



МОСКОВСКИЙ  
ПОЛИТЕХ

Г. В. Векшин, М. М. Лемешева

Электронный «Живой стилистический  
словарь русского языка»:  
результаты и перспективы

# «Живой стилистический словарь русского языка» –

справочный и корректировочный лингвистический сервис для работы с любыми текстовыми данными

Цель:

создание электронного лингвистического справочника – **веб-сервиса**, который позволит **автоматически представлять стилистический портрет** русских слов и выражений, **формировать стилистические пометы**, многосторонне отражая контекстуально-ролевые тяготения единиц, динамику социокультурного и эмоционально-оценочного употребления.

<https://livedict.syllabica.com>

# ЖИВОЙ СТИЛИСТИЧЕСКИЙ СЛОВАРЬ РУССКОГО ЯЗЫКА

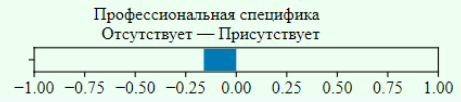
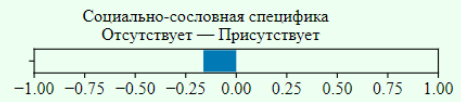
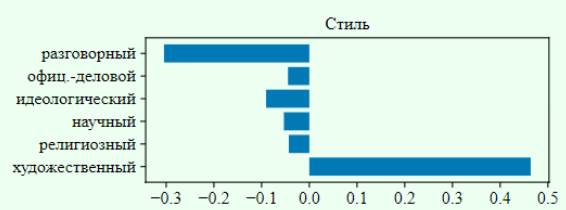
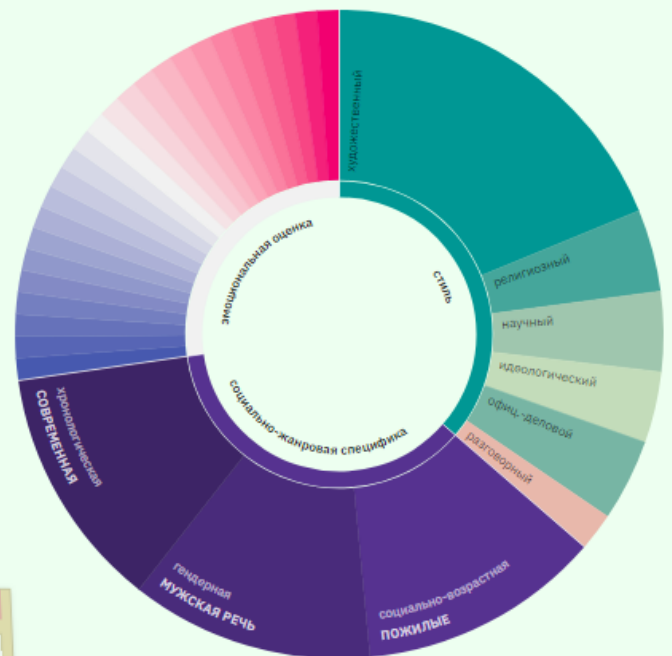
Лингвистическая подготовка и создание электронного  
«Живого стилистического словаря русского языка»  
Linguistic Development and Creation of the Electronic Live Stylistic Dictionary of Russian  
Создан при поддержке РФФИ (грант № 17-04-00421)

- Словарь
- Стилиевой подсказчик
- Просмотр корпуса словаря
- Краткое руководство

## Стилистическая характеристика

брожу

первая строка — заголовки, кодировка — utf8/65001, размер — 1.9Gb(архив — 415Mb).



## Что лежит в основе ЖСС?

- авторская типология стилистической семантики, отражающая ее простейшие компоненты;
- алгоритмы автоматического определения стилистической окраски;
- технология обработки сетевых данных и автоматического пополнения корпуса словарс по принципу Web as a Corpus;
- методы глубокого машинного обучения и создания нейросетей.

<https://livedict.syllabica.com>

# Стилистическая окраска

Ощутимый носителями языка «привкус» типичной среды, а с нею и наиболее типичной, ожидаемой роли говорящего, традиционно действующего в данной среде. **Маркер типичного контекста и** конвенционально привязанных к нему форм поведения, коммуникативных ритуалов и **ролей**.

# Структура социокультурного пространства и функциональные языки общения



# Типические **социокультурные роли** и формируемые ими типы стилистических окрасок – **стили языка**



КОММУНИКАЦИЯ:  
УСТАНОВКИ

**БЫТЬ**  
(ЦЕЛЬ)

**КАЗАТЬСЯ**  
(РОЛЬ)



ЛИТЕРАТУРНЫЕ ПРАКТИКИ

**ОБЛАСТЬ ПОЭТИЧЕСКОГО:  
ИСКУССТВО**

субъективное освоение вечности  
переживание делания мира

**ОБЛАСТЬ ПОЭТИЧНОГО:  
СОЧИНИТЕЛЬСТВО**

социокультурная реализация  
в роли поэта



**ЯЗЫК СЛОВЕСНОГО ИСКУССТВА,  
ПОЭТИЧЕСКИЙ ЯЗЫК**

Стратегии остраняющей  
многозначности  
Худ. тактики, операции, приемы

**ХУДОЖЕСТВЕННЫЙ СТИЛЬ ЯЗЫКА,  
ПОЭТИЧЕСКИЙ СТИЛЬ**

**Поэтизмы**  
(экспоненты роли поэта и контекста  
литературного сочинительства)

**РЕЧЕВАЯ ДЕЯТЕЛЬНОСТЬ**

**ЯЗЫКОВАЯ СИСТЕМА**





# Семантическая структура языковых единиц



Социокультурно обусловленная контекстная и ролевая семантика

Оценочно обусловленная контекстная и ролевая семантика

Стилистическая окраска

## Схема формирования корпусной базы ЖСС

Типичные маркеры социокультурной роли/контекста + тематические элементы – источник стайлсетов для формирования поисковых запросов и последующего извлечения стилистически однородных текстов



Ты не зови меня порочным;  
В моей измученной судьбе,  
Я другом был тебе не вечным,  
Но думал нежно о тебе.  
И часто летнею порою  
За рощей пышной, молодой,  
Ты помнишь ли, гулял с тобою  
Я мрачный, бледный и худой?  
И долго я тоской мятежной  
Не мог прошедшее будить  
И ручкой милой, белоснежной,  
Уста свои я шевелить.

*(Н. Ганецкий, «Ты не зови меня порочным...», 1913)*

- *Ведь ты сама прекрасно знаешь, / какое место занимаешь, / **В моей измученной судьбе**, / да и вообще здесь, на Земле (автор неизвестен);*
- *Мне всё напоминает о тебе: / И солнце, и земля, и небо. / Ведь ты была важнее хлеба / **В моей измученной судьбе** (Елена Венская);*
- *Устал гореть огонь лучисто / **В моей измученной судьбе** / Закат взирает серебристо, / И нет отрадушки нигде (Лидия Лазор);*
- *Устал от пьянок и гуляний. / Устал топить печаль в вине / Оно лишь все усугубляет / **В моей измученной судьбе** (Кирилл Володин);*
- *Теперь же только о Тебе / Все мысли сны мои и настроения / Ты солнца луч **в моей измученной судьбе** / Моя надежда на души спасение (Андрей Петров-Павел);*
- *Ты дитя моё возьми к Себе / Да поближе к сердцу приголубь, / И **в моей измученной судьбе** / Светом веры непременно будь! (Татьяна Маругова);*
- *Ты свет **в моей измученной судьбе**, / Моя дорога к призрачному дому...(Виктор Северенков).*

## Сетевая ПОЭЗИЯ

### СТАТИСТИКА

Всего пользователей: 56 194

Читателей: 24 787

Современников: 31 206

Классиков: 201

Стихотворений: 1 445 833

Стихотворений современников: 1 429 206

Стихотворений классиков: 15 880

Поэмбук



СТИХИ

ПОЭТЫ

ДУЭЛИ

АЛЬБОМ

ДНЕВНИКИ

РЕЙТИНГИ

# ЖИВОЙ СТИЛИСТИЧЕСКИЙ СЛОВАРЬ РУССКОГО ЯЗЫКА

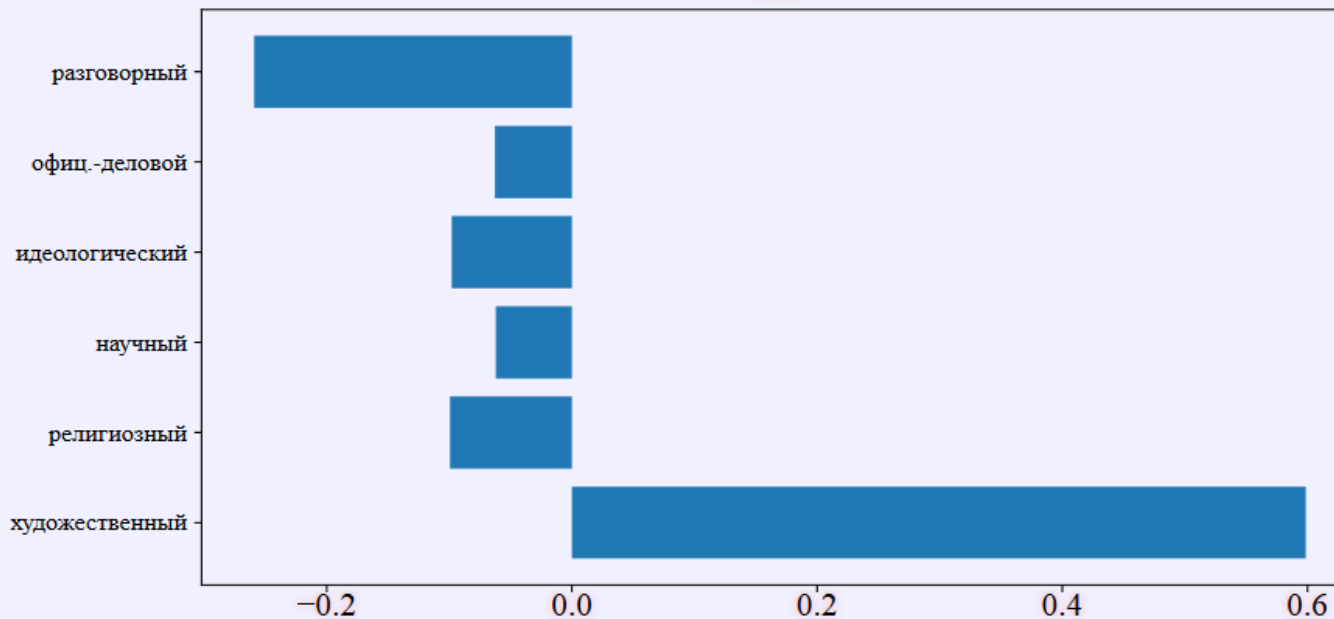
Лингвистическая подготовка и создание электронного  
«Живого стилистического словаря русского языка»  
Linguistic Development and Creation of the Electronic Live Stylistic Dictionary of Russian  
Создан при поддержке РФФИ (грант № 17-04-00421)

Поиск в словаре

## Приложения к отчету по гранту РФФИ № 17-04-00421, 2018 год

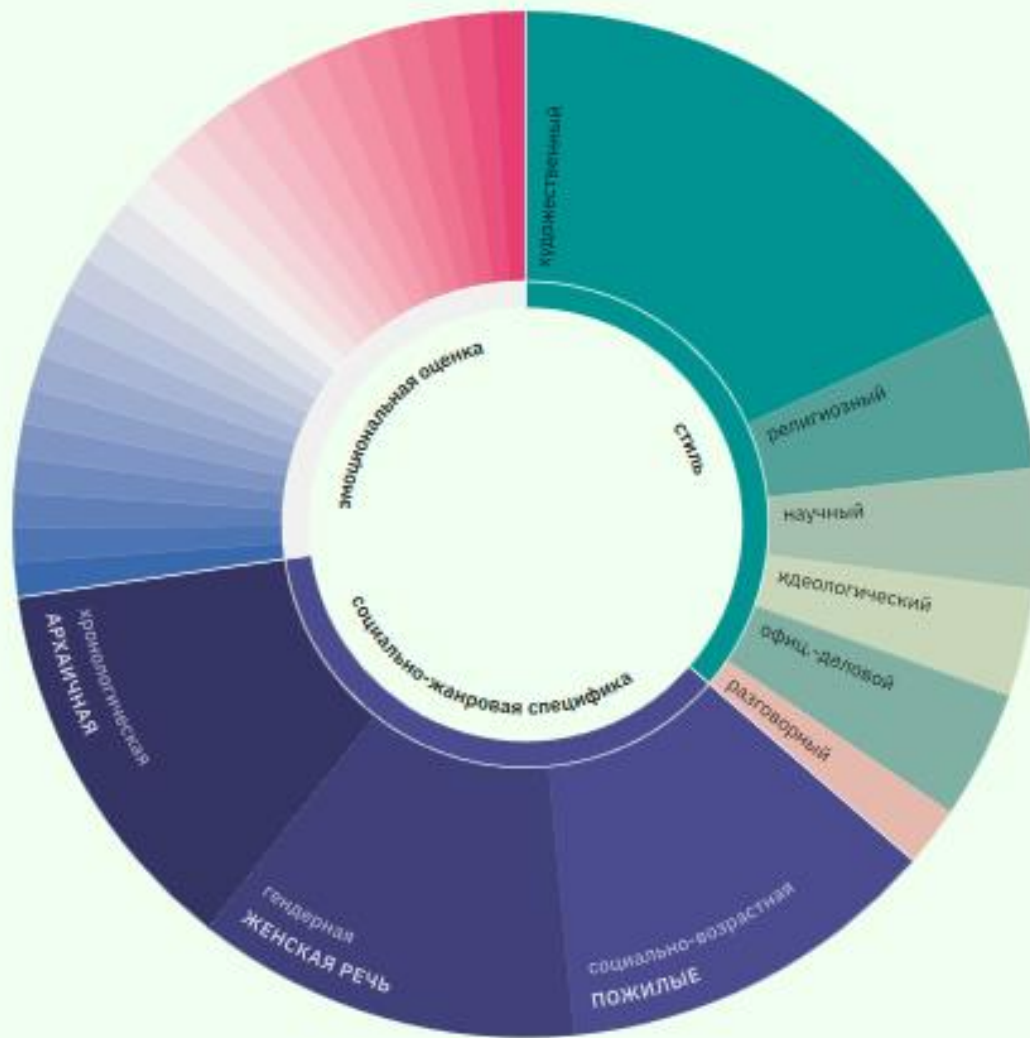
- Векшин Г.В., Кретов А.А., Кукушкина Е.Ю., Леденева В.В., Шилихина К.М.  
Социокультурно обусловленная стилистическая семантика русского языка: социально-жанровый компонент (к алгоритмам описания лексики в проекте «Живого стилистического словаря русского языка»).

МИГ

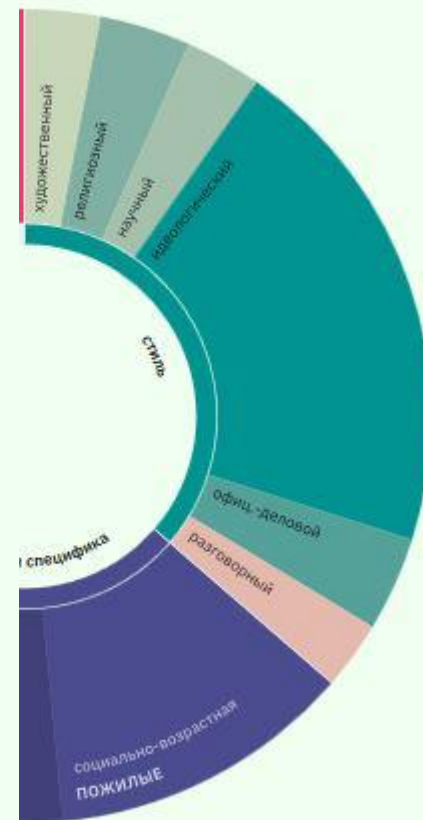




# безумство



# военщина



**Стилистический портрет слова в Живом стилистическом словаре русского языка**

# Machine learning

## Logistic Regression Approach

- Using Bag of Words for text classification
- Moving from Bag of Words to Bag of Morphemes
- The longer a morpheme is the more it is valuable to a model
- Word prefixes and suffixes counted as separate morpheme



$$style\_probability = \sigma \left( \sum_i \omega * M_i \right)$$

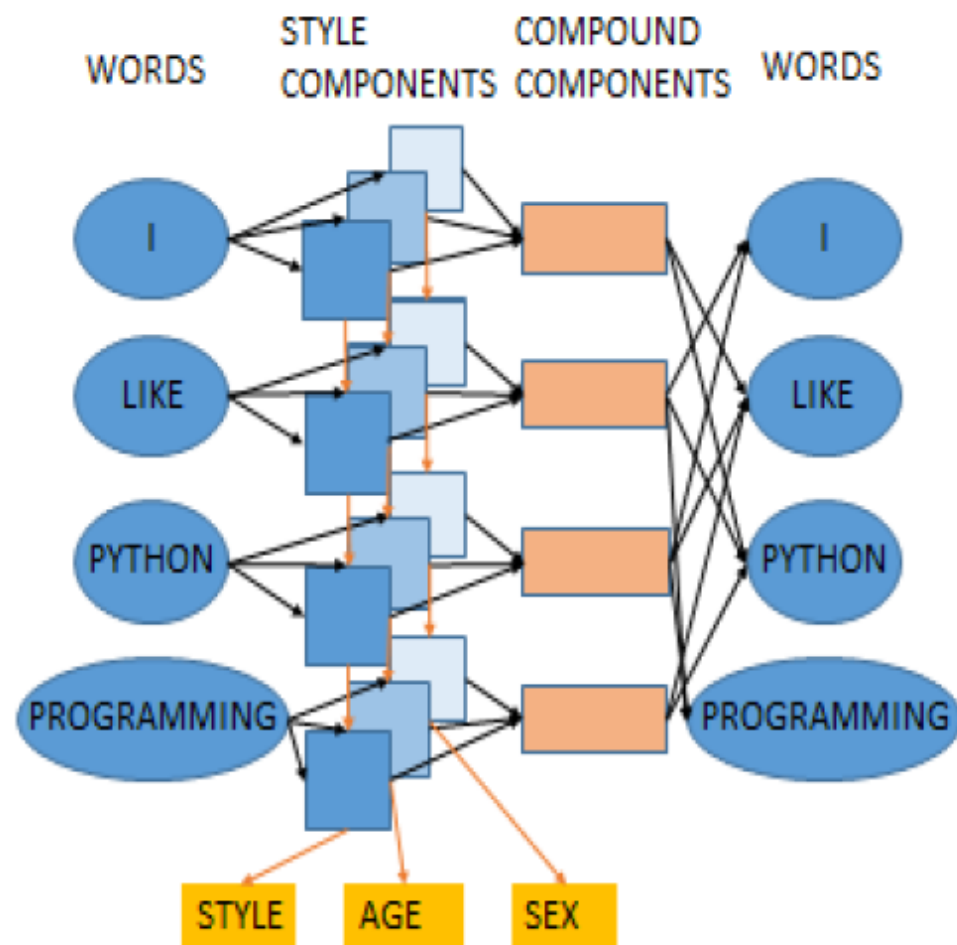
	1	2
fiction	4.13e-05	-1.14e-04
religious	1.20e-04	1.14e-03
scientific	-2.29e-06	7.49e-05
ideological	-2.44e-08	-2.29e-05
business	-1.87e-06	-2.27e-04
colloquial	-3.35e-07	-9.45e-06

	n-1	n	n+1	...	N
	-7.79e-04	2.65e-04	1.10e-04		2.72e-05
	-1.96e-03	2.18e-04	6.29e-04		-3.80e-05
...	4.93e-03	-4.10e-08	7.95e-07	...	-5.65e-07
	-6.82e-05	5.96e-07	2.14e-07		5.29e-08
	-5.12e-03	-1.45e-05	5.14e-06		4.66e-07
	-8.53e-06	3.08e-06	5.46e-07		1.67e-06



# Contextual Word2Vec

- In classical W2V contextual information is distributed along the components
- Fine tuning W2V model for a certain task
- All necessary information should be held in certain components
- **Parallel training of W2V model and text classification.**



# Экспертная оценка результатов проекта в 2017-2019

## Результаты экспертизы

### Экспертное заключение 1

**Обоснование оценки "Уровень полученных результатов" (доступно заявителю, не допускается копирование содержания пунктов, необходимо объяснить сделанный выбор):**

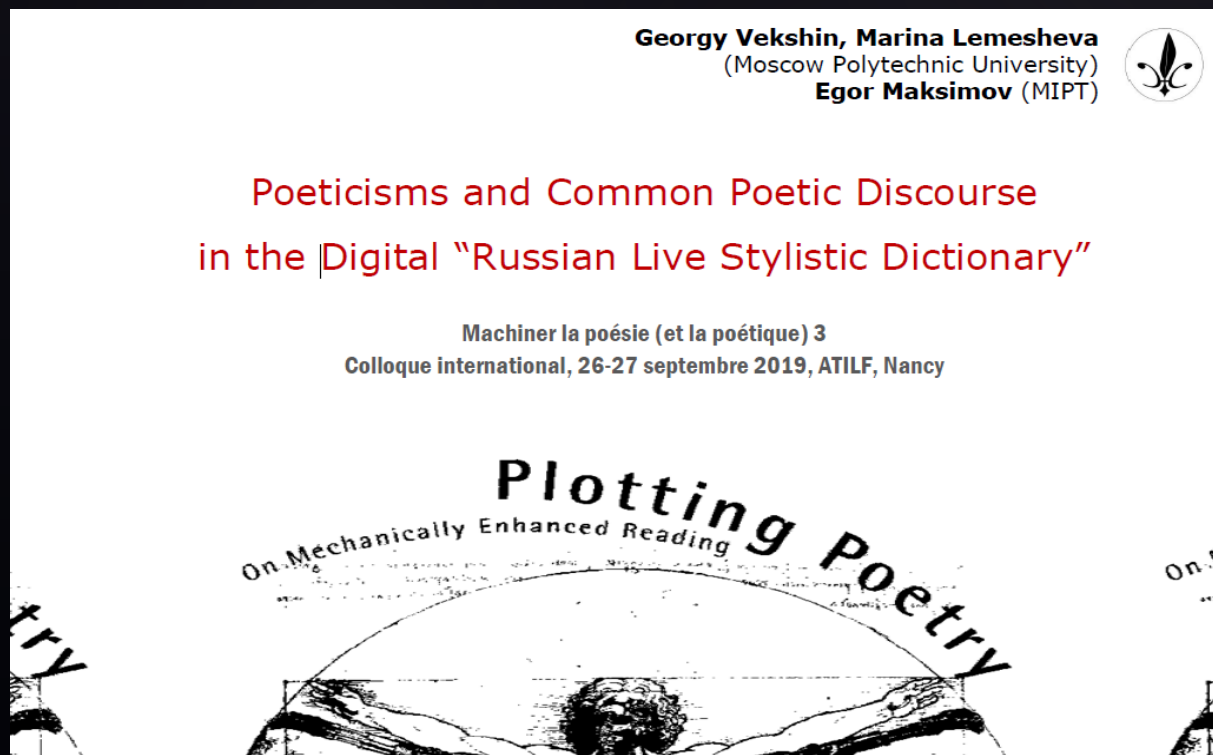
Полученные в ходе реализации данного проекта результаты можно рассматривать как прорывные, поскольку они, во-первых, позволяют получить принципиально новое знание о специфике стилистического аспекта лексики современного русского языка; во-вторых, представляют чёткие критерии описания стилистических черт современного русского лексикона; в-третьих, формируют принципиально новую методологию описания стилистической стороны лексических единиц современного русского языка; в-четвёртых, позволяют формализовать описание специфики стилистического аспекта содержания лексических единиц так, что оказывается возможным применение информационно-коммуникационных технологий для однозначной интерпретации данного аспекта содержания слов; в-пятых, позволяют существенно обогатить представления о лексике современного

**Заключительные замечания эксперта по Отчету: (Ваши замечания будут доступны заявителю.)**

Исследование, проведённое в ходе реализации рассматриваемого проекта, можно рассматривать как масштабное, фундаментальное, принципиально новаторское, обладающее высокой степенью актуальности, своевременное и современное; как обладающее большой теоретической значимостью и колоссальным потенциалом практического применения его результатов в широком диапазоне областей, связанных с практикой современной общественной коммуникации в её сугубо человеческом проявлении либо в аспекте, предполагающем использование

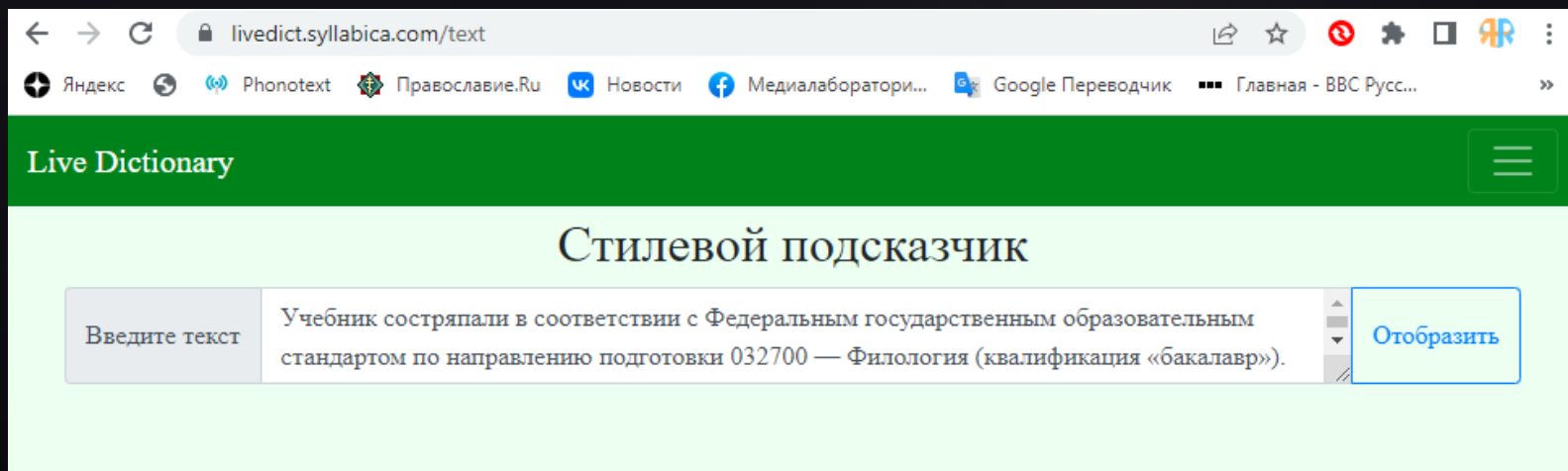
## Апробация технологии и результатов проекта

- Технология «Живого стилистического словаря русского языка» признана международным научным сообществом.
- Проект апробирован на крупных международных конференциях и в научных публикациях в России и за рубежом.
- Проект до сих пор не имеет аналогов в мире.



## Перспективы разработки «Живого стилистического словаря»

### ❖ Style-checker для работы с любыми текстовыми данными



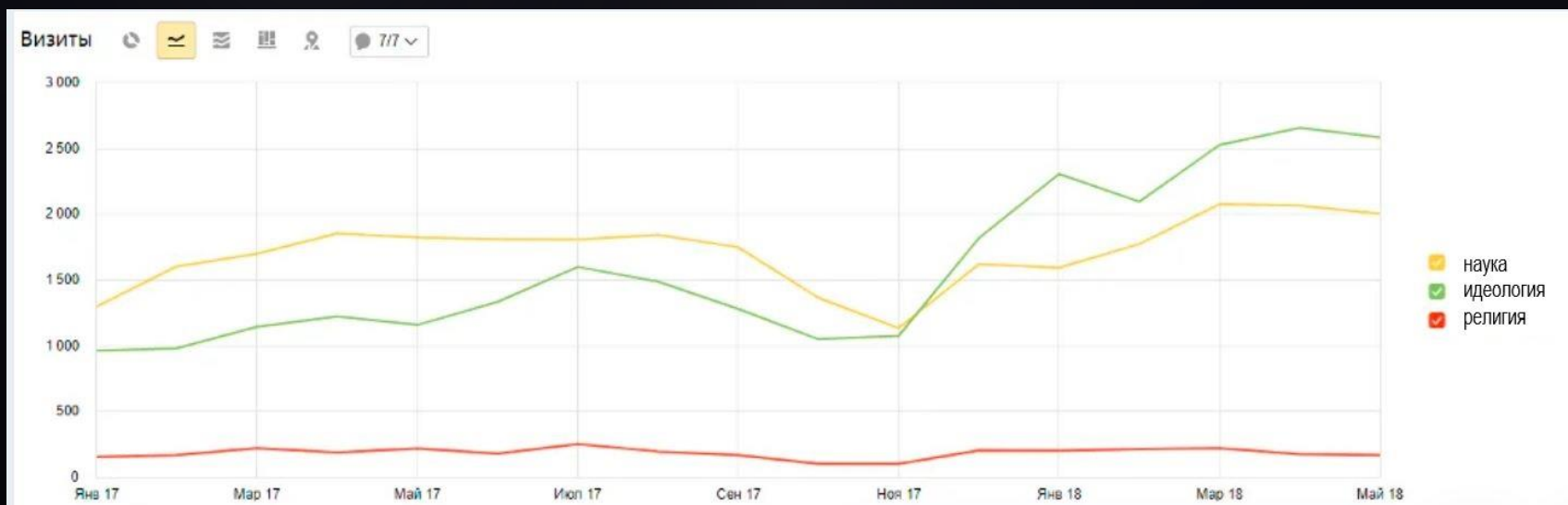
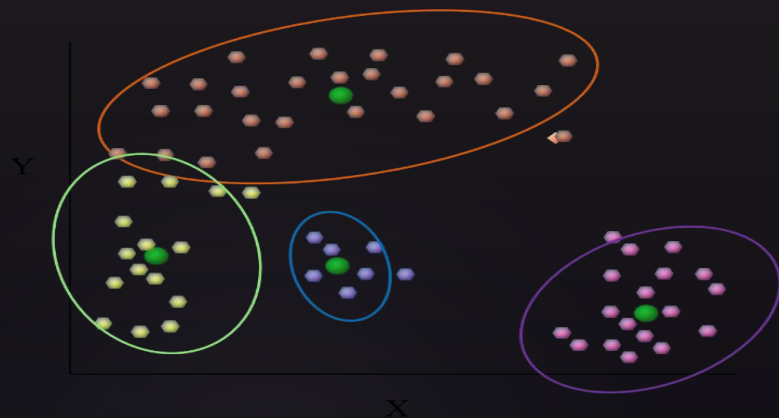
Стилистический конфликт: разг., пренебр., негативн.

Учебник состряпали в соответствии с Федеральным государственным образовательным стандартом по направлению подготовки 032700 — Филология (квалификация «бакалавр»).

## Перспективы разработки «Живого стилистического словаря»

### ❖ Data mining

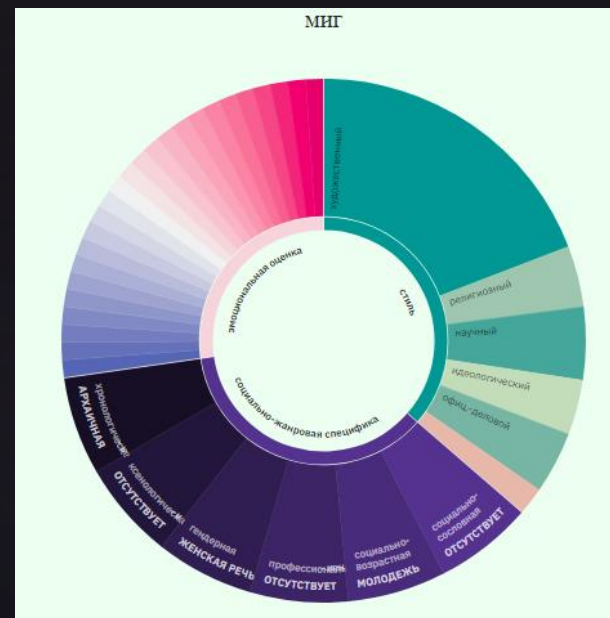
Автоматическое пополнение корпусов словаря. Мониторинг и кластерный анализ текстовых ресурсов сети для установления и уточнения актуальных тенденций словоупотребления, динамики изменения стилистического значения слова.



## Сферы применения

### «Живого стилистического словаря»

- ❑ стилистическая корректировка работы с любыми текстовыми данными
- ❑ редактирование;
- ❑ копирайтинг;
- ❑ обучение языку как родному и иностранному (на всех уровнях);
- ❑ практика перевода;
- ❑ судебная лингвистическая экспертиза;
- ❑ автоматическое реферирование/аннотирование текста;
- ❑ русская лексикография.





# Литература

- ⌘ Georgy Vekshin, Marina Lemesheva, Egor Maximov. Poeticisms and Common Poetic Discourse in the Digital Russian Live Stylistic Dictionary // Computational Stylistics in Poetry, Prose, and Drama. Berlin: De Gruyter Mouton < 2022. Pp. 153-178.  
DOI: [10.1515/9783110781502-009](https://doi.org/10.1515/9783110781502-009)
- ❖ Векшин Г. В., Лемешева М. М. Поэт как речевая роль: к семантике и прагматике русского поэтизма // Вестник РУДН. Серия: Теория языка. Семиотика. Семантика. 2019, № 4. С. 35-44.  
DOI: [10.22363/2313-2299-2019-10-4-1067-1087](https://doi.org/10.22363/2313-2299-2019-10-4-1067-1087)
- ❖ Векшин Г. В. Основы стилистической семантики. - Москва : МИРЭА, 2017. - 127 с.
- ❖ Векшин Г.В., Леденёва В.В. и др. Стилистическая семантика русского языка: Коммуникативная роль – контекст – слово (к основаниям «Живого стилистического словаря»). – М.: Издательский центр «Азбуковник», 2019. – 288 с.
- ❖ Векшин Г.В. Социальная среда и стилистическая семантика (к описанию лексики в проекте "Живого стилистического словаря русского языка") // Проблемы изучения живого русского слова на рубеже тысячелетий. – Воронеж, ВГПУ, 2017. – С. 9-16.
- ❖ Векшин Г. В., Лемешева М. М. Семантика поэтизма и стихотворческий дискурс (из опыта работы над «Живым стилистическим словарем русского языка» // Вестник ВГУ. Серия: Лингвистика и межкультурная коммуникация. 2019, № 4. С.33-43.
- ❖ Авамилова Э. А., Векшин Г. В. и др. Алгоритмы составления корпусов «Живого стилистического словаря русского языка» и его программная разработка // Книга в современном мире: место в культурной парадигме общества в условиях цифровой революции: Сборник научных статей / науч. ред. Ж. В. Грачева. – Воронеж, 2019. С. 4-19.

Благодарим за внимание

©Г.В. Векшин, М.М. Лемешева, 2024  
philologos@yandex.ru  
<https://livedict.syllabica.com>