

*PhytoLex* и полевая лингвистика:  
точки соприкосновения

Кира Коваленко  
Ваня Усалко  
Валерия Колосова

# Этноботаническая база данных PhytoLex



Грант РФФИ, проект 17-06-00376  
«Фитонимия русского языка  
в диахроническом аспекте  
(XI-XVII вв.)» (2017-2019)

Проект ЕУСПб «Растения и люди  
в Российской империи XVIII века:  
сословная дистрибуция знаний и  
практик» (2022-2024)

<https://phytolex.eusp.org>

# Исторические материалы о растениях: текстовые и графические данные

- рукописи
- старопечатные книги
- публикации XVIII века
- гербарные образцы
- берестяные грамоты
- эпиграфика

# Хранение источников



# Доступ к текстовым источникам



# Доступ к визуальным источникам



# Доступ к текстовым источникам

## Электронный каталог (Электронная библиотека рукописей)

[Форма поиска документов](#)

[Результат поиска документов](#)

[Описание документа](#)

[Цифровая копия документа](#)

F.VI.9/1 Травник. XVII в.



Переплет - верхняя крышка



л. 1 - 1



л. 1 об. - 2



л. 2 об. - 3



л. 3 об. - 4



л. 4 об. - 5



л. 5 об. - 6



л. 6 об. - 7



л. 7 об. - 8



л. 8 об. - 9



л. 9 об. - 10



л. 10 об. - 11



л. 11 об. - 12



л. 12 об. - 13



л. 13 об. - 14



л. 14 об. - 15



л. 15 об. - 16



л. 16 об. - 17



л. 17 об. - 18



л. 18 об. - 19



л. 19 об. - 20

# Доступ к визуальным источникам

Гербарий БИН РАН - Образцы - LE 01052282

Просмотреть

"Группы образцов"="Herbarium of Abraham Ens" По всем полям

Всего результатов: 3274

1 2 3 4 5 6 7 Следующие записи

Four herbarium specimens are displayed, each mounted on a rectangular piece of paper. The specimens are arranged in a row. Each specimen shows a dried plant with a stem, leaves, and a terminal inflorescence. The plants are mounted on a small, colorful, decorative base. A vertical scale bar is visible on the right side of each specimen, indicating a height of 150 mm. Handwritten text is visible at the bottom of each specimen's paper.

LE 01052281

LE 01052282

LE 01052283

LE 01052284



# Сведения о растениях в базе данных PhytoLex

- название в русском языке (XI-XVIII вв.)
- современное латинское название
- название в современном русском языке
- функции
- место произрастания,
- место купли/продажи
- сословие
- этнонимы
- источник
- выходные данные публикации
- жанр
- ...

# Типы полей в базе данных PhytoLex

Поля со свободным вводом

**Ботаническое название, лат.**

- [Anchusa officinalis L. - воловик лекарственный](#)

**Куратор**

Валерия Колосова

**Комментарий**

BUGLOSSUM. ... This plant, *Anchusa officinalis* of Linnaeus (Hooper 1817: 130) [яловой языкъ] яловеа языкъ. buglossum; b. 'растение воловий язык, воловик'. Запись яловой языкъ вместо воловий языкъ, ср. Травник Любч. Буглоса. Воловей языкъ (Исач.249) (Коннова)

Поля со списком значений, формируемых кураторами

- чирьи
- дшидшо (больнь отъ червей во ртѣ)
- отъ глисть
- сердце дѣлается здоровѣе и крѣпче
- сердце болитъ отъ простуды
- глисть выгоняеть
- отъ застарѣвагося жару щемить сердце
- сильный кашель съ одышкою
- вождение на языкъ
- ранки на языкъ
- если гниеть во ртѣ
- больные зубы
- слухъ возвращается
- больнь въ ухѣ
- если во ртѣ язычокъ распухнеть
- если лопаются губы
- если сердце горячо
- ломъ въ плечахъ (отъ внутренняго жара)
- кровотечение изъ носу
- боли в ушахъ
- голова болитъ (отъ горячихъ вѣтровъ)
- головная больнь (от простуды)

Поля со списком значений, формируемых администратором базы

- огородничество
- полеводство
- пчеловодство
- рыболовство
- садоводство
- скотоводство
- социальная
- атрибут вождя (венок и т.п.)
- атрибут победителя (венок и т.п.)
- курение
- наказание
- нюханье табака
- подарок
- торговля, обмен
- укрытие
- фураж, корм
- верблюд
- кролики
- медведи
- обезьяна
- птица
- голуби
- русский
- сербский
- сингальский
- средневерхненемецкий
- среднегреческий
- среднелатинский
- среднижненемецкий
- среднеперсидский
- старонитальянский
- старорусский (XV-XVII вв.)
- татарский
- турецкий
- украинский
- французский
- хинди
- церковнославянский древнеболгарского извода
- церковнославянский русского извода
- чешский
- шведский
- японский

# Компромиссы

## Частичное дублирование данных

Функция	<input type="text"/>
Функция растения как в источнике	<input type="text"/> <small>Выберите функцию из списка. Если нужная функция отсутствует, введите ее и нажмите Ввод.</small>
Болезнь человека	<input type="text"/>
Болезнь как в источнике	<input type="text"/> <small>Выберите название болезни из списка. Если нужная болезнь отсутствует, введите название и нажмите Ввод</small>

<input type="checkbox"/> EТHNONYM
<input type="checkbox"/> "азиаты"
<input type="checkbox"/> "американцы"
<input type="checkbox"/> "восточные народы"
<input type="checkbox"/> "древние славяне"
<input type="checkbox"/> "израильтяне"
<input type="checkbox"/> "индейцы"
<input type="checkbox"/> "римляне"
<input type="checkbox"/> "сибиряки"
<input type="checkbox"/> "славяне варяжские"
<input type="checkbox"/> абугач

Пополнение стандартизованных списков значениями, не вписывающимися в современную научную классификацию (пишем в кавычках)

<input type="checkbox"/> MEDICINAL FORM	<input type="checkbox"/> ЯЗЫК
<input type="checkbox"/> "вода"	<input type="checkbox"/> "азиатский"
<input type="checkbox"/> "водка"	<input type="checkbox"/> "бразильский"
<input type="checkbox"/> "дух"	<input type="checkbox"/> "восточный"
<input type="checkbox"/> "спирт"	<input type="checkbox"/> "индейский"
<input type="checkbox"/> аэрозоль	<input type="checkbox"/> "славянский"
<input type="checkbox"/> бальзам	<input type="checkbox"/> авестийский

# Типы региональных источников

Полевые дневники

Аудиокассеты

Видеокассеты

Электронные записи

Расшифровки

=====

Excel (фитонимы)

Тамбовские губернские ведомости

<https://elibrary.tambovlib.ru/?id=ebook.tgv>

Тамбовские епархиальные ведомости

<https://elibrary.tambovlib.ru/?id=ebook.tev>

Другие печатные материалы

Архивные материалы

# Полевые материалы (1994 – ). 312 строк

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Код информант	Село / деревня	Район	местное	научное	использу	дикораст	Хранени	Хранение II	Приготовлени	Применени	медицин	ветерина	животно	пища (не тип обря	обряд (Т; быт (стро	кост	
2	СМИ2007Тамб	Пахотный Угол	Бондарский	лебеда			дикорастущее			вареное					суп (щи)			
3	СМИ2007Тамб	Пахотный Угол	Бондарский	чувель	щавель		дикорастущее			сырое					перекус			
4	СМИ2007Тамб	Пахотный Угол	Бондарский	одуванчик			дикорастущее				растирание	болезнь ног						
5	СМИ2007Тамб	Пахотный Угол	Бондарский	огурцы		плод	культурн	засол		засол	внутри	лихорадка						
6	СМИ2007Тамб	Пахотный Угол	Бондарский	дуб		ствол	дикорастущее				протаскивал	грыжа						
7	СМИ2007Тамб	Пахотный Угол	Бондарский	дуб		растение	дикорастущее				обходили	грыжа						
8	СМИ2007Тамб	Пахотный Угол	Бондарский	лебеда			дикорастущее					куриная слепота						
9	СМИ2007Тамб	Пахотный Угол	Бондарский	куриная слепота			дикорастущее							крупный рогатый скот				
10	ТКИ2007Тамб	Пахотный Угол	Бондарский	купарис	кипарис	древесин	дикорастущее											
11	ТАД2007Тамб	Пахотный Угол	Бондарский	вилок	капуста	листья	культурн	засол	деревянная	засол		рак						
12	ТАД2007Тамб	Пахотный Угол	Бондарский	картошка	картофель	клубни	культурное			вареное		рак			гарнир			
13	ТАД2007Тамб	Пахотный Угол	Бондарский	травки (с.н.)			дикорастущее											
14	СЕК2007Тамб	Кривополянье	Бондарский	конопи	конопля	семечки	культурное											насыпают у двора предполагаем
15	ПВД1992Тамб	Знаменка	Знаменский	картошка	картофель	клубни	культурное											как подручно
16	ПВД1992Тамб	Знаменка	Знаменский		конопля	коноплян	культурное											фитиль для к
17	ПВД1992Тамб	Знаменка	Знаменский		лен	льняные	культурное											фитиль для к
18	ФСН1993Тамб	Прудовка	Сампурский	полянья	полян	растение	дикорастущее											подстилка дл
19	ФСН1993Тамб	Прудовка	Сампурский	конопля		стебли	культурное											прядение и т
20	ФНД1993Тамб	Прудовка	Сампурский	конопля		стебли	культурное											прядение и т
21	ПАМ1993Тамб	Пановы Кусты	Сампурский	лен		стебли	культурное											прядение и т
22	ПАМ1993Тамб	Пановы Кусты	Сампурский	конопля		стебли	культурное											прядение и т
23	СЕК2007Тамб	Кривополянье	Бондарский	табак (ма табак		листья	культурное				прикладыва	сибирка						
24	КАТ1993Тамб	Пановы Кусты	Сампурский	ёлка	ель	ветвь	дикорастущее											семейны похороны

# Типы лингвистических карт

- Лексические (от значения к слову) [*Viola tricolor* L.]
- Семантические (от слова к значению) [пижма]
- Лексико-словообразовательные [объединяются с лексическими]
- Словообразовательные
- Мотивационные [подорожник]
- Номинативные (однословные / составные номинации) [объединяются с лексическими]

**лексико-словообразовательные карты:** практически любая карта, построенная по принципу «от значения к слову» эксплицирует и лексические, и словообразовательные различия

# Анализ

Данные полевых исследований диалектов имеют следующие особенности:

Код  
информанта

Село /  
деревня

Район

местное  
название

научное /  
литературное  
название (если  
отличается от  
диалектного)

используемая  
часть (кора, лист,  
древесина, семя,  
ветвь, луб, почка,  
бутон, цветок,  
побег, корнеплод)

miro

дикорастущее  
/ культурное /  
покупное

Атрибут у Usages

Хранение  
СПОСОБ  
(сушка, засол,  
настойка,  
толчение...)

Хранение  
МЕСТО

Атрибут у Usages

Приготовление (сырое,  
отвар, ферментация,  
засол, вымачивание,  
заваривание, жарка,  
выпечка, стирание,  
толчение, расчесывание)

Применение (внутри,  
мазь, растирание,  
прикладывание,  
битье, вдыхание,  
закапывание,  
полоскание, ванны)

медицина  
(местное  
название  
болезни)

miro

# Анализ (продолжение)

Каждая карточка (колонка в excel файле)

ветеринария  
(болезнь)

ЖИВОТНОЕ

пища (перекус, закуска, ароматизаторы, специи, приправы, суп, каша, пироги, выпечка, гарнир, варенье, компот, морс, кисель, наливка, настойка, настой)

тип обряда  
(календарь, семейный, животноводческий, магия...)

обряд (Троица, выгон скота, свадьба, похороны, родины, крестины, эпидемия, эпизоотия)

быт  
(строительство, ремесла, корзиноплетение, инсектицид...)

miro

косметика,  
гигиена

Фольклор (легенды, сны, быличка, мемораты, песни, считалки, загадки, проклятия, благопожелания, табу)

Атрибут у Usages

Игры  
(номинация, в игровом тексте, названии команд), игрушки

когда  
использовали  
(всегда / только сейчас / давно / недавно бросил / в детстве / недавно научился)

Атрибут у Usages

источник (мама, бабушка, сосед(ка), подруга, ТВ, книга, газета, радио, интернет)

Атрибут у Usages

КТО  
ИСПОЛЬЗОВАЛ

Атрибут у Usages

miro



# Анализ (продолжение)

Синим цветом отмечены концепты представленные в phytalex

цитата

комментарии  
собирателя

Атрибут у Citation

Латынь

семейство

Нужно ли его хранить?

По латыни из  
внешнего классификатора

Гербарный  
образец

индекс  
района по  
сетке  
ЛАРНГ

год  
рождения

пол  
респондента

год  
записи

ИМЯ  
папки

формат записи  
материала  
(полевая,  
аудиозапись,  
видеозапись)

порядковый  
номер  
респондента  
на кассете

# Результаты анализа

Из 35 особенностей данных  
26 представлены в PhytoLex  
концептами. (60%)

Остальные 9 могут быть  
добавлены как  
дополнительные атрибуты

Концепты представлены в виде отношений в  
реляционной базе

**Для загрузки данных необходима функция импорта**  
**Функция импорта должна включать в себя возможность**  
**повторного импорта и обладать свойством идемпотентности** (т.е.  
Повторная загрузка одинаковой информации не приводит к ее  
дублированию).

# Список доработок

- Новые атрибуты
- Функция импорта

## Персональные данные?

- Это боль (ФЗ-152 и GDPR) дают всем по рукам
- Необходимо управляемое обезличивание
  - Часть данных (ФИО) хранится отдельно в Excel файле у владельца данных
  - Код информанта используется везде где требуется ссылка на источник информации или исследователя
  - Поиск только по коду информанта

Вопросы?

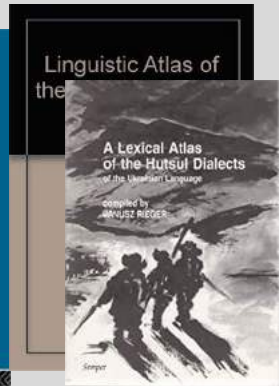
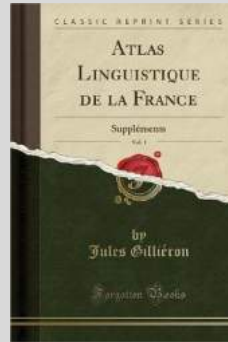
# Лингвистические карты

- Зачем?
- Кому может помочь?
- Как?



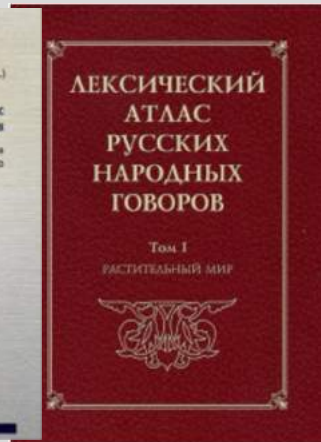
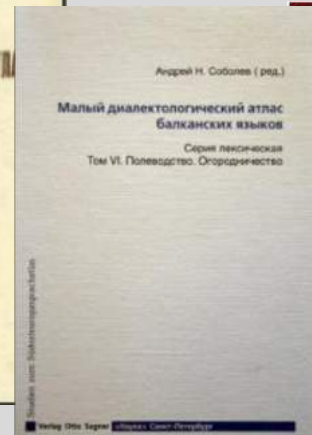
# NonDigital:

- Atlas linguistique de la France (Jules Gilliéron and Edmond Edmont) 1902-1910
- Linguistic Atlas of New England (Hans Kurath) 1939
- The Linguistic Atlas of Scotland (James Y. Mather and Hans-Henning Speitel) 1975
- Linguistic Atlas of the Upper Midwest (Harold Allen) 1973
- Linguistic Atlas of the Gulf States (Lee Pederson) 1991
- A lexical atlas of the Hutsul dialects of the Ukrainian language (Rieger, Janusz) 1996
- Atlas Linguarum Europae inspired by Roman Jakobson (1930) and Mario Alinei (1965) (UNESCO) from 1975 to 2007



# NonDigital:

- Общеславянский лингвистический атлас(1958)
- Общекарпатский диалектологический атлас (1987)
- Малый диалектологический атлас балканских языков (2003)
- Лексический атлас русских народных говоров ([2004], 2017, 2022)



# Digital:

- Treasury of Linguistic Maps Online (academic access only) 2022
- Linguistic Atlas Project (<http://www.lap.uga.edu/>)
- Digital Language Atlas of China (Crissman, Lawrence W.) 2012
- Lingtypology + Glottolog database (George Moroz) 2016-2023
- [www.muturzikin.com](http://www.muturzikin.com) (2005-2019)



Data  
Download  
Center

LICHEN  
Graphical  
User Interface

Commons  
Uploads  
to Share

[Home](#)  
[Atlas Projects](#)  
[News](#) Updated!  
[Help](#)  
[Links](#)  
[About Us](#)  
[Contact](#)  
[The Old Site](#)  
[Conditions of Use](#)  
[of LAP Materials](#)

## The site for information about American English.

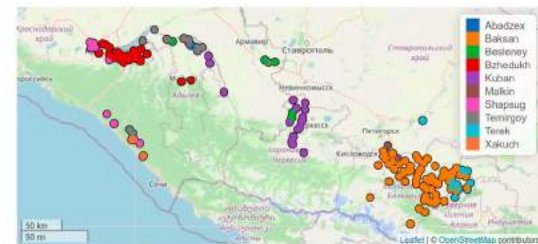
We offer information about English as it is spoken in the United States. Most of the projects included here present the results of survey research carried out between 1930 and 1980; some are more recent. Click the Atlas Projects link to the left for more information.

## Data Download Center

Click the above button if you want to navigate our collection of thousands of text and sound files available to download. This works best if you already know about what we have.

## LICHEN Graphical User Interface

When this tool appeared, creating maps became a matter of one line of code. This opened up new horizons: as with any visualization, linguistic maps were often used as a tool to illustrate an idea. Now maps have become a research tool, that allows the visualization of different linguistic variables. The map below, for example, shows the distribution of Circassian dialects in Russia. If a researcher wants to investigate a feature based on a representative sample with different dialects, they can use a map to choose maximally distant settlements from each dialect group.

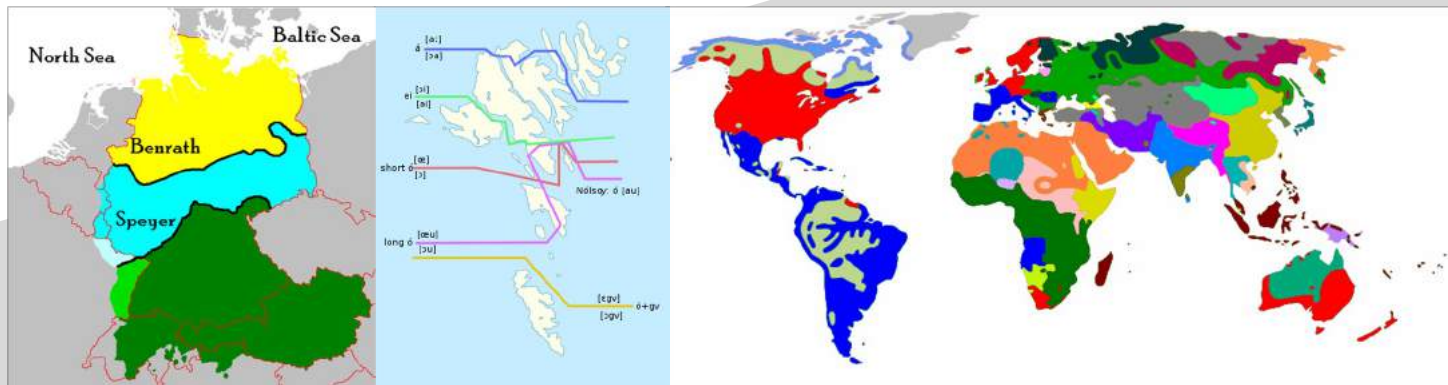


Another function of the package that many linguists are not aware of yet, is that it provides access to linguistic databases. Currently *lingtypology* supports eight databases, created by international collectives of linguists. The most widely known ones are *Glottolog*, *WALS*, *AUTOTYP*, *PHOIBLE* and *Affix Borrowing database*. With *lingtypology* researchers can download data from these sources with a single command and combine these data with their own to visualize the relationships.



# Зачем все это?

- Для проверки гипотез, которые сложно проверить другим способом
  - Например, гипотеза о магическо-религиозной мотивации формирования языка (Mario Alinei)
- Для оценки лексической эрозии (без визуального представления на карте трудно сделать логические выводы)
- Для возможности вывода новых знаний. Привязка лексической информации к географии обогащает данные



# Кому?

- Диалектологам и всем, кто пытается посмотреть на лексические данные в географическом разрезе

# Как?

- Через адаптивность в представлении данных
  - После выхода non-digital атласов со стороны исследователей посыпались критические замечания о способах представления данных
- Через многослойные представления
  - Перевод данных в цифровой формат позволяет делать сопоставления через наложение слоев, это расширяет возможности для insight'a

# СЛОЖНОСТИ:

- Источники карт (много, но все разного качества)
  - Так в большинстве источников (Google, Yandex, MapBox) отсутствует из коробки функция управляемого обеднения, а svg tiles которые могли бы ее заменить (есть у Google и MapBox) требуют вложений в разработку svg фильтров.
- Исторический разрез, представлен в основном в виде оцифрованных старых карт (трудности с извлечением исторических данных)
- Геокодинг - это боль, оптимальное решение включить эту функцию на этапе импорта данных, что усложняет импорт

# Технологии:

- HTML5 - наше все, ориентируемся только на браузеры с его полноценной поддержкой (включая WebGL API).
- WebGL - для возможности быстрого наложения слоев и сглаживания переходов
- Deck.gl - фреймворк который позволяет превращать карту в нечто большее, добавляет возможности интерактива и упрощает реализацию идей
- React.js - для упрощения интеграции с существующим решением.

Вопросы?