

Исследование вариативности в рутульском языке для курса «Основные приложения математики», НИУ ВШЭ

Георгий Мороз

Международная лаборатория языковой конвергенции
НИУ ВШЭ

на материале

презентации А. П. Алексеевой, А. Д. Ивановой, Т. А. Майсака,
презентации Н. Р. Добрушиной,

тезисов, написанных М. А. Даниэлем, Г. А. Морозом, Н. Р.
Добрушиной, К. В. Филатовым, А. Д. Ивановой, М. Г. Меленченко
и А. П. Алексеевой

18.10.2024

Мифы о лингвистике

#ТЫЖЛИНГВИСТ

- знает как правильно: писать, употреблять слова/выражения, ...

#ТЫЖЛИНГВИСТ

- знает как правильно: писать, употреблять слова/выражения, ...
- умеет читать на всех письменностях мира

#ТЫЖЛИНГВИСТ

- знает как правильно: писать, употреблять слова/выражения, ...
- умеет читать на всех письменностях мира
- знает все языки на свете

#ТЫЖЛИНГВИСТ

- знает как правильно: писать, употреблять слова/выражения, ...
- умеет читать на всех письменностях мира
- знает все языки на свете
- умеет распознавать каждый язык на слух

#ТЫЖЛИНГВИСТ

- знает как правильно: писать, употреблять слова/выражения, ...
- умеет читать на всех письменностях мира
- знает все языки на свете
- умеет распознавать каждый язык на слух
- может рассказать о происхождении каждого слова

#ТЫЖЛИНГВИСТ

- знает как правильно: писать, употреблять слова/выражения, ...
- умеет читать на всех письменностях мира
- знает все языки на свете
- умеет распознавать каждый язык на слух
- может рассказать о происхождении каждого слова
- не может разбираться в статистике и программировании

#ТЫЖЛИНГВИСТ

- знает как правильно: писать, употреблять слова/выражения, ...
- умеет читать на всех письменностях мира
- знает все языки на свете
- умеет распознавать каждый язык на слух
- может рассказать о происхождении каждого слова
- не может разбираться в статистике и программировании
- все перечисленное выше — чушь

Обо мне

- полевой исследователь (30 поездок, почти все на Кавказ)
- фонетист, фонолог, квантитативный лингвист, занимаюсь лингвистической географией
- преподаю статистику и R (язык программирования)
- написал несколько лингвистических пакетов для R
 - [lingtypology](#)
 - [phonfieldwork](#)
 - [lingglosses](#)

Прескриптивная vs. дескриптивная лингвистика

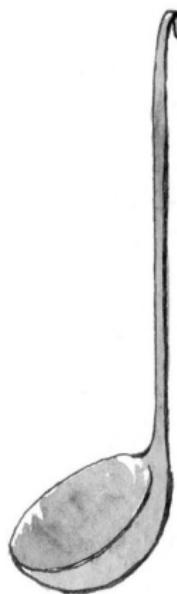
- прескриптивная

Прескриптивная vs. дескриптивная лингвистика

- прескриптивная
- вся остальная (дескриптивная)
 - каталогизация языкового разнообразия, описание языковых контактов
 - исследования и документация грамматики, фонетики и лексики конкретных языков
 - исследования распределения грамматических/фонетических/лексических особенностей в языках мира
 - исследования и документация исторических изменений грамматических/фонетических/лексических особенностей языков
 - исследования когнитивных способностей человека и других животных, связанных с языком (усвоение, потеря языка и др.)
 - языковые аспекты исследования мозга
 - исследования в области синтеза и распознавания речи и языка
 - исследования в области NLP, пробинг языковых моделей и т. п.
 - ...

Прескриптивная vs. дескриптивная лингвистика

Запишите где-нибудь, что изображено на картинке (рис. Т. Пановой).



Прескриптивная vs. дескриптивная лингвистика

Это часть опроса И. Левина 2021 года:

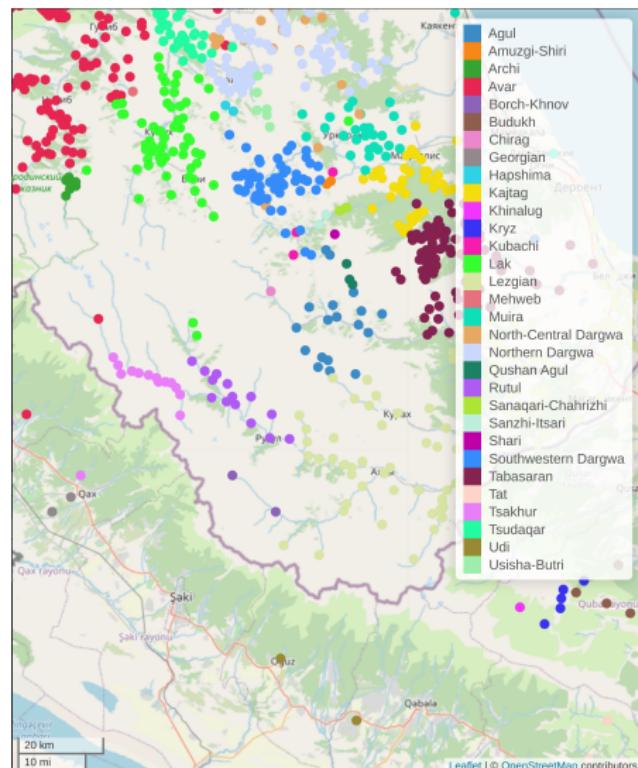


- половник**
- поварёшка**
- черпак**
- ополовник**
- ополонник**

Рутульский язык

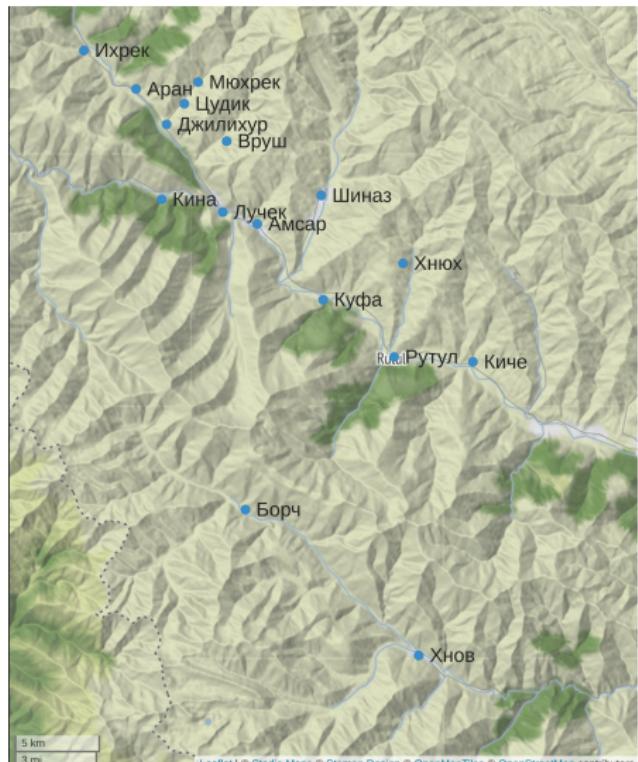
Рутульский язык

- Лезгинская ветвь
нахско-дагестанской семьи
- Южный Дагестан
- Число носителей
 - 33 000 (по данным последней переписи)
 - в селах Рутульского и Ахтынского района Республики Дагестан



Рутульский язык

- Лезгинская ветвь нахско-дагестанской семьи
 - Южный Дагестан
 - Число носителей
 - 33 000 (по данным последней переписи)
 - в селах Рутульского и Ахтынского района Республики Дагестан
 - Еще есть селение Шин в Азербайджане



Традиционное диалектное членение

Пять традиционных диалектов

[Ибрагимов, 2004]:

- ихрекский
- борчинско-хновский
- мюхрекский
- шиназский
- мухадский (литературный язык)

Говоры остальных сел считаются
“смешанными”.



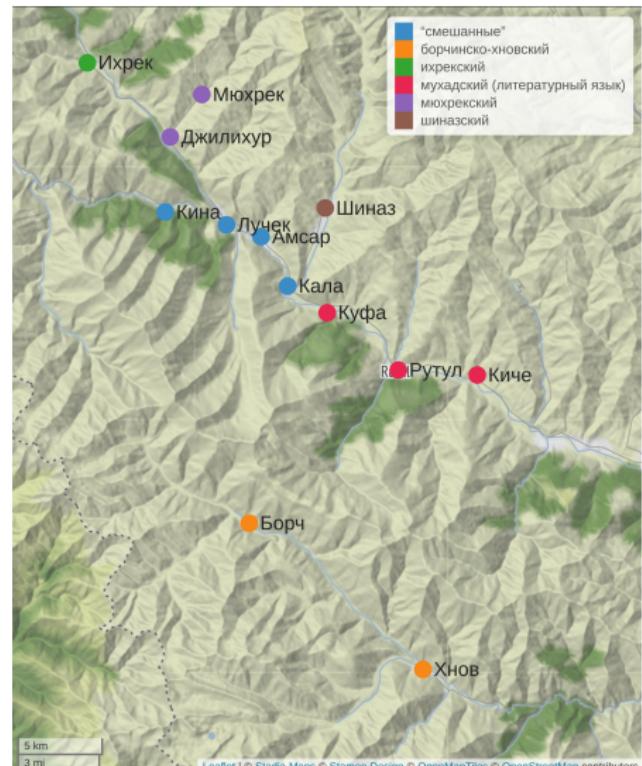
Традиционное диалектное членение

Пять традиционных диалектов

[Ибрагимов, 2004]:

- ихрекский
- борчинско-хновский
- мюхрекский
- шиназский
- мухадский (литературный язык)

Говоры остальных сел считаются
“смешанными”.



Традиционное диалектное членение

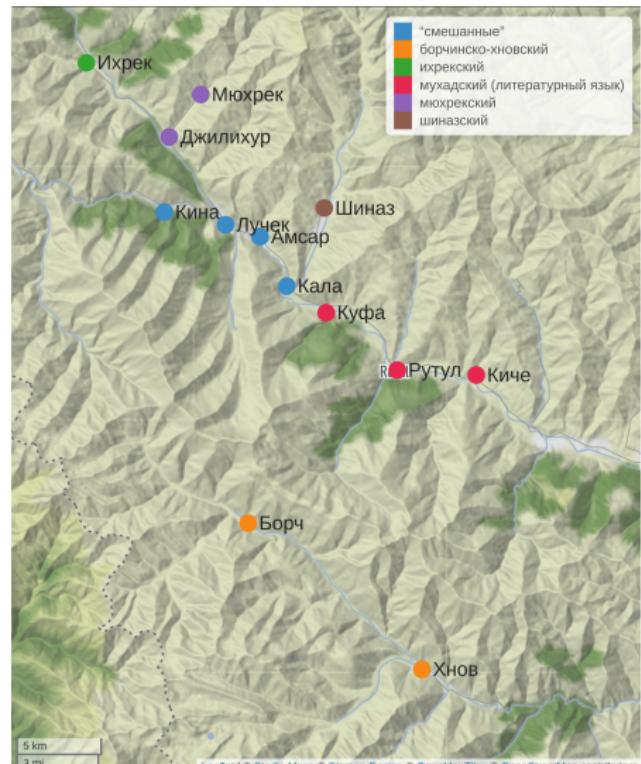
Пять традиционных диалектов

[Ибрагимов, 2004]:

- ихрекский
- борчинско-хновский
- мюхрекский
- шиназский
- мухадский (литературный язык)

Говоры остальных сел считаются
“смешанными”.

Некоторые специалисты считают
борчинско-хновский отдельным
языком.

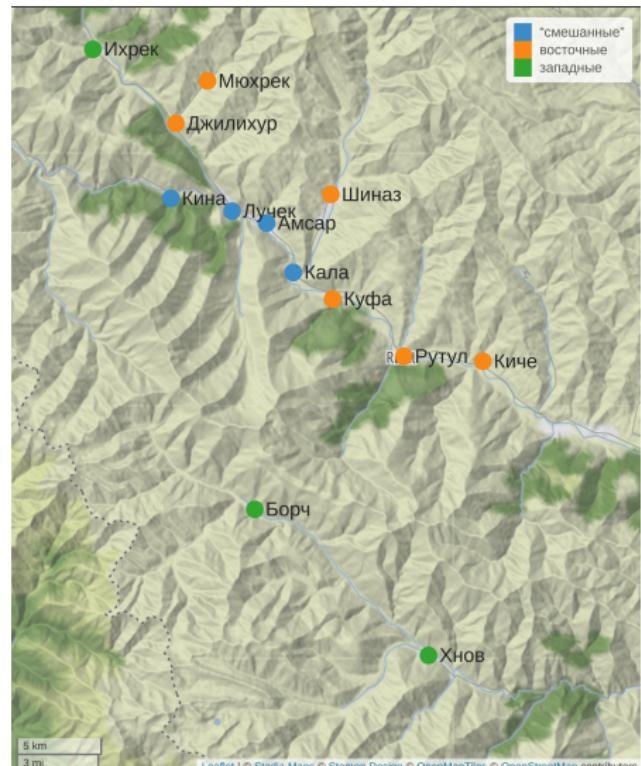


Традиционное диалектное членение

Также в [Ибрагимов, 2004]

выделяются:

- Западная диалектная группа
 - ихрекский
 - борчинско-хновский
- “Смешанные диалекты”
- Восточная диалектная группа
 - мюхрекский
 - шиназский
 - мухадский (основа литературного языка)



Проблемы традиционной классификации

- Импрессионистическая классификация, основанная на качественных данных
Вообще-то надо заметить, что эти наречия похожи больше на изоглоссы, чем на настоящие диалекты. В большинстве случаев диалектные демаркации здесь затрудняются. [Ибрагимов, 2004, 24].
- Непонятно, что значит “смешанные” говоры
 - “смешанные” между западными и восточными?
 - “смешанные” внутри восточных?

Рутульские села на картах

- На карте 1870 обозначены Ихрек,
Лучек, Рутул, Борч, Хнов



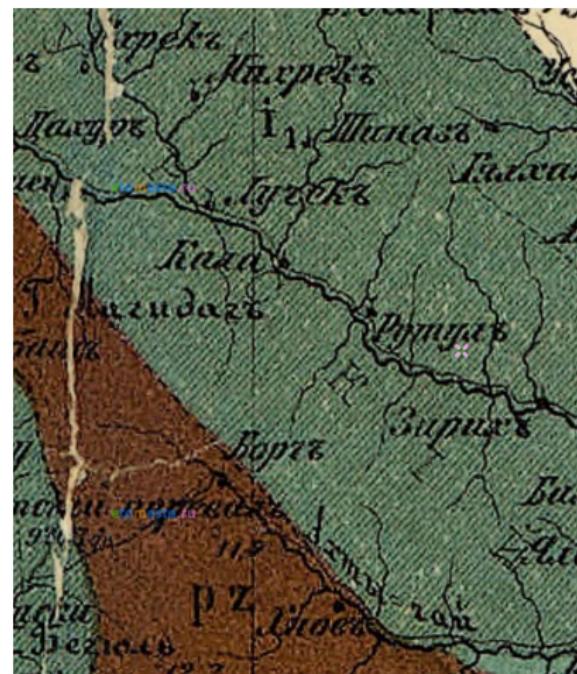
Рутульские села на картах

- На карте 1870 обозначены Ихрек, Лучек, Рутул, Борч, Хнов и Шин
- На карте 1903 обозначены Ихрек, Михрек, Джилихур, Лучек, Кина, Амсар, Шиназ, Кала, Рутул, Борч, Хнов и Шин



Рутульские села на картах

- На карте 1870 обозначены Ихрек, Лучек, Рутул, Борч, Хнов и Шин
- На карте 1903 обозначены Ихрек, Мюхрек, Джилихур, Лучек, Кина, Амсар, Шиназ, Кала, Рутул, Борч, Хнов и Шин
- На карте 1909 обозначены Ихрек, Мюхрек, Лучек, Шиназ, Кала, Рутул, Борч, Хнов



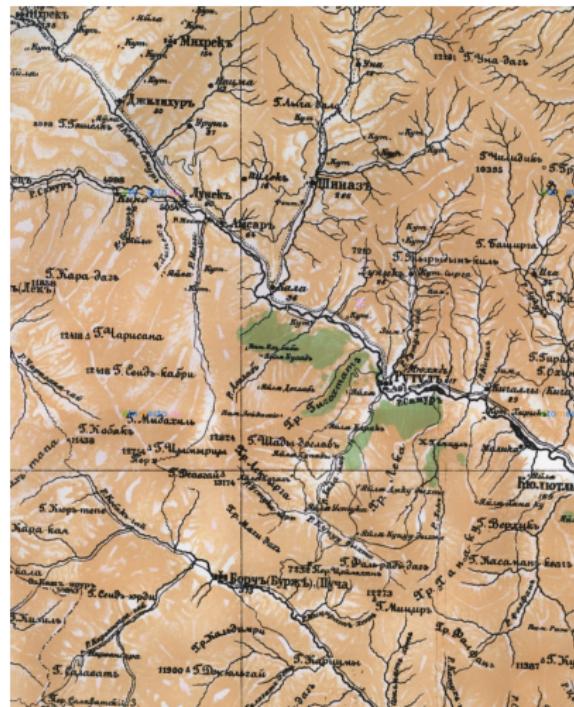
Рутульские села на картах

- На карте 1870 обозначены Ихрек, Лучек, Рутул, Борч, Хнов и Шин
- На карте 1903 обозначены Ихрек, Мюхрек, Джилихур, Лучек, Кина, Амсар, Шиназ, Кала, Рутул, Борч, Хнов и Шин
- На карте 1909 обозначены Ихрек, Мюхрек, Лучек, Шиназ, Кала, Рутул, Борч, Хнов
- На карте 1915 обозначены Ихрек, Амсар, Шиназ, Рутул, Борч, Хнов, Шин



Рутульские села на картах

- На карте 1870 обозначены Ихрек, Лучек, Рутул, Борч, Хнов и Шин
 - На карте 1903 обозначены Ихрек, Мюхрек, Джилихур, Лучек, Кина, Амсар, Шиназ, Кала, Рутул, Борч, Хнов и Шин
 - На карте 1909 обозначены Ихрек, Мюхрек, Лучек, Шиназ, Кала, Рутул, Борч, Хнов
 - На карте 1915 обозначены Ихрек, Амсар, Шиназ, Рутул, Борч, Хнов, Шин
 - На карте 1926 обозначены Ихрек, Мюхрек, Джилихур, Лучек, Кина, Амсар, Шиназ, Кала, Рутул, Борч, Хнов и Шин



Из истории селений

Согласно работе [[Лавров, 1962](#)], есть

старые селения:

- Рутул
- Ихрек
- Лучек
- Мюхрек
- Амсар
- Хнов
- Шиназ

новые селения:

- Киче
- Кина
- Борч
- Джилихур
- Кала
- Куфа
- совсем мелкие кутаны (выселки)
 - Хунюк
 - Уна
 - Вруш
 - Пилек
 - Фартма
 - Нацма
 - Фчек

Рутульская экспедиция

Рутульская экспедиция

С 5–27 июля 2022 года 11 лингвистов объездили 14 рутульских сел.



Рутульская экспедиция

С 5–27 июля 2022 года 11 лингвистов объездили 14 рутульских сел.



Рутульская экспедиция

С 5–27 июля 2022 года 11 лингвистов объездили 14 рутульских сел.



Цели экспедиции

- Описание диалектного варьирования методом элицитации единого набора стимулов во всех селах
 - Описание ранее не засвидетельствованных диалектных вариантов переменных
 - Уточнение традиционной диалектной классификации
 - Количественная оценка степени схожести диалектов

Стимулы для элицитации

- Анкеты по лексике, фонологии, морфологии, синтаксису — около 900 стимулов (слов или предложений) на русском языке
 - Список исследуемых параметров междиалектного варьирования подготовлен на основе информации из описательной и сравнительно-исторической литературы по рутульскому языку ([Дирр, 1912, Ибрагимов, 1978, Махмудова, 2001, 2002, Ибрагимов, 2004, Джамалов and Семедов, 2006, Ибрагимова, 2009, Исмаилова, 2011, Исламов, 2014, Алисултанов and Сулейманова, 2019, Ибрагимова, 2019, Алисултанов and Исламов, 2022] и др.) а также на основе опыта экспедиционеров в описании кининского диалекта

Сбор данных

- Анкеты были распределены между участниками экспедиции, каждый человек собирал одну и ту же анкету во всех сёлах
- Сессии с носителями записывались на диктофон, ответы на стимулы фиксировались в таблице
- Исследователи дальше производили кодирование, полученных значений

параметр	село	стимул	данные	кодирование
Лексема со значением "волк"	Хнов	Волк задрал овцу.	джанавар	djanawar
Лексема со значением "волк"	Ихрек	Волк задрал овцу.	убул	ubul
Лексема со значением "клюв"	Ихрек	клюв	кlyф	k'uf / k'ufl / k'ux ^w
Лексема со значением "клюв"	Лучек	клюв	кlyфъ	k'uf / k'ufl / k'ux ^w
Лексема со значением "клюв"	Шиназ	клюв	кlyф'	k'uf / k'ufl / k'ux ^w
Лексема со значением "клюв"	Кала	клюв	кlyхъв	k'uf / k'ufl / k'ux ^w

Атлас

Атлас диалектов рутульского языка

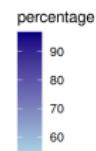
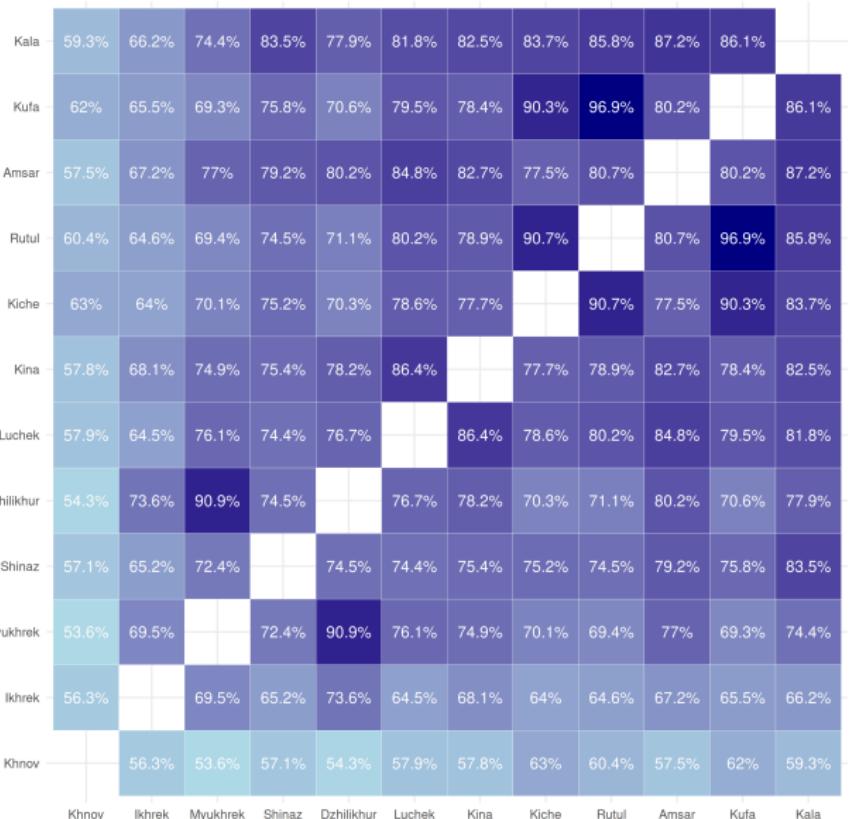
https://lingconlab.github.io/rutul_dialectology/ru/

Лингвистическая близость диалектов

Матрица расстояния

После того, как был закодирован весь материал атласа, мы посчитали, для каждой пары деревень, какая доля форм в данной паре деревень одинакова. Получившуюся матрицу я буду называть матрицей (лингвистического) расстояния.

Матрица расстояния

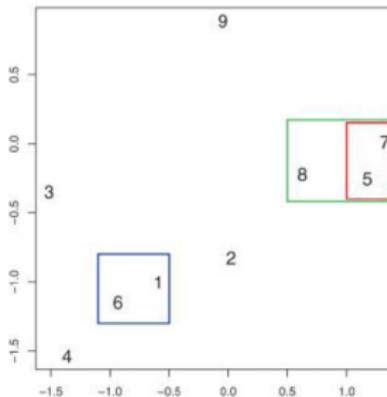
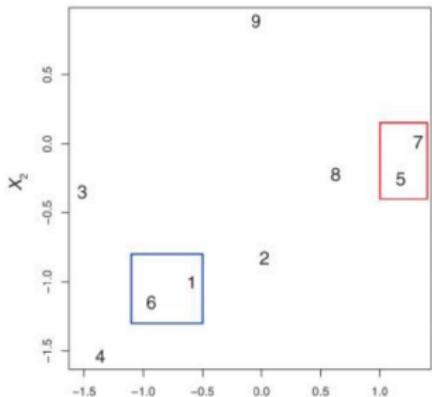
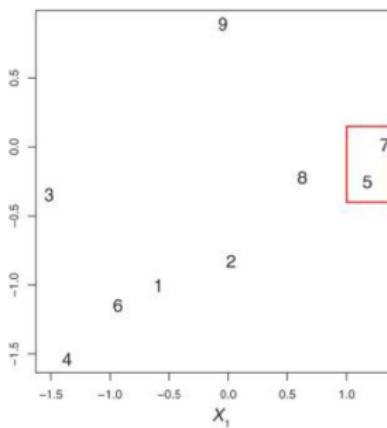
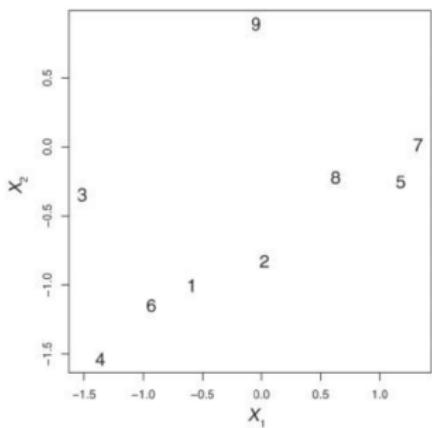


Иерархическая кластеризация

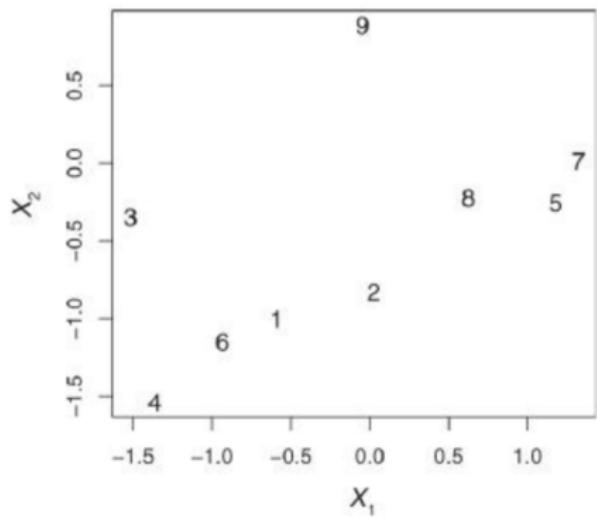
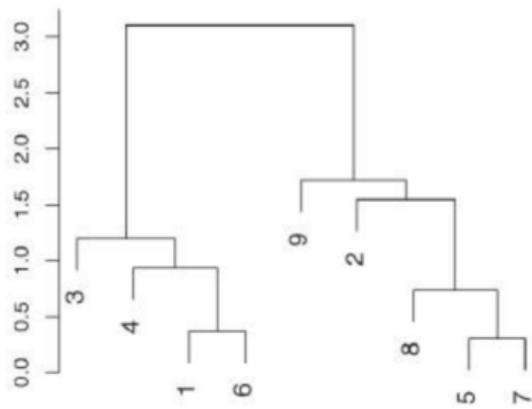
Существует достаточно много методов кластеризации, мы применили иерархическую кластеризацию:

- На первом шаге все наблюдения — отдельный кластер
- Объединяем пару кластеров с наименьшим расстоянием
- Повторяем предыдущие шаги, пока не останется один кластер

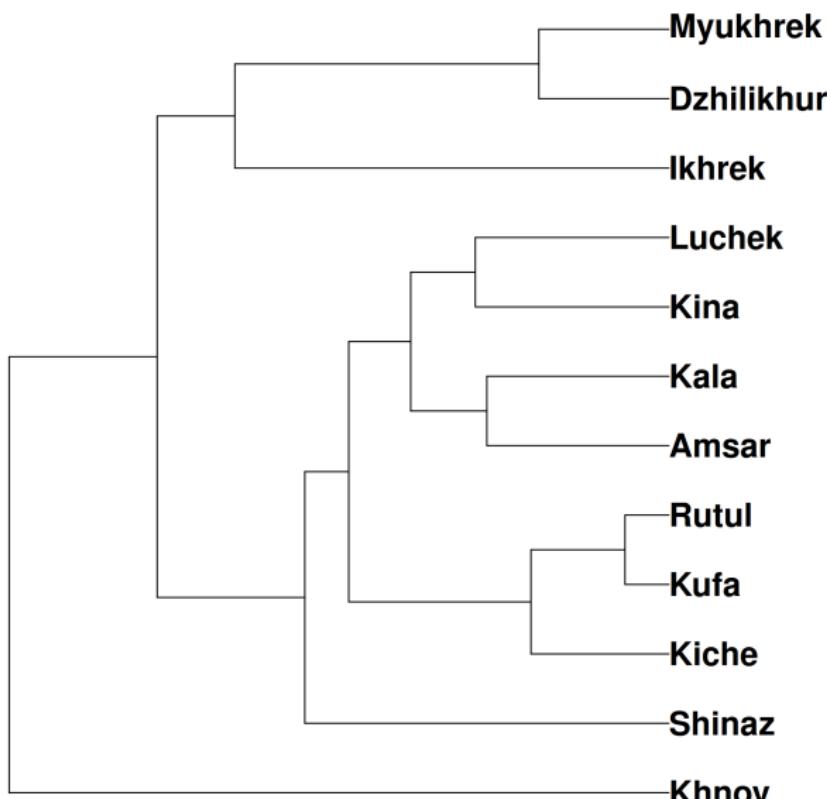
Иерархическая кластеризация: [James et al., 2013, 400]



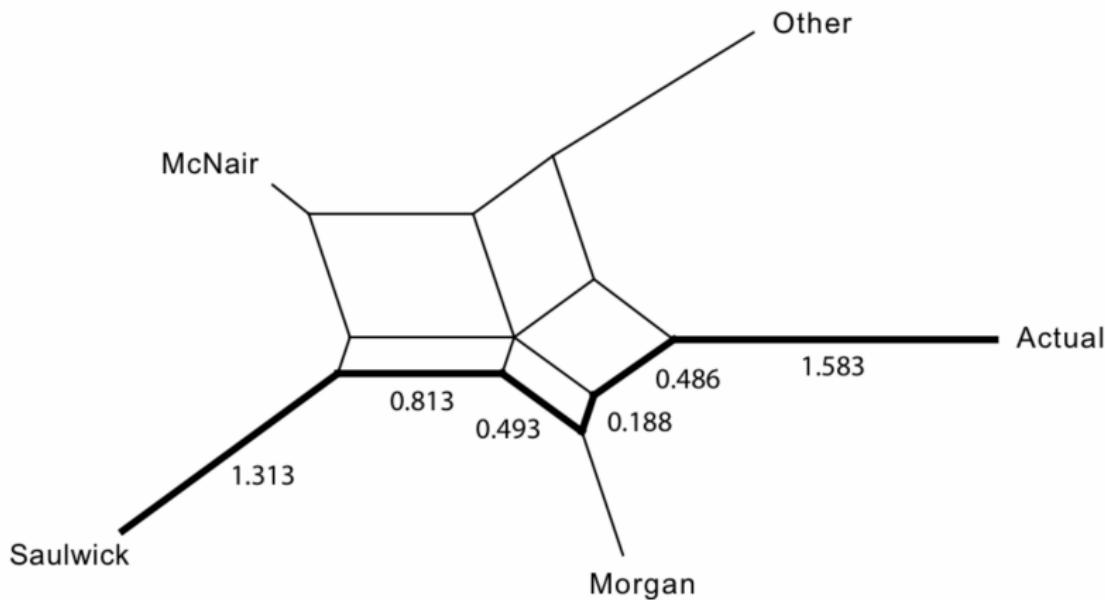
Иерархическая кластеризация: [James et al., 2013, 398]



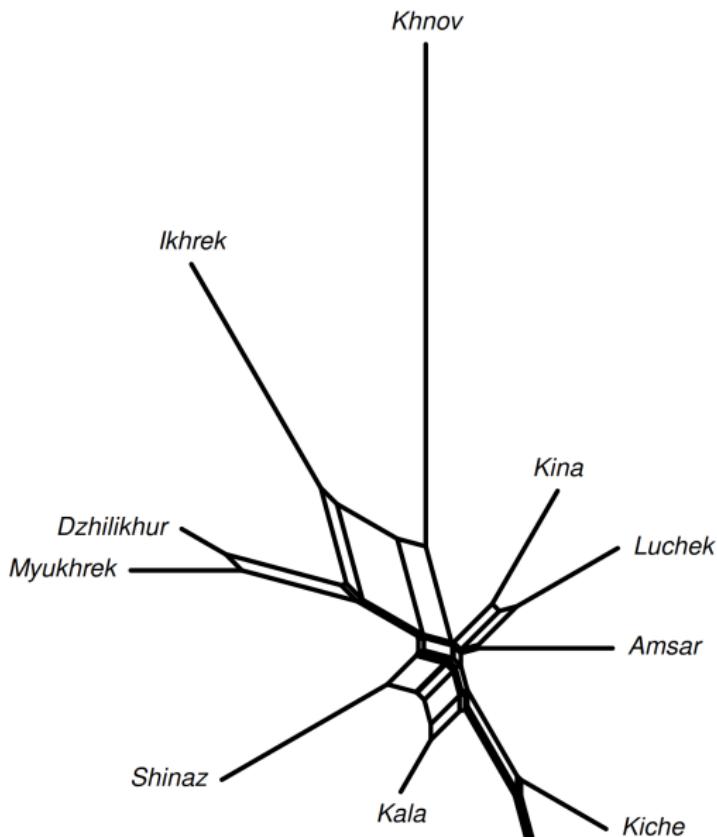
Иерархическая кластеризация: рутульские данные



Нейборнет



Нейборнет: рутульские данные

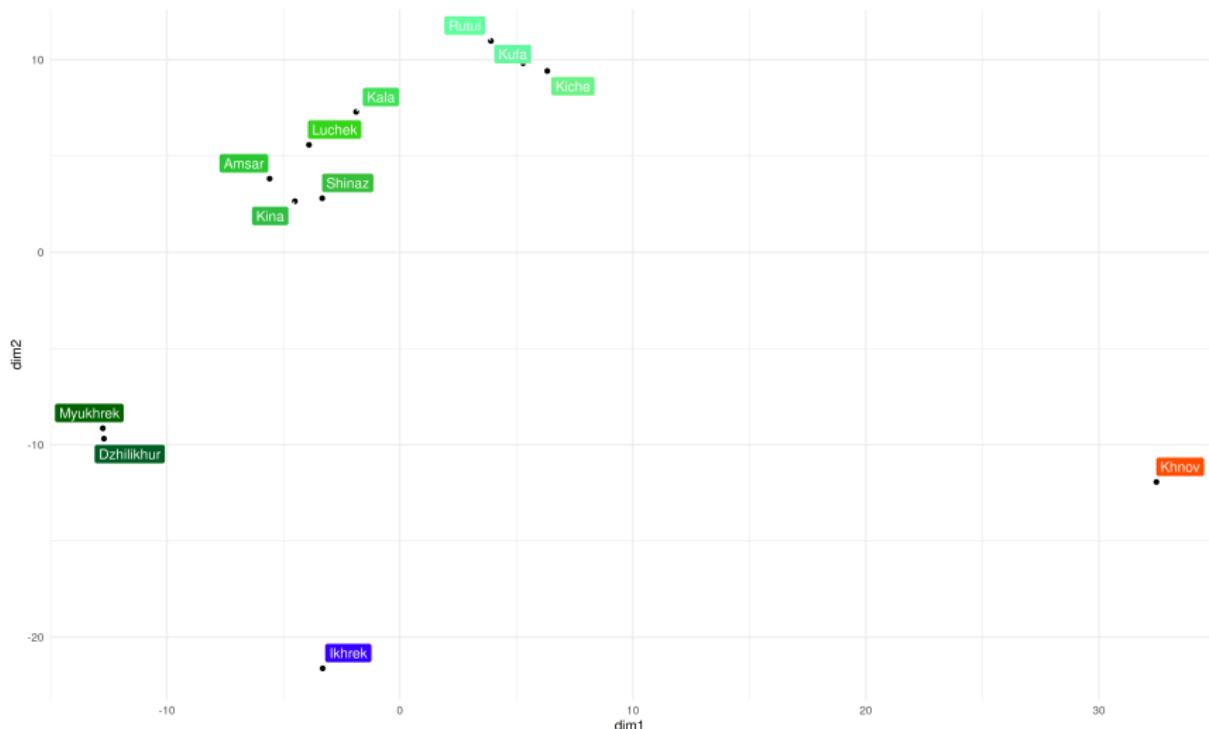


Многомерное шкалирование

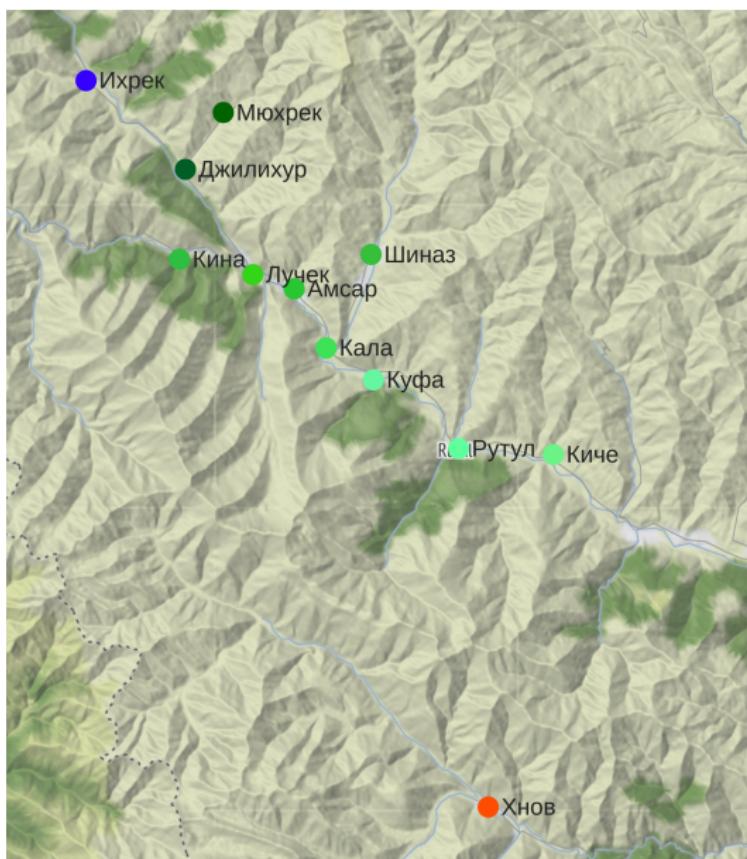
Многомерное шкалирование — один из методов уменьшения размерности. Основная цель методов уменьшения размерностей — сделать так, чтобы в пространстве меньшей размерности отобразить наблюдения, сохраняя расстояние между ними.

Чтобы совместить полученное расстояние с физическим расстоянием, лингвисты используют многомерное шкалирование для уменьшения до размерности 3, а дальше каждую полученную размерность преобразуют в цвета, используя RGB. Результат можно отобразить на обычной карте.

Многомерное шкалирование: рутульские данные



Многомерное шкалирование: рутульские данные



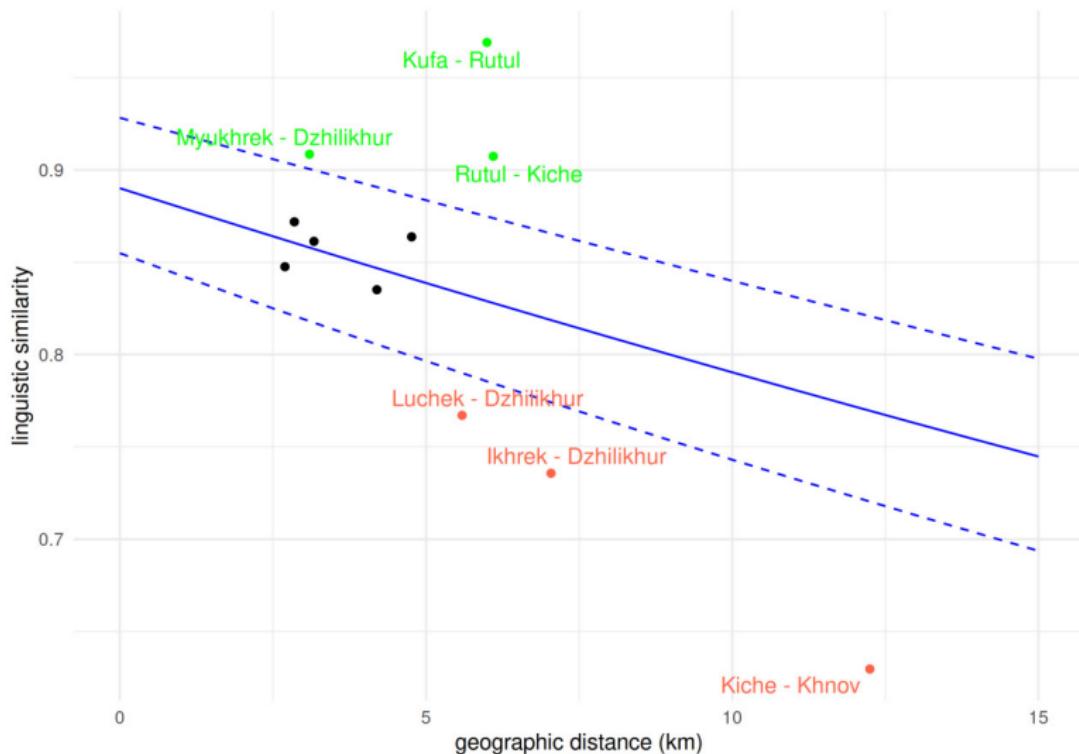
Насколько лингвистическое сходство неожиданно?

Мы задались вопросом, насколько неожиданным является лингвистическая разница между соседними деревнями.

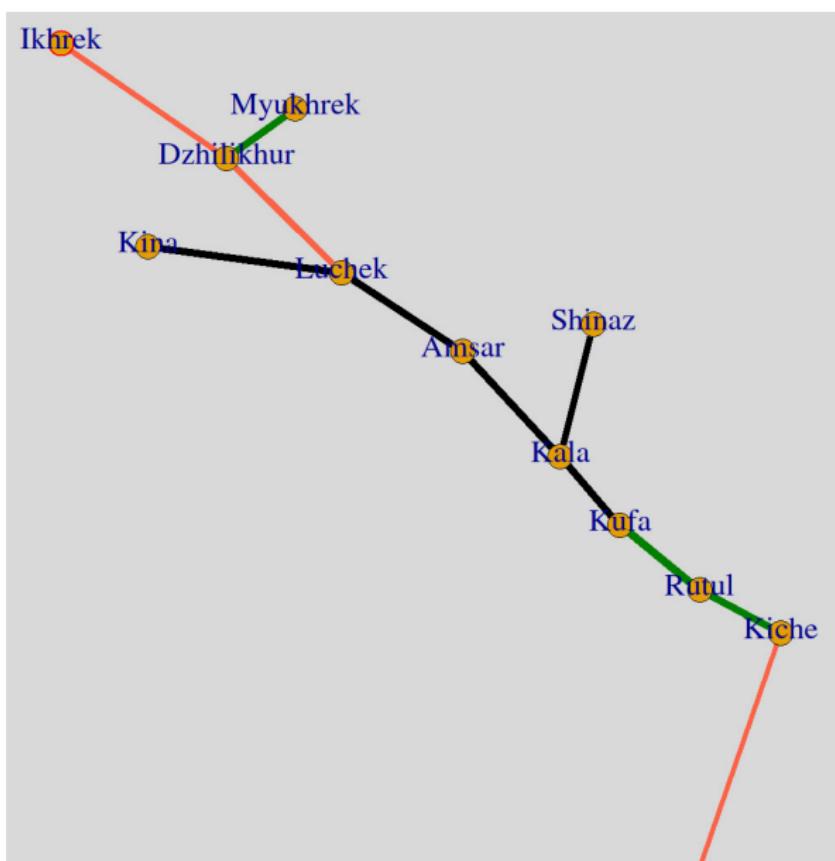
Для этого мы

- построили бета-регрессию, предсказывающую связь между лингвистическим расстоянием и физическим расстоянием
- построили граф соседства, где каждое ребро раскрасили на основе 60% доверительного интервала вокруг регрессионной линии.

Насколько лингвистическое сходство неожиданно?



Насколько лингвистическое сходство неожиданно?



Список литературы I

- G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*. Springer, 2013.
- А. С. Алисултанов and Р. А. Исламов. *Хновский (хинаад) говор рутульского языка*. АЛЕФ, Махачкала, 2022.
- А. С. Алисултанов and Т. А. Сулейманова. *Рутульско-русский словарь*. АЛЕФ, Махачкала, 2019.
- К. Э. Джамалов and С. А. Семедов. *Рутульско-русский словарь (ихрекский диалект)*. Экон-Информ, Moscow, 2006.
- А. М. Дирр. *Рутульский язык*. Издание управления кавказского учебного округа, Тифлис, 1912.
- Г. Х. Ибрагимов. *Рутульский язык*. Академия Наук СССР, Москва, 1978.

Список литературы II

Г. Х. Ибрагимов. *Рутульский язык. Синхрония и диахрония.*

Дагестанский научный центр РАН, Дагестанский
государственный педагогический университет, Махачкала, 2004.

Л. М. Ибрагимова. *Мюхрекский диалект рутульского языка.* PhD
thesis, ДГПУ, Махачкала, 2009.

М. О. Ибрагимова. *Категория падежа и морфология склонения в
рутульском языке: диалектная стратификация.* Алеф,
Махачкала, 2019.

Р. А. Исламов. *Хновский говор борчинско-хновского диалекта
рутульского языка: фонетика, морфология, лексика.* PhD thesis,
ДГПУ, Махачкала, 2014.

Список литературы III

- Э. И. Исмаилова. *Русско-рутульский словарь*. ДНЦ РАН Институт языка, литературы и искусства им. Г. Цадасы и Институт эволюционной антропологии им. М. Планка, Махачкала, 2011.
- Л. И. Лавров. Рутульцы в прошлом и настоящем. *Кавказский этнографический сборник*, 3:110–157, 1962.
- С. М. Махмудова. *Морфология рутульского языка*. Советский писатель, Москва, 2001.
- С. М. Махмудова. *Грамматические классы слов и грамматические категории рутульского языка*. PhD thesis, ДГУ, 2002.