

Кто есть кто в лингвистике?

Проект по каталогизации лингвистических (и не только) журналов

Г. А. Мороз

17 марта 2022

# Outline of the talk

Введение

Работа с журналами

Работа со статьями

Репозиторий

- Государство и университеты ставят своей целью финансировать науку, однако напрямую это сделать нельзя, так что приходится финансировать ученых. Так как результат деятельности ученых далеко не всегда очевиден, то приходится придумывать какие-то меры эффективности ученых.

- Государство и университеты ставят своей целью финансировать науку, однако напрямую это сделать нельзя, так что приходится финансировать ученых. Так как результат деятельности ученых далеко не всегда очевиден, то приходится придумывать какие-то меры эффективности ученых.
- Оценивать принято через публикации, но тут обнаруживается, что какую бы меру не придумали, всегда находятся хакеры, которые ломают систему.

- Государство и университеты ставят своей целью финансировать науку, однако напрямую это сделать нельзя, так что приходится финансировать ученых. Так как результат деятельности ученых далеко не всегда очевиден, то приходится придумывать какие-то меры эффективности ученых.
- Оценивать принято через публикации, но тут обнаруживается, что какую бы меру не придумали, всегда находятся хакеры, которые ломают систему.
- Появляется направление под названием *наукометрия* (scientometrics), которое аккумулирует в себе анализ данных в области цитирований, журналов и науки (ср. с другими частными направлениями: биоинформатика, дата журналистика и др.)

## Чем занимаются наукометры

В основном они анализируют метаданные научных публикаций:

- авторов
- название, аннотации, ключевые слова
- список литературы (исходящие цитирование)
- где процитирована данная работа (входящие цитирования)
- где опубликовано (импакт фактор, квартили и другие характеристики журнала и т. п.)

В Вышке они еще и влияют на составление списка хороших и плохих журналов, за которые платят надбавки.

По лингвистике они выдали список из 2662 журналов, которые категорезированы по группам  $a > b > c > d$ .

Лингвистика входит в тег ФИЛОЛОГИЯ, ЛИНГВИСТИКА И МЕДИАКОММУНИКАЦИИ, так что в списке много мусора.

# Outline of the talk

Введение

**Работа с журналами**

Работа со статьями

Репозиторий

- руководители
  - Г. Мороз
  - Б. Орехов
- студенты
  - А. Алексеева (б)
  - Т. Дедов (б)
  - А. Орехов (м)
  - К. Сидоров (м)
  - А. Степанова (б)

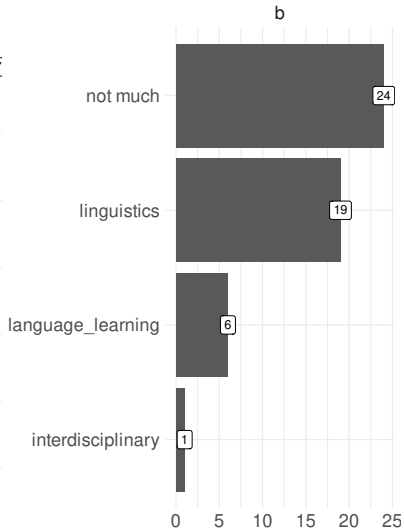
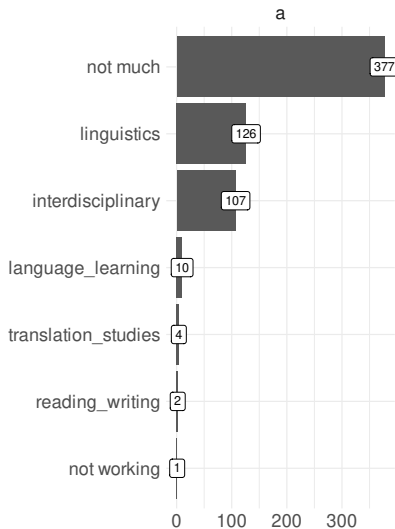


# Что мы сделали

Мы начали работать со списками a и b:

- Таблица с журналами содержит поля
  - title: AAC: AUGMENTATIVE AND ALTERNATIVE COMMUNICATION
  - issn: 0743-4618; 1477-3848
  - hse\_level: a
  - category: биология, медицина и здравоохранение; филология, лингвистика и медиакоммуникации
- Мы добавили поля:
  - only\_PhLiMed: 0
  - url: <https://www.tandfonline.com/loi/iaac20>
  - (ручная разметка) linguistics: not much
  - (ручная разметка) literary\_studies: not much
  - (ручная разметка) aims\_scope: ...
  - url\_domain: tandfonline.com
  - служебные мета-данные: finished, day, month, year

# Разметка поля linguistics



# Outline of the talk

Введение

Работа с журналами

**Работа со статьями**

Репозиторий

# Библиометрические базы

- Web of Science
- Scopus
- eLIBRARY.ru (РИНЦ)
- Google Scholar
- `crossref`, см. R пакет `rcrossref`
- `openalex`, см. R пакет `openalexR`
- Semantic Scholar
- The Lens

При помощи rcrossref мы извлекли для всех лингвистических журналов следующие поля:

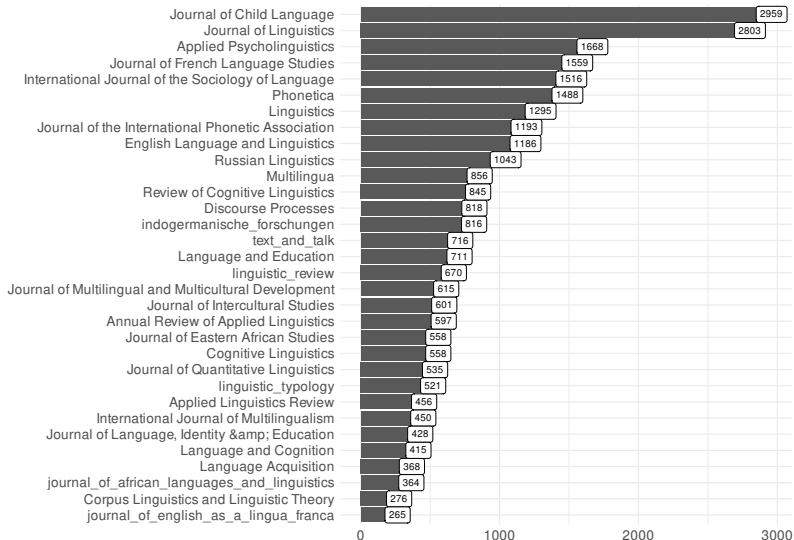
- doi: 10.1515/ang-2012-0005
- author: Müller Patrick
- title: “The impediment that cannot say its name”: Stammering and Trauma in Selected American and British Texts
- url: <http://dx.doi.org/10.1515/ang-2012-0005>
- volume: 130
- issue: 1
- page: -
- reference.count: 0
- container.title: Anglia - Zeitschrift für englische Philologie
- issn: 1865-8938, 0340-5222

# Наконец-то краулер!

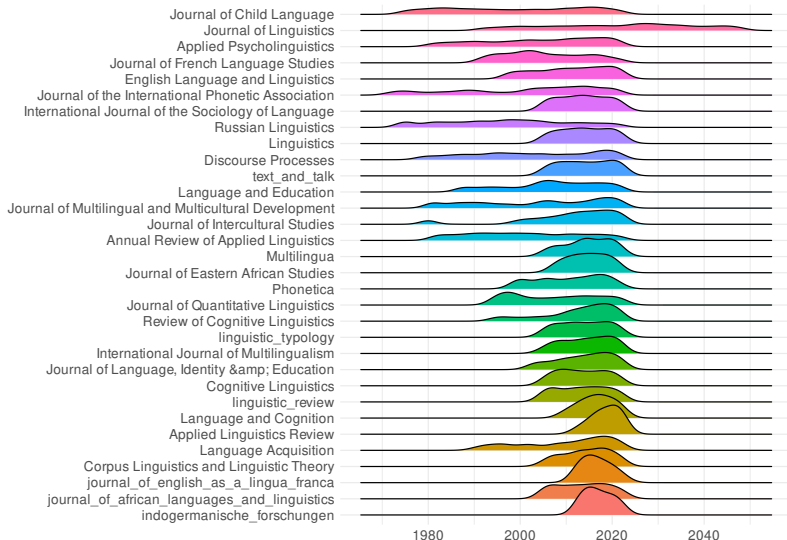
Мы разделили издательства между ребятами и они написали краулеры, чтобы достать следующие поля:

- abstract
- keyword
- year
- source

# Что уже сделано (32 журнала):



# Покрытие по годам





# Outline of the talk

Введение

Работа с журналами

Работа со статьями

Репозиторий

[https://github.com/linguistic-journals/journal\\_list\\_data](https://github.com/linguistic-journals/journal_list_data)

|                            |                                     |
|----------------------------|-------------------------------------|
| — crossref                 | # папка с экспортом из crossref     |
| — with_abstracts           | # папка со сделанными журналами     |
| — scripts                  | # папка со скриптами                |
| — abcd_list_annotated.xlsx | # список журналов                   |
| — helsenki.csv             | # ранкинг журналов из Хелсинской бд |

Спасибо за внимание!