

Опыт зеркальной лаборатории

Г. А. Мороз

Международная лаборатория языковой конвергенции (НИУ ВШЭ,
Москва)

14 декабря 2023

План презентации

О международной лаборатории языковой конвергенции

Что делать с корпусами текстов?

Опыт зеркальной лаборатории с Южным Федеральным Университетом

Потенциал Нижневартовского государственного университета

Международная лаборатория языковой конвергенции

- Открыта в 2017 году



Н. Р. Добрушина и Дж. Николс

Оба исследователя специализируются на славянских языках и языках Кавказа, а также лингвистической типологии

Миссия

Исследование механизмов конвергентных процессов в истории языка, то есть языковых ситуаций, при которых контакт между носителями разных языков ведет к появлению у этих языков общих черт. В лаборатории разрабатываются инструменты для выявления результатов таких процессов по данным электронных корпусов устной речи и создаются каталоги таких явлений на материале малых языков России.

Ресурсы международной лаборатории языковой конвергенции

- lingconlab.ru
- 22 устных корпуса диалектов русского языка
- 8 устных корпусов русского языка билингвов
- 12 корпусов малых языков
- другие ресурсы
 - словари (мегебский, рутульский, тукитинский, хваршинский, даргинский)
 - Типологический атлас языков Дагестана
 - Атлас многоязычия в Дагестане
 - Атлас диалектов рутульского языка
 - Корпус просодии русских диалектов (ПРУД)
 - ...

22 устных диалектных корпуса

Корпус говора Хиславичского района
260,793 ток.

Корпус говора села Спиридонова Буда
70,565 ток.

Корпус говора села Кеба
54,535 ток.

Корпус говора села Церковное
39,469 ток.

Корпус устьянских говоров
959,782 ток.

Корпус говора верхней Пинеги и Виши
70,803 ток.

Корпус говора среднего течения Северной Двины
68,010 ток.

Корпус говора среднего течения Кольмы

Корпус донских говоров
71,600 ток.

Корпус опочечих говоров
68,741 ток.

Лужниковский корпус
68,666 ток.

Корпус говора города Эманжелино
68,324 ток.

Корпус говора деревни Веегора
91,514 ток.

Корпус говора деревни Нехочи
88,965 ток.

Корпус говора Средней Песьи
79,566 ток.

Корпус говора Мантуровского района Костромской области
113,637 ток.

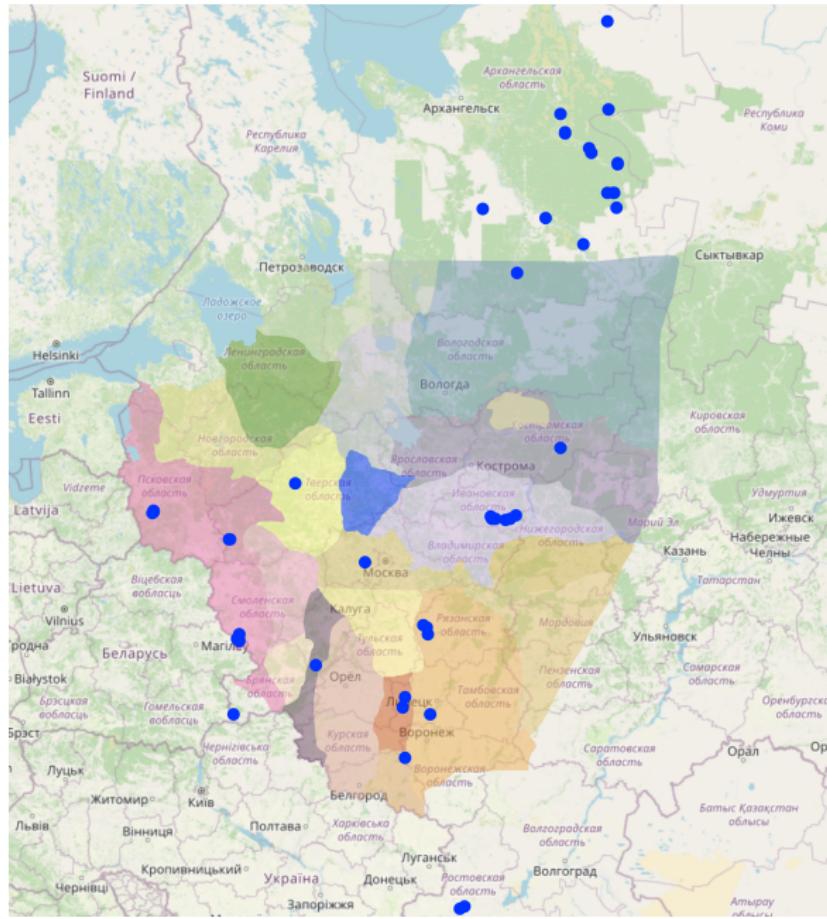
Корпус говора села Роговатка
100,047 ток.

Корпус говора деревни Шаптаково и Манаково
95,335 ток.

Корпус говоров низовья рек Лух и Теза
146,350 ток.

Корпус говора села Малинино
138,943 ток.

22 устных диалектных корпуса



8 устных билингвальных корпусов

Корпус дагестанского русского
376,717 ток.

Якутско-русский корпус переключения кода
15,139 ток.

Корпус русской речи Чувашии
46,307 ток.

Корпус цыганского русского
41,767 ток.

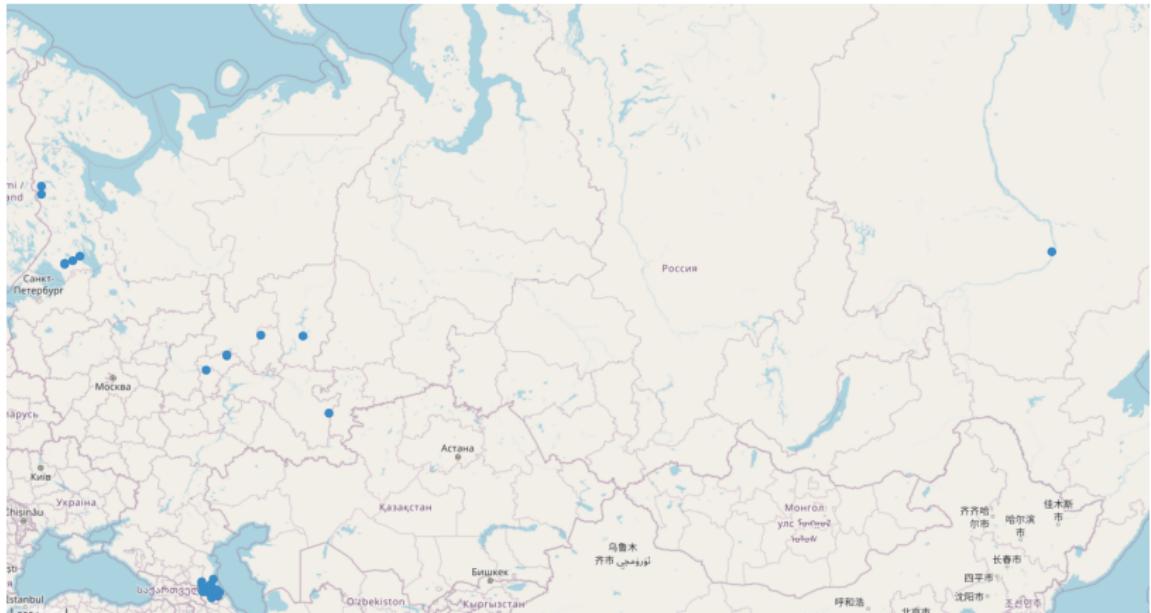
Корпус русской речи Карелии
578,646 ток.

Корпус русской речи Республики Марий Эл
69,109 ток.

Корпус русской речи Башкирии
93,127 ток.

Корпус русской речи бесермян
97,216 ток.

8 устных билингвальных корпусов



12 корпусов малых языков

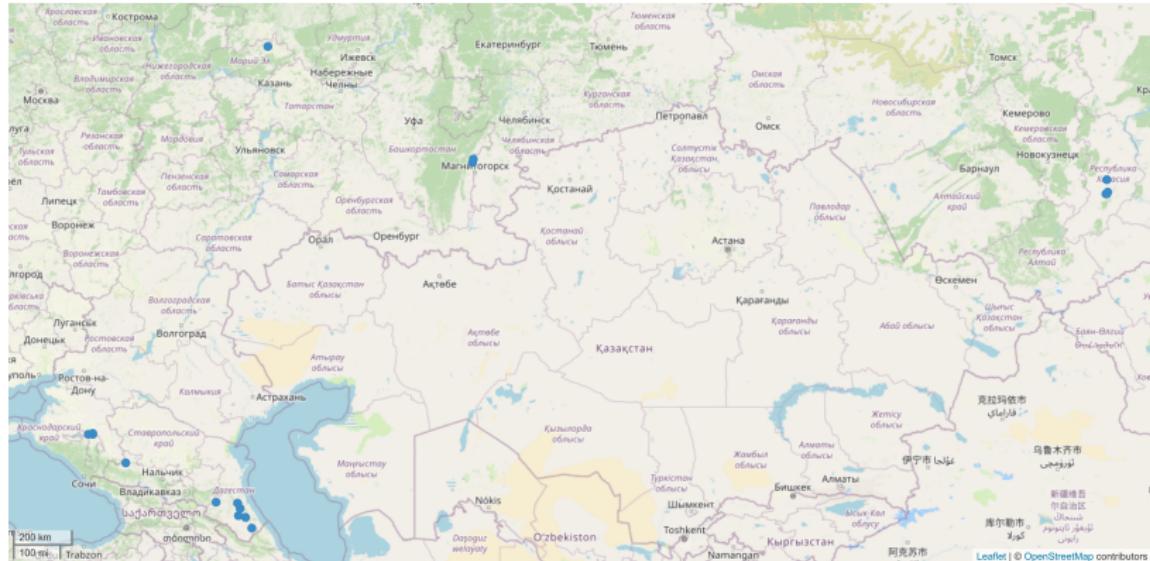
Корпус литературного даргинского
703,988 ток.

Устный корпус кадарского даргинско
12,654 ток.

Устный корпус башкирского языка
25,000 ток.

Устный корпус диалекта хакасского языка
58,000 ток.

12 корпусов малых языков



План презентации

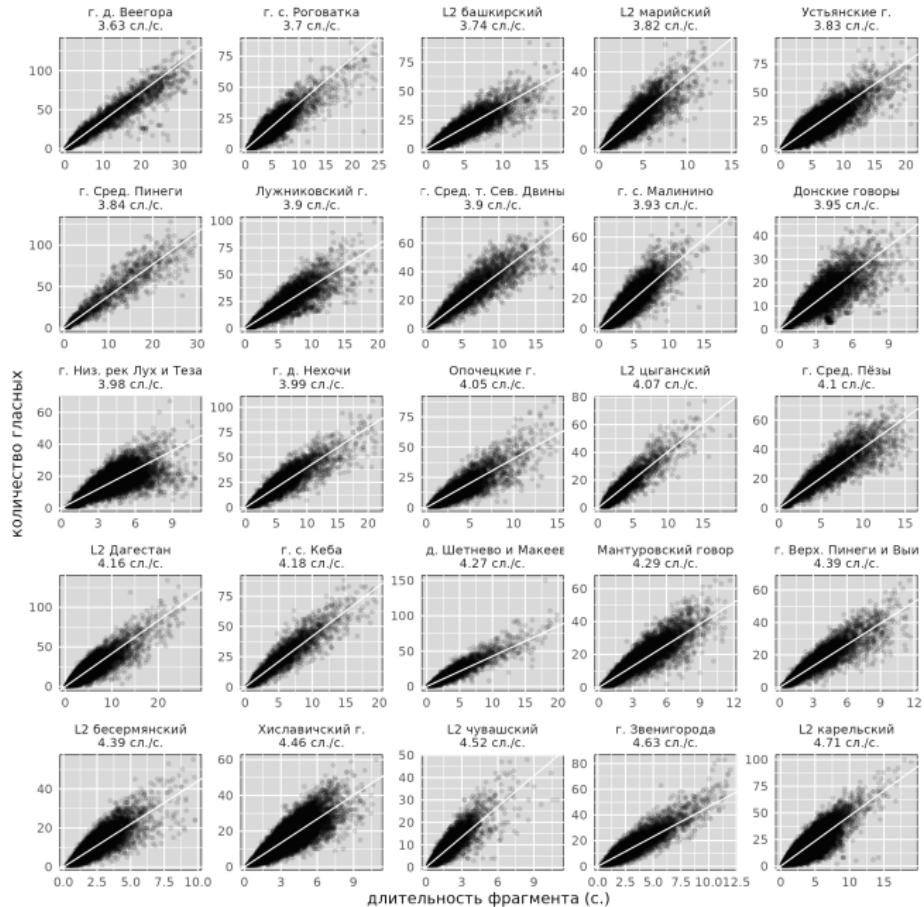
О международной лаборатории языковой конвергенции

Что делать с корпусами текстов?

Опыт зеркальной лаборатории с Южным Федеральным
Университетом

Потенциал Нижневартовского государственного университета

Скорость речи (Мороз 2023: 382)



Корпус просодии русских диалектов (Князев et al. 2024)

PRuD главная о проекте информация о говорах ритмическая структура карта данные



intonation_type	transcription	transcription2	image	speaker	idiom	village	district	region	со
All	со м	All	All	All	All	All	All	All	All
Незавершённость	Дети были со мною маленькие...	H*!H*!H*H%	🕒	ЕАЖ1935	Говор_2	Роговатое	Старооскольский	Белгородская	
Утверждение	Пошли на солому.	%L L+H* L* L%	🕒	?	Говор_3	Клиновское	Коношский	Архангельская	
Незавершённость; завершённость	Беру вот такую чесовочку льну - несу домой.	L+H* H% %L L+H* L* L%	🕒	?	Говор_3	Клиновское	Коношский	Архангельская	
Незавершённость	Само по себе дерево посохло...	L*+H H- H%	🕒	МАН1910	Говор_51	Мосеево	Мезенский	Архангельская	
Утверждение	Сортируем лес.	L* L%	🕒	РИИ1915	Говор_55	Веегора	Пинежский	Архангельская	
Утверждение, взкий фокус	Время семь часов!	H* L%	🕒	РУК	Говор_55	Веегора	Пинежский	Архангельская	

Исследование вариативности



С. С. Земичева, Г. А. Мороз, К. Наккарато, А. В. Яковлева

Грант РНФ (24-28-01097) “Исследование вариативности
билингвального и диалектного русского языка на материале
устных корпусов”

План презентации

О международной лаборатории языковой конвергенции

Что делать с корпусами текстов?

Опыт зеркальной лаборатории с Южным Федеральным
Университетом

Потенциал Нижневартовского государственного университета

Зеркальная лаборатория с ЮФУ

- проходила 2020-2023 под руководством Елены Михайловны Севериной



- мы надеемся, что сможем продолжить сотрудничать с коллегами в 2024-2026

В рамках зеркальной лаборатории велась работа по трем направлениям:

- Проект **Digital Chekhov**: создание многоуровневой (семантической) разметки собрания произведений А. П. Чехова
- Устный корпус донских диалектов: аннотация аудиозаписей и создание интерфейса устного корпуса
- Русский учебный корпус: образцы устной и письменной речи нестандартных говорящих на русском языке, изучающих русский язык как иностранный

Digital Chekhov

- <https://chekhov-digital.sfedu.ru/>

The screenshot shows the homepage of the Digital Chekhov project. At the top, there is a navigation bar with links to 'Главная' (Home), 'Поиск' (Search), 'Произведения' (Works), and 'О нас' (About us). Below the navigation bar is a large, ornate illustration of a woman's face, likely a portrait of Anton Chekhov's mother. Overlaid on this image is a search interface. The main title 'Расширенный поиск' (Advanced search) is displayed above three search buttons: 'по произведению' (by work), 'по сущностям' (by entities), and 'по указателям' (by indicators). Below these buttons is a search input field with the placeholder 'Введите запрос' (Enter query) and a magnifying glass icon. Underneath the search input is a sub-instruction 'Найти любой отрывок произведения по параметрам' (Find any excerpt from a work by parameters). Below this are six search parameters arranged in two rows of three: 'Год' (Year), 'Том' (Volume), 'Виды текста' (Text types), 'Издание' (Edition), 'Место написания/издания' (Place of writing/edition), 'Тип текста' (Text type), 'По объему' (By volume), and 'По языкам' (By languages). At the bottom of the page, there is a footer note: 'Проект Chekhov Digital разрабатывается Центром цифровых гуманитарных исследований ИФЖиМКК ЮФУ совместно с Лабораторией филологии ОГИ ЮНЦ РАН и Центром цифровых гуманитарных исследований НИУ ВШЭ'. To the right of the footer are social media icons for YouTube, VK, and Twitter, along with a blue circular icon containing an '@' symbol.

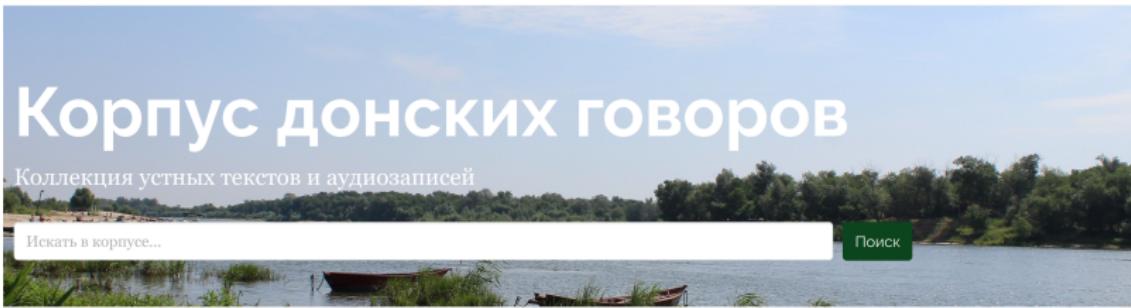
Устный корпус донских диалектов

- http://lingconlab.ru/don_rnd

Корпус донских говоров

Корпус ▾ Статистика ▾ О проекте ▾ Помощь

[rus en]



О проекте

Записи были выполнены в 2021–2022 году в ст. Раздорская (2021 год) и ст. Мелиховская (2022 год) Ростовской области.

Сбор и расшифровка текстов осуществлена при поддержке ИФЖиМКК ЮФУ в рамках проекта «Конвергенция языковых пластов русского языка в зеркале цифровых решений» в рамках соглашения о научном сотрудничестве («Зеркальные лаборатории») между НИУ ВШЭ и ЮФУ № 6.13.1-02/250821-1 от 25.08.2021.

Исключительное право на базу данных аудиофайлов принадлежит ФГАОУ ВО "Южный федеральный университет".

Больше информации о корпусе см. [здесь](#).

Устный корпус донских диалектов

- http://lingconlab.ru/don_rnd

Корпус донских говоров

Корпус ▾ Статистика ▾ О проекте ▾ Помощь

[rus en]

Query: *fword='руки'%"c*

Number of results: 5

Recording	Расшифровка текстов
0:00 / 0:00	Да , становится на место руки , и ноги , и все .
0:00 / 0:00	Всё у меня атрофировалось , ни руки , ни ноги , всё . Так поднимали , как Ваньку - встаньку , и опять клади .
0:00 / 0:00	Его рука , твоя рука , моя , все , ложите , ложите все руки .
0:00 / 0:00	Ну , вот на моих детях и хорошо , ни одной . . . ни переломов ноги , ни руки , ничего не было .

Русский учебный корпус

- <http://www.web-corpora.net/RLC>

RLC Главная Поиск Новости Помощь Статистика

ru ✓

RLC

Русский учебный корпус

Русский язык в многоязычном мире
Associated Event of the April HSE International Conference

Подробнее



Что такое Русский учебный корпус?

Метаданные

Партнёры

Языки

Поиск

Что такое Русский учебный корпус?

В Русском учебном корпусе содержатся образцы устной и письменной речи двух категорий нестандартных говорящих на русском языке: изучающих русский язык как иностранный и так называемых эркотажных говорящих. Для первой категории русский язык не является родным, представители же второй категории начали усваивать его в детстве как первый язык, но по разным причинам (в основном, это эмиграция) в качестве основного языка общения используют другой язык. Корпус позволяет производить поиск по лексико-грамматическим свойствам, а также по самым разным типам отклонений от стандартной русской речи - от орфографических ошибок до выбора лексических единиц и конструкций. Лингвистический анализ и разметка осуществляются членами Лаборатории по корпусным исследованиям НИУ ВШЭ под руководством Екатерины Рахилиной.

Создание тренажеров

- https://agricolamz.github.io/language_of_science_website/

Академический русский: экономика

[главная](#) [портрет слова](#) [упражнения](#) [о проекте](#)

рынок



▼ парадигма

падеж ед. ч. мн. ч.

Им.	рынок	рынки
Р.	рынка	рынков
Д.	рынку	рынкам
В.	рынок	рынки
Тв.	рынком	рынками
Пр.	рынке	рынках

[примеры](#)

[упражнения](#)

Монополия может существовать лишь постольку, поскольку вход на _____ (рынок) представляется другим предприятиям невыгодным или невозможным.

[проверить](#)

Нажмите, чтобы посмотреть ответ

Взаимодействие

- активный обмен опытом
- лекции сотрудников НИУ ВШЭ для студентов ЮФУ
- совместная конференция по цифровым гуманитарным исследованиям в 2023 году

План презентации

О международной лаборатории языковой конвергенции

Что делать с корпусами текстов?

Опыт зеркальной лаборатории с Южным Федеральным
Университетом

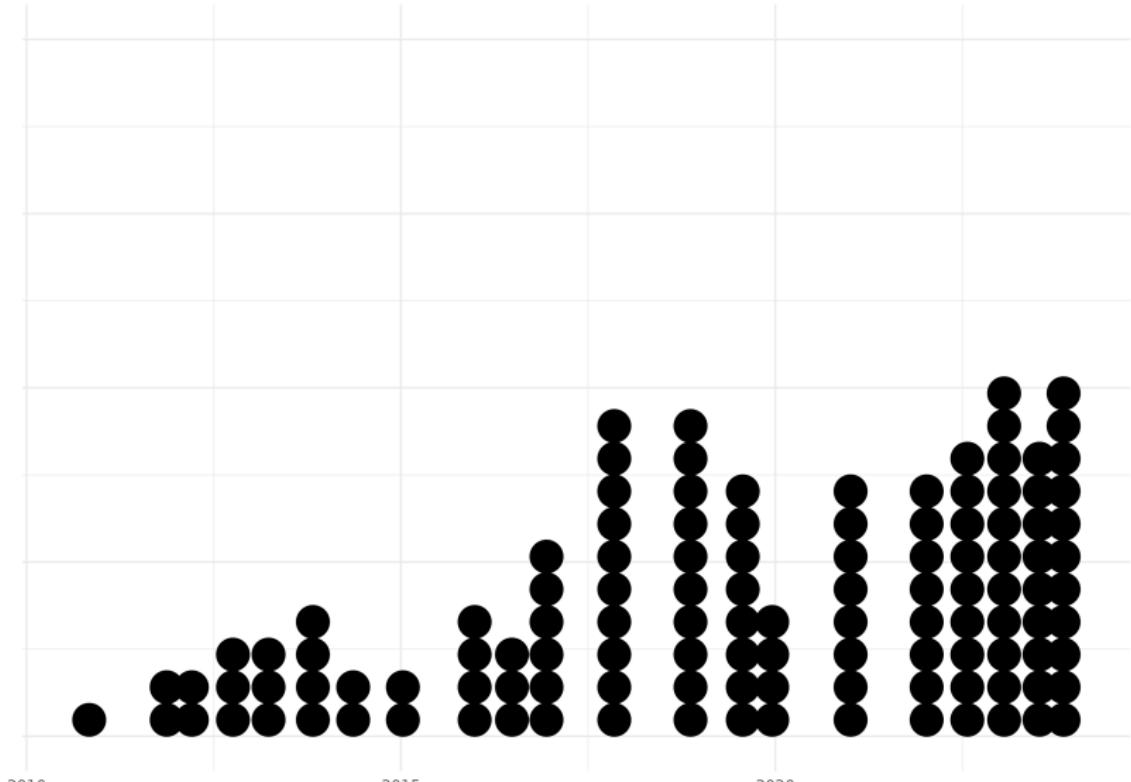
Потенциал Нижневартовского государственного университета

Изучение хантыйского языка

- профиль магистратуры «Хантыйская филология»
- профиль магистратуры «Искусственный интеллект в моделировании речевой деятельности»
- Научно-исследовательские лаборатория этнофилологии народов Севера
- Научно-исследовательские лаборатория лингвопедагогики и проблемного обучения

Изучение хантыйского языка

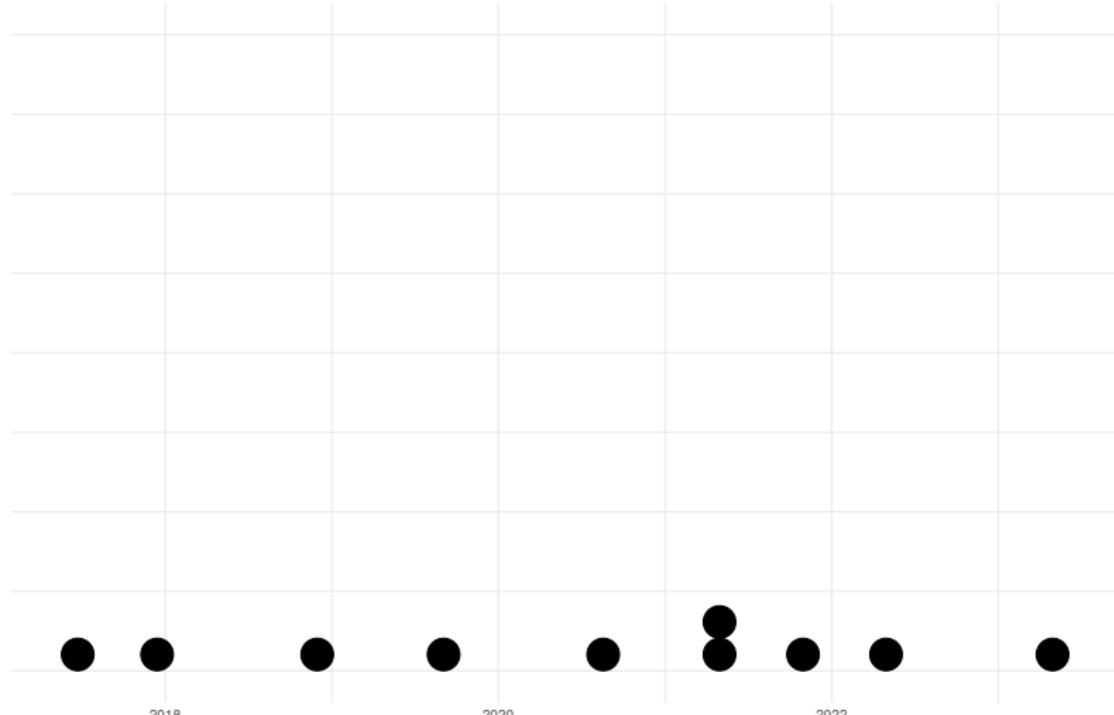
Упоминание хантыйского языка и культуры в новостях НВГУ



новости взяты с сайта <https://nvsu.ru>
в промежуток с 2009 по сегодняшний день

Изучение хантыйского языка

Упоминание хантыйского языка и культуры в Нижневартовский Филологический Вестнике



в промежуток с 2016 по сегодняшний день

Хантыйские корпуса

- Ob-Ugric database, Германия, Университет Людвига-Максимилиана, Мюнхен. Е. К. Скрибник
- <https://wwwbabel.gwi.uni-muenchen.de/>

Хантыйские корпуса

- Ob-Ugric database, Германия, Университет Людвига-Максимилиана, Мюнхен. Е. К. Скрибник
- <https://wwwbabel.gwi.uni-muenchen.de/>
- ЛингвоДок, Россия, ИСП РАН, ИЯз РАН. Норманская Ю. В., Дыбо А. В., Борисенко О.
- http://lingvodoc.ispras.ru/corpora_all?all=&language=508%2C44

Хантыйские корпуса

- Ob-Ugric database, Германия, Университет Людвига-Максимилиана, Мюнхен. Е. К. Скрибник
- <https://wwwbabel.gwi.uni-muenchen.de/>
- ЛингвоДок, Россия, ИСП РАН, ИЯз РАН. Норманская Ю. В., Дыбо А. В., Борисенко О.
- http://lingvodoc.ispras.ru/corpora_all?all=&language=508%2C44
- NorthEuralex, Германия, Университет Тюбингена, Йоханнес Деллерт
- <https://link.springer.com/article/10.1007/s10579-019-09480-6>

Хантыйские корпуса

- Ob-Ugric database, Германия, Университет Людвига-Максимилиана, Мюнхен. Е. К. Скрибник
- <https://wwwbabel.gwi.uni-muenchen.de/>
- ЛингвоДок, Россия, ИСП РАН, ИЯз РАН. Норманская Ю. В., Дыбо А. В., Борисенко О.
- http://lingvodoc.ispras.ru/corpora_all?all=&language=508%2C44
- NorthEuralex, Германия, Университет Тюбингена, Йоханнес Деллерт
- <https://link.springer.com/article/10.1007/s10579-019-09480-6>
- UraTyp – типологический ресурс от института Макса Планка
- <https://uralic.clld.org/languages>

Хантыйские корпуса

- устные диалогические корпуса
- акцент именно на сургутский диалект хантыйского, которым занимаются в НВГУ

Спасибо за внимание!

Литература

С. В. Князев, Г. А. Мороз, and Дьяченко С. В. Корпус Просодии Русских Диалектов (ПРУД), 2024. URL
<https://LingConLab.github.io/PRuD/>.

Г. А. Мороз. Скорость русского речи на основе билингвальных и диалектных устных корпусов. In Н. А. Коротаев and Н. Р. Сумбатова, editors, *Состав науки: Сборник статей к юбилею Веры Исааковны Подлесской*. Буки Веди, М., 2023.