Python Library for Linguistic Typology

Michael Voronov

Higher School of Economics

18.06.2019

Introduction

Problem:

- No Python tools for online linguistic databases queries.
- No Python tools for linguistic interactive mapping.

What exists?

- R package **lingtypology** that does both (Moroz 2017).

Why Python?

- De-facto standard language among linguists.
- A lot of scientific libraries (Pandas, SciPy etc.)
- Unicode out of the box.
- Relatively high speed.
- Versatile language.

Used Tools

- Python (Python Software Foundation 2019)
- Pandas (Augspurger et al. 2019)
- Folium (Filipe et al. 2019)
- Matplotlib (Caswell et al. 2019)
- PyGlottolog (Forkel 2019)
- OpenElevationAPI (Lourenço and Developer66 2019)
- SciPy (Jones, Oliphant, Peterson, et al. 2019)

Project Description

Remote Repository:

- https://github.com/OneAdder/lingtypology

Documentation:

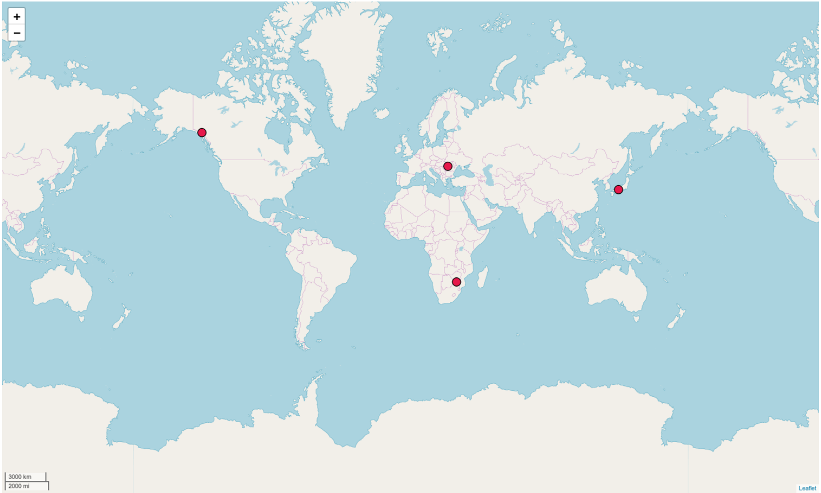
- https://oneadder.github.io/lingtypology/

Modules:

- lingtypology.maps
- lingtypology.datasets
- lingtypology.glottolog

Maps

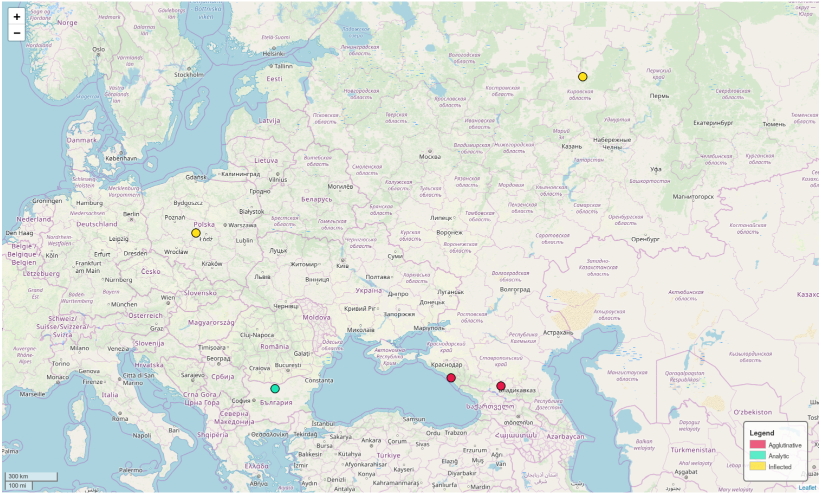
```
languages =(’Romanian’, ’Afrikaans’, ’Tlingit’, ’Japanese’)
m =lingtypology.LingMap(languages)
m.create_map()
```



Maps

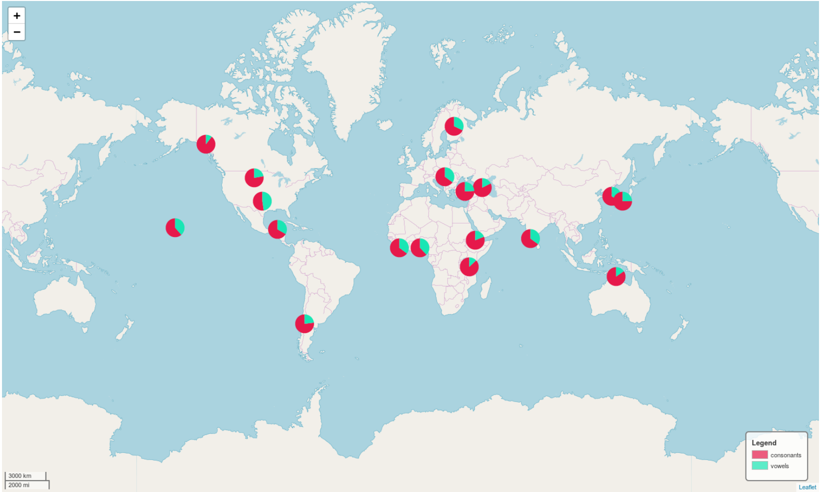
```
languages =[
    ”Adyghe”, ”Kabardian”, ”Polish”,
    ”Russian”, ”Bulgarian”
]
features =[
    ”Agglutinative”, ”Agglutinative”, ”Inflected”,
    ”Inflected”, ”Analytic”
]
m =lingtypology.LingMap(languages)
m.add_features(features)
m.create_map()
```

Maps


+
−
Legend
🟥 Agglutinative
🟩 Analytic
🟨 Inflected
300 km
100 mi
Leaflet
Norge
Oslo
Stockholm
Helsinki
Tallinn
Eesti
Latvija
Lietuva
Danmark
København
Gdańsk
Калининград
Vilnius
Гродно
Могилёв
Беларусь
Москва
Казань
Набережные Челны
Пермь
Екатеринбург
Тюмень
Уфа
Магнитогорск
Оренбург
Groningen
Hamburg
Nederland
Deutschland
Berlin
Bydgoszcz
Poznań
Polska
Warszawa
Łódź
Białystok
Lublin
Den Haag
België / Belgique / Belgien
Köln
Leipzig
Dresden
Erfurt
Frankfurt am Main
Wrocław
Kraków
Луцьк
Житомир
Київ
Lëtzebuerg
Česko
Nürnberg
München
Slovensko
Wien
Львів
Вінниця
Полтава
Суми
Липецк
Воронеж
Україна
Дніпро
Луганськ
Донецьк
Волгоград
Österreich
Schweiz/ Suisse/Svizzera/ Svizra
Graz
Magyarország
Чернівці
Moldova
Кривий Ріг
Запоріжжя
Маріуполь
Миколаїв
Астрахань
Milano
Venezia
Slovenija
Hrvatska
Cluj-Napoca
România
Timișoara
Galați
Torino
Genova
Monaco
Firenze
Città di San Marino
Sarajevo
Београд
Србија
Craiova
București
Constanța
Краснодар
Владикавказ
Marseille
Italia
Crna Gora / Црна Гора
София
България
Северна Македонија
Roma
Bari
Napoli
Shqipëria
Θεσσαλονίκη
Tekirdağ
İstanbul
Samsun
Ordu
Trabzon
საქართველო
თბილისი
Հայաստան
Azərbaycan
Palma
Palermo
Ελλάδα
Αθήνα
Bursa
Balıkesir
İzmir
Sakarya
Ankara
Çorum
Sivas
Kütahya
Afyonkarahisar
Türkiye
Erzurum
Ağrı
Elazığ
Muş
Van
Kayseri
Konya
Denizli
Aydın
Antalya
Kahramanmaraş
Şanlıurfa
Diyarbakır
Mardin
Türkmenistan
Aşgabat
O'zbekiston
Qazaqstan

Maps

```
m =lingtypology.LingMap(data.language)
m.add_minicharts(data.consonants, data.vowels)
m.create_map()
```



Databases

- WALS: The World Atlas of Language Structures (Dryer and Haspelmath 2013).
- Autotyp: an international network of typological linguistic databases (Bickel et al. 2017).
- AfBo: A world-wide survey of affix borrowing (Seifart 2013).
- SAILS: The South American Indigenous Language Structures (Muysken et al. 2016).
- PHOIBLE: ... is a repository of cross-linguistic phonological inventory data (Moran and McCloy 2019).

WALS

```
w =lingtypology.datasets.Wals(’1a’)
w.get_df().head(10)
```

	wals_code	language	genus	family	coordinates	_1A_area	_1A	_1A_num	_1A_desc
0	kiw	Kiwai (Southern)	Kiwaian	Kiwaian	(-8.0, 143.5)	Phonology	1. Small	1	Small
1	xoo	!Xóõ	Tu	Tu	(-24.0, 21.5)	Phonology	5. Large	5	Large
2	ani	//Ani	Khoe-Kwadi	Khoe-Kwadi	(-18.9166666667, 21.9166666667)	Phonology	5. Large	5	Large
3	abi	Abipón	South Guaicuruan	Guaicuruan	(-29.0, -61.0)	Phonology	2. Moderately small	2	Moderately small
4	abk	Abkhaz	Northwest Caucasian	Northwest Caucasian	(43.0833333333, 41.0)	Phonology	5. Large	5	Large
5	acm	Achumawi	Palaihnihan	Hokan	(41.5, -121.0)	Phonology	2. Moderately small	2	Moderately small
6	ach	Aché	Tupi-Guaraní	Tupian	(-25.25, -55.1666666667)	Phonology	1. Small	1	Small
7	aco	Acoma	Keresan	Keresan	(34.9166666667, -107.583333333)	Phonology	5. Large	5	Large
8	adz	Adzera	Oceanic	Austronesian	(-6.25, 146.25)	Phonology	2. Moderately small	2	Moderately small
9	agh	Aghem	Bantoid	Niger-Congo	(6.666666666669999, 10.0)	Phonology	3. Average	3	Average

WALS

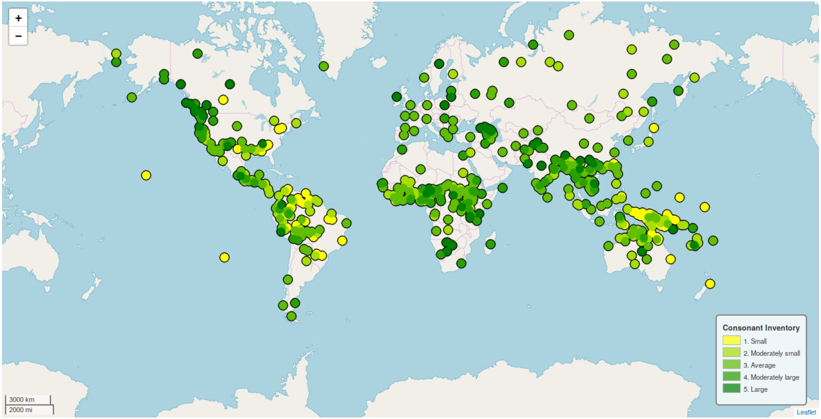
```
w =lingtypology.datasets.Wals(’1a’, ’2a’)
w.get_df().head()
```

	language	...	_1A	...	_2A	...
0	Kiwai (Southern)	...	1. Small	...	2. Average (5-6)	...
1	!Xóõ	...	5. Large	...	2. Average (5-6)	...
2	//Ani	...	5. Large	...	2. Average (5-6)	...
3	Abipón	...	2. Moderately small	...	2. Average (5-6)	...
4	Abkhaz	...	5. Large	...	1. Small (2-4)	...

Examples: WALS Features

```
wals_page =lingtypology.datasets.Wals(’1a’).get_df()
m =lingtypology.LingMap(wals_page.language)
m.add_custom_coordinates(wals_page.coordinates)
m.add_features(
    wals_page._1A,
    colors=lingtypology.gradient(5, ’yellow’, ’green’)
)
m.legend_title =’Consonant Inventory’
m.create_map()
```

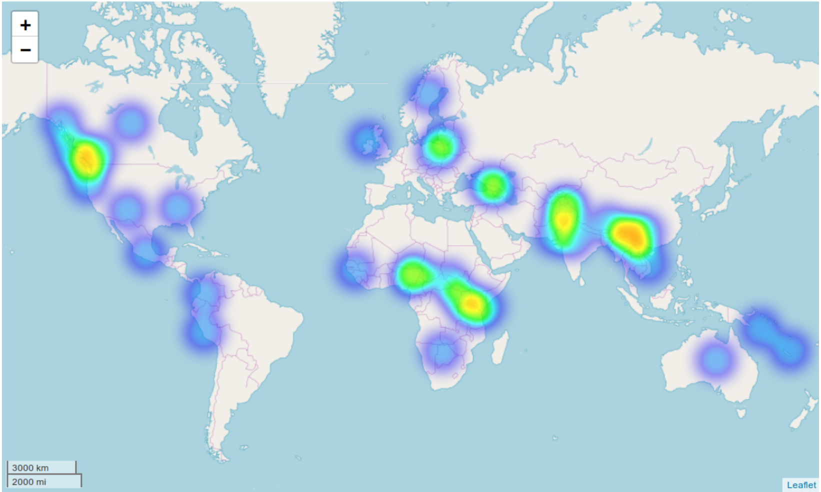
Examples: WALS Features



Examples: WALS Heatmap

```
wals =lingtypology.datasets.Wals(’1A’)
data =wals.get_df()
m =lingtypology.LingMap()
m. add_heatmap(data[data._1A_desc ==’Large’].coordinates)
m.create_map()
```

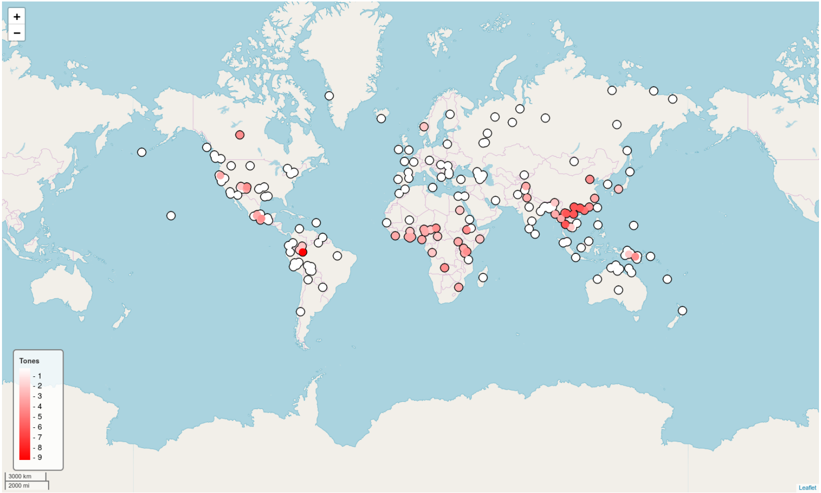
Examples: WALS Heatmap



Examples: PHOIBLE Tones

```
p =lingtypology.datasets.Phoible(subset=’SPA’)
df =p.get_df(strip_na=[’tones’])
m =lingtypology.LingMap(df.language)
m.add_features(df.tones, numeric=True)
m.colormap_colors =(’white’, ’red’)
m.legend_title =’Tones’
m.legend_position =’bottomleft’
m.create_map()
```

Examples: PHOIBLE Tones



???