

Использование картографирования в лингвистических исследованиях

Г. А. Мороз

30.08.2024

Лингвистическая география

Лингвистическая география

- Лингвистическая география исследует распределение языков и языковых признаков в пространстве

Лингвистическая география

- Лингвистическая география исследует распределение языков и языковых признаков в пространстве
- Лингвистическая география объединяет разные области лингвистики, где картографирование языков и языковых признаков или составление геоинформационной системы является полноценным методом получения нового лингвистического знания

Лингвистическая география

- Лингвистическая география исследует распределение языков и языковых признаков в пространстве
- Лингвистическая география объединяет разные области лингвистики, где картографирование языков и языковых признаков или составление геоинформационной системы является полноценным методом получения нового лингвистического знания
- По-английски можно встретить два термина: *language geography* и *geolinguistics*.

Лингвистическая география

- Лингвистическая география исследует распределение языков и языковых признаков в пространстве
- Лингвистическая география объединяет разные области лингвистики, где картографирование языков и языковых признаков или составление геоинформационной системы является полноценным методом получения нового лингвистического знания
- По-английски можно встретить два термина: *language geography* и *geolinguistics*.
- Я исключаю из обсуждения работы по топонимике

Картографирование в лингвистике

- визуализация какого-то идиома/языковых признаков
 - в пространстве
 - во времени
 - относительно других идиомов/языковых признаков
 - относительно нелингвистических параметров (например, высотность, температура, социополитика)

Журналы

Журналы по лингвистической географии

- Раздел Linguistic Geography в журнале PLOS ONE — (Q₁)
- Dialectologia et Geolinguistica — (Q₂–Q₄)
- Journal of Linguistic Geography

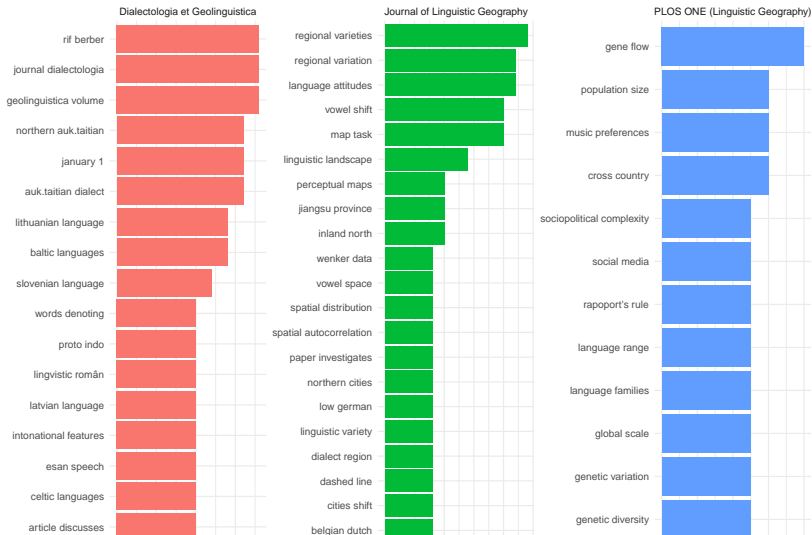
Журналы по лингвистической географии

Я **выкачал аннотации** из этих журналов (266 аннотаций):



Журналы по лингвистической географии

Топ-10 типичных биграм на основе tf-idf (266 аннотаций):



Инструменты

Инструменты для картографирования

- ESRI (ArcGIS)
- QGIS
- R пакеты
 - ggplot2
 - leaflet
 - лингвистические
 - `lingtypology`
 - `glottospace`
 - `lingtypr`
- Python пакеты
 - Arcpy
 - Geopandas
 - лингвистические
 - `lingpy`
 - `LingTypology`

Проекты Международная лаборатория языковой конвергенции

- страница ресурсов
- Корпус Просодии Русских Дialectов
- Атлас диалектов рутульского языка
- Типологический атлас языков Дагестана

Геостатистика

Обусловлено ли распределение лингвистического признака географически?

Меры для измерения

- Пространственная автокорреляция:

Обусловлено ли распределение лингвистического признака географически?

Меры для измерения

- Пространственная автокорреляция:
 - Moran's I [[Moran, 1948](#), [Odland, 1988](#)]. Мера I, изменяется от -1 до 1, где o обозначает произвольное пространственное распределение. Положительное значение означает тенденцию похожих значений кластеризоваться пространственно вместе. Отрицательное значение означает тенденцию непохожих значений кластеризоваться пространственно вместе.

Обусловлено ли распределение лингвистического признака географически?

Меры для измерения

- Пространственная автокорреляция:
 - Moran's I [[Moran, 1948](#), [Odland, 1988](#)]. Мера I, изменяется от -1 до 1, где 0 обозначает произвольное пространственное распределение. Положительное значение означает тенденцию похожих значений кластеризоваться пространственно вместе. Отрицательное значение означает тенденцию непохожих значений кластеризоваться пространственно вместе.
 - Getis-Ord G_i^* [[Ord and Getis, 1995](#)]. Эта мера тестирует локальную пространственную автокорреляцию, которая пытается ответить на вопрос, насколько каждая из точек статистически значимо отличается от соседей.

Обусловлено ли распределение лингвистического признака географически?

Меры для измерения

- Пространственная автокорреляция:
 - Moran's I [Moran, 1948, Odland, 1988]. Мера I, изменяется от -1 до 1, где 0 обозначает произвольное пространственное распределение. Положительное значение означает тенденцию похожих значений кластеризоваться пространственно вместе. Отрицательное значение означает тенденцию непохожих значений кластеризоваться пространственно вместе.
 - Getis-Ord G_i^* [Ord and Getis, 1995]. Эта мера тестирует локальную пространственную автокорреляцию, которая пытается ответить на вопрос, насколько каждая из точек статистически значимо отличается от соседей.
- Примеры использования в лингвистике: [Tamminga, 2013, Grieve et al., 2013].

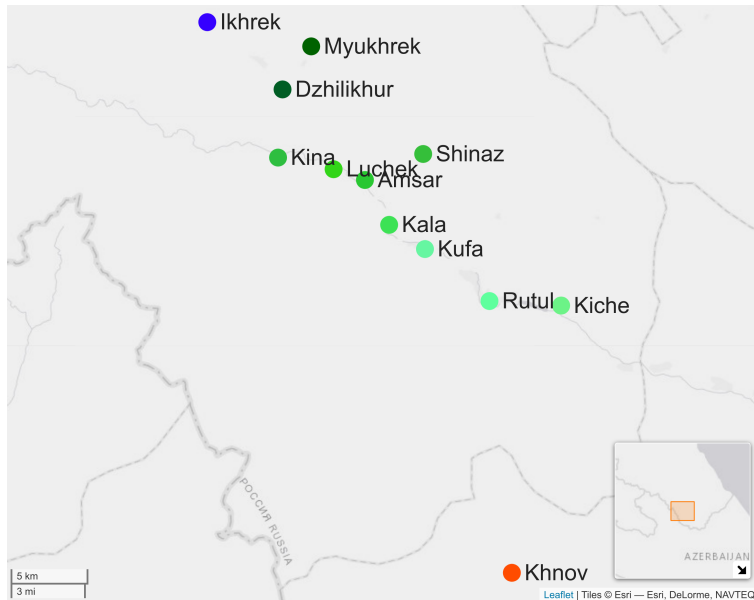
Автоматическое географическое объединение

Часто бывает так, что у лингвистов есть какое-то количество наблюдений по n переменным из k географически отдельных пунктов. В таком случае появляется задача предложить bottom-up объединения. Одно из возможных решений:

- взять некоторое подмножество признаков
- ввести меру расстояния между географическими пунктами
- сделать уменьшение размерности (чаще всего используют MDS)
- из полученного множества взять 3 размерности и преобразовать их в RGB
- оценить визуально кластеры на карте

Например, [[Wieling and Nerbonne, 2015](#), [Marchenko, 2023](#), [Sung et al., 2024](#)].

Автоматическое географическое объединение



- J. Grieve, D. Speelman, and D. Geeraerts. A multivariate spatial analysis of vowel formants in american english. *Journal of Linguistic Geography*, 1(1):31–51, 2013.
- I. A. Marchenko. *Building A Classification of Russian Dialects Using Multidimensional Scaling*. PhD thesis, HSE, 2023.
- P. A. P. Moran. The interpretation of statistical maps. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2):243–251, 1948.
- J. D. Odland. *Spatial autocorrelation*. Beverly Hills, CA: Sage Publications, 1988.
- J. K. Ord and A. Getis. Local spatial autocorrelation statistics: distributional issues and an application. *Geographical analysis*, 27(4): 286–306, 1995.

- Ho Sung, Jelena Prokić, and Yiya Chen. A new dataset for tonal and segmental dialectometry from the yue-and pinghua-speaking area. In *Proceedings of the 6th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, pages 25–36, 2024.
- M. Tamminga. Phonology and morphology in dutch indefinite determiner syncretism: Spatial and quantitative perspectives. *Journal of Linguistic Geography*, 1(2):115–124, 2013.
- Martijn Wieling and John Nerbonne. Advances in dialectometry. *Annu. Rev. Linguist.*, 1(1):243–264, 2015.