

# lingtypology: R пакет для лингвистического картографирования

Г. Мороз

Лаборатория языковой конвергенции, НИУ ВШЭ

18 декабря 2018 г.

Открытые лекции — «Городские данные»  
Софт Культура и Инфокультура

#ТЫЖЛИНГВИСТ



# #ТЫЖЛИНГВИСТ

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

# #ТЫЖЛИНГВИСТ

- 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

- прескриптивная



- прескриптивная
- вся остальная
  - исследования грамматики языка и языкового разнообразия
  - исследования распределения грамматических особенностей в языках мира
  - исследования когнитивных способностей человека и других животных, связанных с языком
  - исследования в области NLP и их приложения
  - исследования в области синтеза и распознавания речи и языка
  - создание компьютерных инструментов для решения самых разных задач

Еще бывает *компьютерная лингвистика*, но это обобщенный термин, которым объединяют совсем несвязанные области:

- вспомогательные инструменты лингвистического исследования и документации
- Computational linguistics
- NLP

# лингвистические базы данных



# Лингвистические базы данных

Корпуса — базы данных языкового материала

- корпус литературных текстов
- Русский национальный корпус
- аудио и видео корпуса

# Лингвистические базы данных

## Корпуса — базы данных языкового материала

- корпус литературных текстов
- Русский национальный корпус
- аудио и видео корпуса
  - настоящая речь, а не тексты
  - см., например, [корпус русского жестового языка](#)
  - см., например, [Уошо](#)

## Базы данных — базы данных языковых структур

- The World Atlas of Language Structures (WALS)
- World Atlas of Varieties of English (eWAVE)
- [Glottolog](#) — comprehensive reference information for the world's languages
- ... и другие