исследованием языковых явлений на материале некоторых собраний языкового материала. В большинстве случаев это письменные тексты, однако это может быть аудио и даже видео корпус.

Среди корпусов русского языка можно назвать:

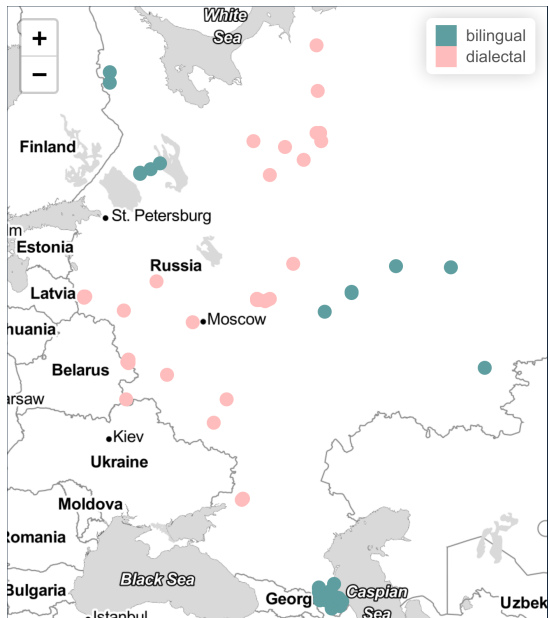
- **Национальный корпус русского языка**
  - более 1.5 млрд слов
  - много подкорпусов (газетный, устный, параллельный, диалектный, поэтический, исторические)
- Google Books Ngram Viewer
- ...

# Отложить в ... ящик



на основе Google Books Ngram Viewer

# Диалектные устные корпуса лаборатории языковой конвергенции



# Диалектные устные корпуса лаборатории языковой конвергенции

Malinino Corpus

Corpus ▾ Statistics ▾ About ▾ Help | rus en |

Query / lemma: 'рука'[c]  
Number of results: 35

Recording

Transcription

	Так, дай <b>руку</b> покажите.
	А теперь всё, а говорю, до того износилось, вот как это работешь <b>рукой</b> ... < смеется >
	Тоже пил, а попалась <b>и</b> <b>руки</b> отперлась.
	Без <b>рук</b> остался, всё равно пил.
	Под <b>руку</b> тебе чего, чего любишь... И ешь.
	Вот на эту, на себелу, мы говорим, право чтоб слезть, и прямо сразу наберем, и вот прямо <b>рукой</b> постелить, и вылезешь, близко.
	Вот, потом там вырезанная тулупка - человек вот, лежит, прав вот так <b>рука</b> .

201008e\_mgd

[Слушать в одном файле \(mp3\)](#)  
First transcription: 0:0:0 Last transcription: 0:18:10

Interviewer: <>!

mgd: <Рисовала>... Русалка!

Interviewer: Ага! А масленницу делали у вас?

mgd: А?

Interviewer: А на масленницу не было?

mgd: О, а как же! И масленница бывает! Тогда, бывало, катались с горы.

Interviewer: С горы?

mgd: Бывало-ка воровали лошадиные сани, это сейчас лошадей-то уж нету! а тогда были лошадиные сани, унесут где-нибудь, и с горы катаются до самой реки. Вон сядут вот оттуда - пошёл!

Interviewer: Ага!

mgd: До самой реки едут!

Interviewer: На санях на этих, да?

mgd: Да, на этих санях катались!

Interviewer: Ага! А саму масленницу как-то тоже изображали куклой такой, нет?

mgd: Да и - как изображали?!



Malinino Corpus

Corpus ▾ Statistics ▾ About ▾ Help | rus en |

Corpus of the Russian dialect spoken in the village Malinino

Collection of Spoken Texts and Recordings

Search in the corpus...

Search

## About the project

The Malinino dialect corpus contains texts recorded in the village Malinino (Khlevenskiy rayon, Lipetsk Oblast) in July-August of 2010. The dialect is part of the interzonal group B, of the South Russian dialect group (following the classification of K.F. Zakharova and V.G. Orlova). It combines

# План доклада

Лингвистика: мифы и реальность

Обо мне

Корпусная лингвистика

Использование причастий и деепричастий в устной речи

## Цель работы

- работа по извлечению примеров и разметке примеров была сделана нашим постдоком Светланой Сергеевной Земичевой
- исследовать использование причастий и деепричастий в устной речи на материале диалектных и билингвальных корпусов
- исследовать соотношение литературных и диалектных форм (например, *евши*)
  - диалектные формы на *-ши* могут иметь разные функции, иногда они ведут себя как финитные формы:

*Пока делаешь, придёшь — он уже вставши.* (Макеево, 1953, f)

Спасибо за внимание!