

---

**Research Data**

---

07 Apr 2022

**Creating a corpus of all possible word-forms of modified Russian sound verbs using web-scraping methodology. Compilation and adjustment of summary tables for the future tense, imperative, and gerund forms. Forms of “dual action”. Experimental multi-dimensional scaling of web-scraping results.**

Irina V. Ivliyeva

*Missouri University of Science and Technology, ivliyeva@mst.edu*

Perry Koob

*Missouri S&T Information Technology, koobp@mst.edu*Follow this and additional works at: [https://scholarsmine.mst.edu/research\\_data](https://scholarsmine.mst.edu/research_data) Part of the [Russian Linguistics Commons](#)

---

**Recommended Citation**

Ivliyeva, Irina V., Koob, Perry. Creating a corpus of all possible word-forms of modified Russian sound verbs using web-scraping methodology. Compilation and adjustment of summary tables for the future tense, imperative, and gerund forms. Forms of “dual action”. Experimental multi-dimensional scaling of web-scraping results. June 2021 – April 2022. Working theses. March 2022.

This Data is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Research Data by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact [scholarsmine@mst.edu](mailto:scholarsmine@mst.edu).

---

**Creating a corpus of all possible word-forms of modified Russian sound verbs using web-scraping methodology. Compilation and adjustment of summary tables for the future tense, imperative, and gerund forms. Forms of “dual action”. Experimental multi-dimensional scaling of web-scraping results.**

**Abstract**

Modern Russian dictionaries do not include all possible forms of words. To compile such an index, even for one part of a speech, on paper is not technically feasible. With the appearance of electronic versions of dictionaries, however, for the first time we can try to create an inventory of all possible forms for the lexical-semantic group of Russian sound verbs, using the web-scraping methodology. This project attempts to develop a number of comprehensive tables for the prefixed (semantically modified at the word-formation level) sound verbs and all their forms. A novel, four-position system of numbering the verbal forms to support experimental multi-dimensional scaling of results have been introduced. The research output takes into the account not only all documented (recorded) modifications of the source verbs, but also all potential derivative forms. The compilation of the tables, the labelling of the elements and the subsequent analysis of their correlations revealed some technical execution and documenting challenges that are directly related to the semantics of verbal modifications (e.g. the concurrence of the future tense forms, the imperatives of verbs of sound and forms of “dual action”). The project outcomes may positively impact the development of diverse web-based applications for gathering, visualizing, or analyzing data from a variety of digital lexicographic sources across a single or multiple language, corpora, or from other digital text collections.

**Disciplines**

Russian Linguistics

**Comments**

Ивлиева, И.В., Куб, Перри. Создание корпуса всех возможных словоформ модифицированных русских глаголов звучания методом веб-извлечения. Составление и корректировка сводных таблиц для форм будущего времени, повелительного наклонения, и деепричастий. Формы совместного действия. Экспериментальное многомерное шкалирование результатов веб-извлечения. Июнь 2021 – апрель 2022. Рабочие тезисы. Март 2022.

Ивлиева, И.В., Куб, Перри. **Создание корпуса всех возможных словоформ модифицированных русских глаголов звучания методом веб-извлечения. Составление и корректировка сводных таблиц для форм будущего времени, повелительного наклонения, и деепричастий. Формы совместного действия. Экспериментальное многомерное шкалирование результатов веб-извлечения.** Июнь 2021 – апрель 2022. Рабочие тезисы. Март 2022.

Ivliyeva, Irina V., Koob, Perry. **Creating a corpus of all possible word forms of modified Russian sound verbs using web-scraping methodology. Compilation and adjustment of summary tables for the future tense, imperative, and gerund forms. Forms of “dual action”. Experimental multi-dimensional scaling of web-scraping results.** June 2021 – April 2022. Working theses. March 2022.

## **ПОСТАНОВКА ПРОБЛЕМЫ.**

Система глагольных форм в русском языке имеет **матричную структуру** (существуют как полные парадигмы глаголов, так и дефектные - типа «*побежу*», «*пропылесосу*»). Для чего необходимы таблицы глагольных форм? Прежде всего, чтобы заполнить «матрицы» (таблицы) каждого глагола всеми имеющимися в русском языке актуальными словоформами.

Известно, что современные словари не приводят **все** возможные формы слов. Воспроизвести в бумажном издании формы слов даже одной части речи - технически неисполнимая задача. Однако с появлением электронных версий словарей впервые возникает шанс её решить.

На базе электронных словарей в 2020 году работа велась по составлению таблиц, включающих формы всех времен (наст., буд., прош.) и наклонений (изъявительное, повелительное), учитывая дифференциацию по родам (м., ж., ср.), числам (ед., мн.) и лицам (1-е лицо, 2-е лицо, 3-е лицо).

Подробнее об этом смотрите: Ivliyeva, Irina V., Koob, Perry. “*Lacunae Matrices, Web Scraping. Theses.*” (2021). Russian Linguistics Research. 1. [Electronic resource]. – Access mode: [https://scholarsmine.mst.edu/russian\\_linguistics\\_research/1](https://scholarsmine.mst.edu/russian_linguistics_research/1)

**ХРОНОЛОГИЯ ПРОЕКТА в период с мая 2021 по апрель 2022 года**

Сроки	Описание, задачи, результаты
Май 2021	Анализ проблемы, постановка задач, анализ устройства электронного грамматического словаря А.А. Зализняка <a href="https://morfologija.ru/">https://morfologija.ru/</a>
Июнь 2021	<b>Задача I.</b> Создание первого варианта компьютерной программы поиска для веб-извлечения.
30 июня 2021	Обращение к электронному словарю <a href="https://morfologija.ru/">https://morfologija.ru/</a> , веб-извлечение <i>последовательными поисками</i> а) форм будущего времени и б) форм повелительного наклонения модифицированных глаголов звучания. Результаты: <a href="#">Приложение 2</a> , <a href="#">Приложение 2А</a> , <a href="#">Приложение 3</a> , <a href="#">Приложение 3А</a> .
Июль – октябрь 2021	Анализ результатов, усовершенствование программы поиска, оптимизация доступа к источнику
16 ноября 2021	Обращение к электронному словарю <a href="https://morfologija.ru/">https://morfologija.ru/</a> , веб-извлечение <i>одноразовым совместным поиском</i> форм будущего времени, повелительного наклонения и деепричастий модифицированных глаголов звучания. Результаты: <a href="#">Приложение 4</a> .
Ноябрь 2021 – февраль 2022	<b>Задача II.</b> Анализ результатов, дальнейшая отладка программы поиска, оптимизация представления результатов, расширение поиска (одновременно в словаре Зализняка и НКРЯ).
21 февраля 2022	Обращение одновременно к электронному словарю Зализняка <a href="https://morfologija.ru/">https://morfologija.ru/</a> и к Национальному Корпусу русского языка <a href="https://ruscorpora.ru/new/index.html">https://ruscorpora.ru/new/index.html</a> ; веб-извлечение <i>одновременным параллельным поиском</i> форм будущего времени, повелительного наклонения и деепричастий модифицированных глаголов звучания. Результаты: <a href="#">Приложение 5</a> . Анализ результатов, дальнейшая отладка программы поиска, оптимизация представления результатов.
24 февраля 2022	Обращение к электронному словарю <a href="https://morfologija.ru/">https://morfologija.ru/</a> : <b>доступа нет</b> . См. Примечание 1.
Март – апрель 2022	Анализ результатов, дальнейшая отладка программы поиска, оптимизация представления результатов (экспериментальное многомерное шкалирование).

### **Примечание 1. О доступе в базу электронных ресурсов.**

Результаты, зафиксированные в сводной Таблице 5, не являются точными, полными или окончательными. Данные результаты можно назвать промежуточными, так как получены они были в процессе обращения к источникам (Зализняк и НКРЯ) 21 февраля 2022 года, после окончательной отладки поисковой компьютерной программы.

Попытка получения обобщающих результатов в полном объёме 24 февраля 2022 года не принесла успеха: поисковая компьютерная программа **не получила доступа** к основному источнику - [Грамматическому словарю А.А. Зализняка](#). Авторы проекта планируют регулярно (21-ого числа ежемесячно) проверять доступ к данному источнику и документировать наличие или отсутствие доступа к нему при помощи алгоритма веб-извлечения.

### **ЗАДАЧА I.**

В 2021 году исследователями была создан первый вариант компьютерной программы для заполнения индивидуальных таблиц формами *будущего времени* ([Приложение 2](#), [Приложение 2А](#)) и *повелительного наклонения* ([Приложение 3](#), [Приложение 3А](#)), а также *деепричастными* формами глаголов звучания на материале следующего источника: Грамматический словарь А.А. Зализняка. The grammatical dictionary of the Russian language A.A. Zalyzniak. Access mode: <https://morfologija.ru/>.

### **ТРУДНОСТИ.**

В процессе заполнения индивидуальных таблиц, маркирования элементов и последующего анализа их состава выявлены некоторые особенности технического характера, которые связаны непосредственно с семантикой модификаций:

1. Совпадение форм будущего времени и императива глаголов звучания и форм совместного действия;
2. Омонимия форм модификаций у глаголов близкой семантики;
3. Необходимость введения четырехступенчатой нумерации элементов таблицы.

Все это требует дополнительных пояснений.

### 1. Совпадение форм будущего времени и императива глаголов звучания и форм совместного действия.

В [Таблице 2, 2А](#) собраны все формы будущего времени глаголов звучания, которые с помощью специально разработанной программы извлечены из грамматического словаря А.А. Зализняка.

Пример 1. Формы будущего времени глаголов звучания

A	B	C	D	F
Number Номер	Root Основное слово	Derivative Дериват	Pronoun Местоимение	Future tense Будущее время
1.0.0	ахать	<b>ахнуть</b>	--	
1.0.1.0	ахать	ахать	Я	--
1.0.2.0	ахать	ахать	Ты	--
1.0.3.0	ахать	ахать	Он/она	--
1.0.4.0	ахать	ахать	Мы	--
1.0.5.0	ахать	ахать	Вы	--
1.0.6.0	ахать	ахать	Они	--
<b>1.1.0.0</b>	<b>ахать</b>	<b>ахнуть</b>	--	<b>ахнуть</b>
1.1.1.0	ахать	ахнуть	Я	ахну
1.1.2.0	ахать	ахнуть	Ты	ахнешь
1.1.3.0	ахать	ахнуть	Он/она	ахнет
1.1.4.0	ахать	ахнуть	Мы	ахнем
1.1.5.0	ахать	ахнуть	Вы	ахнете
1.1.6.0	ахать	ахнуть	Они	ахнут

Нетрудно заметить, что при заполнении графов с местоимением **мы** появляются формы, на первый взгляд, дублирующие друг друга: **мы ахнем**, **мы ахнемте**, **мы поахаем**, **мы поахаемте** и т.д.

Это явление распространено в русском языке: одно и то же слово в разных контекстах выполняет разные функции. Так, форма *ахнем* может выступать в одном контексте в роли будущего времени, а в другом – в роли императива. Поисковая же программа считает данную информацию как дублиеты.

Возникает вопрос, стоит ли дифференцировать данные явления в рамках таблицы, и если да, то как.

Ниже приведены примеры из [таблицы 3, 3А](#), которые демонстрируют вариативность и синонимию глагольных форм: *мы ахнем*, *(мы) ахнемте*, *мы проахаем*, *(мы) проахаемте* и т.п. Глаголы на *-те*, как видим, на первый взгляд, дублируют формы будущего времени.

*Пример 2. Формы будущего времени и формы совместного действия*

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>F</b>	<b>G</b>
<b>Number Номер</b>	<b>Root Основное слово</b>	<b>Derivative дериват</b>	<b>Pronoun Местоимение</b>	<b>Future tense Будущее время</b>	<b>Imperative Повелительное наклонение</b>
1.1.4.1	ахать	ахнуть	Мы	ахнем	ахнем
1.1.4.2	ахать	ахнуть	Мы	ахнем	ахнемте
1.3.4.1	ахать	заахать	Мы	заахаем	заахаем
1.3.4.2	ахать	заахать	Мы	заахаем	заахаемте
1.4.4.1	ахать	поахать	Мы	поахаем	поахаем
1.4.4.2	ахать	поахать	Мы	поахаем	поахаемте
1.6.4.1	ахать	проахать	Мы	проахаем	проахаем
1.6.4.2	ахать	проахать	Мы	проахаем	проахаемте

Это так называемые **формы совместного действия** – то есть формы мн. числа глаголов, не относящиеся к какому-либо времени.

Повелительное наклонение у подобных форм образуется с помощью постфикса *-те* к форме 1 л. мн. числа: **мы споем / споемте**, **мы скажем / скажемте** и т.п. Сочетание одного местоимения с разными формами глагола вызывает у «неносителя» языка вопрос о дублиетности данных форм, что не совсем так.

У форм совместного действия, которые очень редко употребляются в современном языке (исключение – *споемте* и еще несколько глаголов незвуковой семантики), существуют дублиетные образования со значением побудительности, но они

образованы аналитически: с помощью частицы *давай(те)* и некоторых др. + форма буд. времени: **давайте сыграем (сыграемте), давайте споем (споемте), давайте крикнем (крикнемте).**

Формы совместного действия образуются не у всех глаголов: **мы говорим**, но не *\*говоримте*, **мы кричим**, но не *\*кричимте* и т.п., а только у глаголов со значением однократного действия (совершенного вида). Ср., например: **мы играем**, но не *\*играемте* (если нет значения однократности, то и нет формы совместного действия); **мы сыграем** - присутствует значение кратности и появляется форма совместного действия - **сыграемте**.

Формы совместного действия в некоторых случаях, например, если отсутствует постфикс **-те**, могут совпадать с формами **будущего времени** – **сыграем!** (*давайте сыграем! Сыграемте!*).

Ср., например, формы буд. времени - **Мы сыграем концерт** и побудительное обращение: - **Сыграем в две руки! Сыграемте в две руки! Давайте сыграем в две руки!**

В подобных случаях формы совместного действия выступают как формы **повелительного наклонения**. Еще одно отличие конструкций с **ахнем** и **ахнемте** - в количестве участников ситуации.

## **АХНЕМ**

1. Буд. время (*когда-нибудь мы ахнем*)

2. Императив, побудительное действие:

**а)** с одним участником и говорящим: *давай ахнем!* (здесь опущено словосочетание «**мы с тобой**», но оно предполагается в контексте: *давай (мы с тобой) ахнем!*)

## **АХНЕМТЕ**

Императив, побудительное действие:

**б)** с несколькими участниками и говорящим: **давайте ахнем – ахнемте!** (опущено, но предполагается местоимение *мы*: **давайте (мы) ахнем – ахнемте!**)



В формах **будущего времени** число предполагаемых участников действия **не дифференцируется**, это могут быть как два участника, так и несколько: *мы побарабаним* (может быть и с одним участником – *мы (с тобой) побарабаним*, и с несколькими – *мы (с вами) побарабаним*; *мы (с тобой) как ахнем!* (2 участника); *мы (все) как ахнем!* (неопределенное количество участников + сам говорящий).

В рамках нашей программы по составлению полной таблицы форм глагольных модификаций все формы совместного действия можно поместить в отдельную таблицу, технически это нетрудно сделать. Но зависимость данных форм от контекста не дает возможности однозначно их интерпретировать. В зависимости от речевой ситуации и интонации говорящего данные формы могут выступать в совершенно разных значениях.

Все эти доводы не позволяют нам на этапе составления таблиц дифференцировать *формы совместного действия* по времени и наклонению. Корректным решением в данной ситуации будет размещение этих единиц в таблицы будущего времени и императива.

*Пример 3. Императив. Формы совместного действия*

A	B	C	D	G
Number Номер	Root Основное слово	Derivative дериват	Pronoun Местоимение	Imperative Повелительное наклонение
1.0.0.0	ахать	ахать	--	ахать
1.0.1.0	ахать	ахать	Я	--
1.0.2.0	ахать	ахать	Ты	ахай
1.0.3.0	ахать	ахать	Он/она	--
1.0.4.0	ахать	ахать	Мы	--
1.0.5.0	ахать	ахать	Вы	ахайте
1.0.6.0	ахать	ахать	Они	--
1.1.0.0	ахать	ахнуть	--	ахнуть
1.1.1.0	ахать	ахнуть	Я	--

1.1.2.0	ахать	ахнуть	Ты	ахни
1.1.3.0	ахать	ахнуть	Он/она	--
1.1.4.1	ахать	ахнуть	Мы	ахнем
1.1.4.2	ахать	ахнуть	Мы	ахнемте
1.1.5.0	ахать	ахнуть	Вы	ахните

## 2. Омонимия форм глаголов близкой семантики

В таблице приведены примеры (случаи) омонимии в производных формах глаголов звучания: *звенеть* и *звонить* --> *позванивать, позванивай, позванивайте*; *хрустать* и *хрустеть* --> *похрустывать, похрустывай, похрустывайте*.

Пример 4. Омонимия форм префиксально-суффиксальных модификаций

<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>F</b>
<b>Number Номер</b>	<b>Root Основное слово</b>	<b>Derivative дериват</b>	<b>Pronoun Местоимение</b>	<b>Future tense Будущее время</b>
45.4.1.0	<b>звенеть</b>	<b>позванивать</b>	--	<b>позванивать</b>
45.4.2.0	звенеть	позванивать	Ты	позванивай
45.4.5.0	звенеть	позванивать	Вы	позванивайте
46.11.0.0	<b>звонить</b>	<b>позванивать</b>	--	<b>позванивать</b>
46.11.2.0	звонить	позванивать	Ты	позванивай
46.11.5.0	звонить	позванивать	Вы	позванивайте
127.2.0.0	<b>хрустать</b>	<b>похрустывать</b>	--	<b>похрустывать</b>
127.2.2.0	хрустать	похрустывать	Ты	похрустывай
127.2.5.0	хрустать	похрустывать	Вы	похрустывайте
128.3.0.0	<b>хрустеть</b>	<b>похрустывать</b>	--	<b>похрустывать</b>
128.3.2.0	хрустеть	похрустывать	Ты	похрустывай

128.3.5.0	хрустеть	похрустывать	Вы	похрустывайте
-----------	----------	--------------	----	---------------

Омонимия производных форм в русском языке может появляться по разным причинам.

Так, например, неразличимость форм у *звонить* и *звенеть* (*позванивай, позванивайте*) связана с историческим чередованием **О** и **Е** в корне исходных слов *звонѣ* и *звънѣти*:

1. ЗВОН: старосл. звънѣти (*болг. звеню*) --> **звенеть** [звени, звените]

2. ЗВОН: старослав. звонѣ (*болг. звънец*) --> **звонить** [звони, звоните]

Одно и то же слово *звон* послужило основой для образования глаголов близкой семантики *звенеть* и *звонить*. С течением времени контекстные значения *звенеть* и *звонить* существенно разошлись.

Глагол *звонить* в современном языке употребляется преимущественно в значениях: а) издавать звон; б) подавать сигнал; в) звонить по телефону.

Глагол *звенеть* используется только в ситуациях описания звуков металлического тембра: *звенит колокольчик, звенеть шпорами, звенит топор, пули звонят, голос звенит* и т.п.

При этом употребление форм глагола *звонить* в тех же контекстах исключено. Нельзя сказать *\*звонит колокольчик, \*звонить шпорами, \*звонит топор, пули \*звонят, \*голос звонит*.

Аналогичные расхождения в сочетаемости существуют и у пары глаголов близкой (но не одинаковой) семантики - ХРУСТЕТЬ и ХРУСТАТЬ.

Контекстная омонимия производных форм таких глаголов для носителей языка не является препятствием к пониманию текста: *похрустывает снег* или *похрустывает конь, мальчик похрустывает яблоком* – одинаково понятные контексты, так как глаголы *хрустеть* и *хрустать* описывают совершенно различные ситуации. Омонимия форм в подобных случаях для владеющих языком перестает быть релевантной.

Однако омонимию следует учитывать при фиксации искомых форм производных глаголов, где требуется детальное обозначение и маркирование всех элементов системы включая дублиеты, варианты и омонимы.

С этой целью в разрабатываемых таблицах введена четырехступенчатая (четырёхпозиционная) нумерация элементов, где учитываются не только все модификации исходных глаголов, но и все возможные формы дериватов.

В нашем тексте слова четырехступенчатая и четырёхпозиционная используются как дублиеты.

*Пример 4А.*

<b>127.2.0.0</b>	<b>хрустать</b>	<b>похрустывать</b>	--	<b>похрустывать</b>
127.2.2.0	хрустать	похрустывать	Ты	похрустывай
127.2.5.0	хрустать	похрустывать	Вы	похрустывайте

## **ЗАДАЧА II.**

На последующем этапе данного проекта была предпринята попытка создать наиболее полную, обобщающую сводную таблицу приставочных словоформ глаголов звучания -- будущего времени, повелительного наклонения, и их деепричастных форм.

Решение этой задачи планировали в два этапа:

16 ноября 2021	Обращение к эл. словарю <a href="https://morfologija.ru/">https://morfologija.ru/</a> , веб-извлечение <i>одноразовым поиском</i> форм будущего времени, повелительного наклонения и деепричастий модифицированных глаголов звучания. <a href="#">Приложение 4.</a>
21 февраля 2022	Обращение одновременно к эл. словарю <a href="https://morfologija.ru/">https://morfologija.ru/</a> и к НКРЯ <a href="https://ruscorpora.ru/new/index.html">https://ruscorpora.ru/new/index.html</a> ; веб-извлечение <i>одноразовым поиском</i> форм будущего времени, повелительного наклонения и деепричастий модифицированных глаголов звучания. <a href="#">Приложение 5.</a>

Анализ результатов, дальнейшая отладка программы поиска, оптимизация представления результатов.
---

**ОРГАНИЗАЦИЯ СВОДНОЙ ТАБЛИЦЫ 5 (Зализняк и НКРЯ). [Приложение 5.](#)**

**Колонка А** – порядковый номер слова

**Колонка В** – основное слово (центр модификационного гнезда)

**Колонка С** – модифицированный дериват (например, *ахать* - *ахнуть*)

**Колонка D** – личные местоимения, в традиционном порядке

**Колонка E** – словоформы настоящего времени (колонка не заполнена, нет доступа к источнику)

**Колонка F** – словоформы будущего времени

**Колонка G** - словоформы повелительного наклонения

**Колонка H** – словоформы деепричастий несовершенного вида (колонка заполнена с пробелами, нет доступа к источнику)

**Колонка I** – словоформы деепричастий совершенного вида (соответствует трёхступенчатый номер 1.0.0, 2.0.0 в колонке А)

Иллюстрация 1.

	A	B	C	D	E	F	G	H	I
1	number	root	word	pronoun	Настоящее время	Будущее время	Повелительное наклонение	деепричастие Несовершенный вид	деепричастие Совершенный вид
2	1.0.0	ахать	ахнуть	--	--	--	--	--	ахнув; ахнувши
3	1.0.0.1	ахать	ахнуть	--	--	--	--	--	ахнув
4	1.0.0.2	ахать	ахнуть	--	--	--	--	--	ахнувши
5	1.0.1.0	ахать	ахнуть	Я	--	ахну	--	--	--
6	1.0.2.0	ахать	ахнуть	Ты	--	ахнешь	ахни	--	--
7	1.0.3.0	ахать	ахнуть	Он/она	--	ахнет	--	--	--
8	1.0.4.1	ахать	ахнуть	Мы	--	ахнем	ахнем	--	--
9	1.0.4.2	ахать	ахнуть	Мы	--	ахнем	ахнемте	--	--
10	1.0.5.0	ахать	ахнуть	Вы	--	ахнете	ахните	--	--
11	1.0.6.0	ахать	ахнуть	Они	--	ахнут	--	--	--
12	1.1.0.0	ахать	ахать	--	--	--	--	--	--

Колонка В – основное слово (центр модификационного гнезда)

Колонка С – модифицированный дериват (например, *ахать* - *ахнуть*)

Иллюстрация 2.

	A	B	C	D	E	
29	1.3.1.0	ахать	заахать	Я	--	заахаю
30	1.3.2.0	ахать	заахать	Ты	--	заахаешь
31	1.3.3.0	ахать	заахать	Он/она	--	заахает
32	1.3.4.1	ахать	заахать	Мы	--	заахаем
33	1.3.4.2	ахать	заахать	Мы	--	заахаем
34	1.3.5.0	ахать	заахать	Вы	--	заахаете

**Система нумерации по позициям в колонке А.**

В колонке А имеются трёхступенчатые/трехпозиционные и четырёхступенчатые/четырёхпозиционные номера.

Позиция 1 – основное слово (немодифицированный глагол, колонка В)

Позиция 2 – модифицированный дериват (модифицированный глагол, колонка С)

Позиция 3 – словоизменительная парадигма (местоимения, колонка D)

### Иллюстрация 3.

A	B	C	D	E	F	G	H	I
1.1.0.0	ахать	ахать	--	--	--	--	--	--
1.1.1.0	ахать	ахнуть	Я	--	ахну	--	--	--
1.1.2.0	ахать	ахнуть	Ты	--	ахнешь	ахни	--	--
1.1.3.0	ахать	ахнуть	Он/она	--	ахнет	--	--	--
1.1.4.1	ахать	ахнуть	Мы	--	ахнем	ахнем	--	--
1.1.4.2	ахать	ахнуть	Мы	--	ахнем	ахнемте	--	--
1.1.5.0	ахать	ахнуть	Вы	--	ахнете	ахните	--	--
1.1.6.0	ахать	ахнуть	Они	--	ахнут	--	--	--
1.2.0.0	ахать	ахать	--	--	--	--	--	--
1.2.1.0	ахать	взахать	Я	--	ахну	--	--	--
1.2.2.0	ахать	взахать	Ты	--	ахнешь	ахни	--	--
1.2.3.0	ахать	взахать	Он/она	--	ахнет	--	--	--
1.2.4.1	ахать	взахать	Мы	--	ахнем	ахнем	--	--
1.2.4.2	ахать	взахать	Мы	--	ахнем	ахнемте	--	--
1.2.5.0	ахать	взахать	Вы	--	ахнете	ахните	--	--

Позиция 4 – дубликаты, омонимия (представлены как 1.3.4.1 (F) и 1.3.4.2 (G)). Например:



Иллюстрация 4.

	A	B	C	D	E	F	G	H	I	J	K
28	1.3.0.0	ахать	ахать	--	--	--	--	--	--		
29	1.3.1.0	ахать	заахать	Я	--	заахаю	--	--	--		
30	1.3.2.0	ахать	заахать	Ты	--	заахаешь	заахай	--	--		
31	1.3.3.0	ахать	заахать	Он/она	--	заахаёт	--	--	--		
32	1.3.4.1	ахать	заахать	Мы	--	заахаем	заахаем	--	--		
33	1.3.4.2	ахать	заахать	Мы	--	заахаем	заахаемте	--	--		
34	1.3.5.0	ахать	заахать	Вы	--	заахаёте	заахайте	--	--		
35	1.3.6.0	ахать	заахать	Они	--	заахают	--	--	--		
36	1.4.0.0	ахать	ахать	--	--	--	--	--	--		
37	1.4.1.0	ахать	поахать	Я	--	поахаю	--	--	--		

**ПРИМЕРЫ СООТВЕТСТВИЙ номеров и колонок.**

Трёхпозиционный номер 1.0.0, 2.0.0, 3.0.0 и т.д., соответствует колонке I - словоформы деепричастий совершенного вида, например: *1.0.0* *ахнув, ахнувши*.

Иллюстрация 5.

	A	B	C	D	E	F	G	H	I	J	K
1	number	root	word	pronoun	Настоящее время	Будущее время	Повелительное наклонение	деепричастие Несовершенный вид	деепричастие Совершенный вид		
2	1.0.0	ахать	ахнуть	--	--	--	--	--	ахнув; ахнувши		

Четырёхпозиционный номер 1.0.0.1, 2.0.0.1 и т.д., соответствует колонке I - словоформа деепричастий совершенного вида с суффиксом -В, например: *1.0.0.1* *ахнув*.

Иллюстрация 6.

	A	B	C	D	E	F	G	H	I	J
1	number	root	word	pronoun	Настоящее время	Будущее время	Повелительное наклонение	деепричастие Несовершенный вид	деепричастие Совершенный вид	
2	1.0.0	ахать	ахнуть	--	--	--	--	--	ахнув; ахнувши	
3	1.0.0.1	ахать	ахнуть	--	--	--	--	--	ахнув	

Четырёхпозиционный номер 1.0.0.2, 2.0.0.2 и т.д., соответствует колонке I - словоформа деепричастий совершенного вида с суффиксом -ВШИ, например: *1.0.0.1 ахнувши*.

Иллюстрация 7.

	A	B	C	D	E	F	G	H	I	J
1	number	root	word	pronoun	Настоящее время	Будущее время	Повелительное наклонение	деепричастие Несовершенный вид	деепричастие Совершенный вид	
2	1.0.0	ахать	ахнуть	--	--	--	--	--	ахнув; ахнувши	
3	1.0.0.1	ахать	ахнуть	--	--	--	--	--	ахнув	
4	1.0.0.2	ахать	ахнуть	--	--	--	--	--	ахнувши	
5	1.0.1.0	ахать	ахнуть	я	--	ахну	--	--	--	

Четырёхпозиционные номера 1.0.1.0, 1.0.2.0, 1.0.3.0, 1.0.4.0, 1.0.5.0, 1.0.6.0 связан с результатами поиска в колонке F, где номер третьей позиции соответствует определённому местоимению:

Иллюстрация 8.

	A	B	C	D	E	F	G	H	I	J	K
1	number	root	word	pronoun	Настоящее время	Будущее время	Повелительное наклонение	деепричастие Несовершенный вид	деепричастие Совершенный вид		
2	1.0.0	ахать	ахнуть	--	--	--	--	--	ахнув; ахнувши		
3	1.0.0.1	ахать	ахнуть	--	--	--	--	--	ахнув		
4	1.0.0.2	ахать	ахнуть	--	--	--	--	--	ахнувши		
5	1.0.1.0	ахать	ахнуть	Я	--	ахну	--	--	--		
6	1.0.2.0	ахать	ахнуть	Ты	--	ахнешь	ахни	--	--		
7	1.0.3.0	ахать	ахнуть	Он/она	--	ахнет	--	--	--		
8	1.0.4.1	ахать	ахнуть	Мы	--	ахнем	ахнем	--	--		
9	1.0.4.2	ахать	ахнуть	Мы	--	ахнем	ахнемте	--	--		
10	1.0.5.0	ахать	ахнуть	Вы	--	ахнете	ахните	--	--		
11	1.0.6.0	ахать	ахнуть	Они	--	ахнут	--	--	--		

Четырёхпозиционный номер 1.1.0.0, 1.2.0.0, 1.3.0.0 используются как разделители между группами словоформ.

Иллюстрация 9.

A44    X    ✓    fx    1.5.0.0

	A	B	C	D	E	F	G	H	I
20	1.2.0.0	ахать	ахать	--	--	--	--	--	--
21	1.2.1.0	ахать	взахать	Я	--	ахну	--	--	--
22	1.2.2.0	ахать	взахать	Ты	--	ахнешь	ахни	--	--
23	1.2.3.0	ахать	взахать	Он/она	--	ахнет	--	--	--
24	1.2.4.1	ахать	взахать	Мы	--	ахнем	ахнем	--	--
25	1.2.4.2	ахать	взахать	Мы	--	ахнем	ахнемте	--	--
26	1.2.5.0	ахать	взахать	Вы	--	ахнете	ахните	--	--
27	1.2.6.0	ахать	взахать	Они	--	ахнут	--	--	--
28	1.3.0.0	ахать	ахать	--	--	--	--	--	--
29	1.3.1.0	ахать	заахать	Я	--	заахаю	--	--	--
30	1.3.2.0	ахать	заахать	Ты	--	заахаешь	заахай	--	--
31	1.3.3.0	ахать	заахать	Он/она	--	заахаёт	--	--	--
32	1.3.4.1	ахать	заахать	Мы	--	заахаем	заахаем	--	--
33	1.3.4.2	ахать	заахать	Мы	--	заахаем	заахаемте	--	--
34	1.3.5.0	ахать	заахать	Вы	--	заахаёте	заахайте	--	--
35	1.3.6.0	ахать	заахать	Они	--	заахают	--	--	--
36	1.4.0.0	ахать	ахать	--	--	--	--	--	--
37	1.4.1.0	ахать	поахать	Я	--	поахаю	--	--	--

Четырёхступенчатый номер 1.1.1.0, 1.2.1.0, 1.3.1.0 используются для обозначения словоформы первого лица единственного числа модификационных дериватов, образованных от основного слова.

Иллюстрация 10.

	A	B	C	D	E	Formula Bar	G	H	I	J
45	1.5.1.0	ахать	поахивать	Я	--		поахаю	--		
46	1.5.2.0	ахать	поахивать	Ты	--		поахаешь	поахай		
47	1.5.3.0	ахать	поахивать	Он/она	--		поахает	--		
48	1.5.4.1	ахать	поахивать	Мы	--		поахаем	поахаем		
49	1.5.4.2	ахать	поахивать	Мы	--		поахаем	поахаемте		
50	1.5.5.0	ахать	поахивать	Вы	--		поахаете	поахайте		
51	1.5.6.0	ахать	поахивать	Они	--		поахают	--		
52	1.6.0.0	ахать	ахать	--	--		--	--		
53	1.6.1.0	ахать	проахать	Я	--		проахаю	--		
54	1.6.2.0	ахать	проахать	Ты	--		проахаешь	проахай		

### ЗАДАЧА III.

Создать алгоритм поиска (прогонки) каждой словоформы, зарегистрированной словаре Зализняка и представленной в [сводной таблице 5](#), с присоединением результатов поиска в базе данных Национального корпуса русского языка; далее создать интерактивную сноску для каждой словоформы (многомерное шкалирование результатов веб-извлечения).

Таким образом, в сводной таблице каждая словоформа *будущего времени, повелительного наклонения*, а также *деепричастные* формы глаголов звучания сопровождаются **электронной сноской** на результаты поиска в НКРЯ. Исследователь нажимает на словоформу мышкой и попадает на соответствующую страницу поиска в [Национальном корпусе русского языка](#).

### Пример 5.

1.0.2.0	<b>ахнешь</b>
	<a href="http://search.ruscorpora.ru/search.xml?mycorp=&amp;mysent=&amp;mysize=1000&amp;dpp=1000&amp;spp=1000&amp;spd=10&amp;t=1000&amp;text=lexform&amp;mode=main&amp;sort=gr_tagging&amp;lang=en&amp;r">http://search.ruscorpora.ru/search.xml?mycorp=&amp;mysent=&amp;mysize=1000&amp;dpp=1000&amp;spp=1000&amp;spd=10&amp;t=1000&amp;text=lexform&amp;mode=main&amp;sort=gr_tagging&amp;lang=en&amp;r</a> Найдено 77 документов, 85 контекстных употреблений. Found 77 documents, 85 contexts.

Поскольку НКРЯ является закрытой системой, изменений статистических результатов поиска не предполагается. Трудностей доступа при помощи компьютерной программы (веб-извлечения) к Национальному корпусу русского языка *не обнаружено*.

## **Примечание 2. Алгоритм поиска, принципы веб-извлечения, компьютерная программа (код).**

За подробностями о компьютерной программе (web-scraping algorithm) и по вопросам методологического характера просьба обращаться непосредственно к научно-исследовательской группе:

Ирина Валентиновна Ивлиева, [ivliyeva@mst.edu](mailto:ivliyeva@mst.edu), профессор кафедры искусств, иностранных языков и философии  
Перри Куб, [koobp@mst.edu](mailto:koobp@mst.edu), эксперт-администратор баз данных

## **РЕСУРСЫ. RESOURCES.**

Ивлиева И.В. Потенциальные единицы глагольного синтеза: свойства и возможности (на материале русских глаголов звучания) // Интерактивная наука. – 2021. – № 5 (60). – С. 11-19. – ISSN 2414-9411. [Electronic resource]. - Access mode:

[https://interactive-science.media/ru/article/554295/discussion\\_platform](https://interactive-science.media/ru/article/554295/discussion_platform)

Ivliyeva, I., Koob, Perry. "Method of web-extraction (web scraping) of Russian verb paradigms from electronic dictionaries and databases. Matrix organization of lacunae, their codification and classification (on the material of the verbs of sound)" (2021). [Electronic resource]. Research Data. 7. – Access mode: [https://scholarsmine.mst.edu/research\\_data/7](https://scholarsmine.mst.edu/research_data/7)

Ivliyeva, Irina V., Koob, Perry. "Lacunae Matrices, Web Scraping. Theses." (2021). Russian Linguistics Research. 1. [Electronic resource]. – Access mode: [https://scholarsmine.mst.edu/russian\\_linguistics\\_research/1](https://scholarsmine.mst.edu/russian_linguistics_research/1)

Koob, Perry B., and Ivliyeva, Irina V. "Russian National Corpus Web Scraping Project 2019-2020." [Electronic resource]. February 1, 2020. – Access mode: [https://scholarsmine.mst.edu/artlan\\_phil\\_facwork/158/](https://scholarsmine.mst.edu/artlan_phil_facwork/158/)

Ивлиева И.В. Особенности семантики лакун-модификаций. К идее создания словаря лакун (на материале глаголов звучания) // Интерактивная наука. – 2020. – № 6 (52). – С. 57-70. – ISSN 2414-9411. [Electronic resource]. - Access mode: [https://interactive-science.media/ru/article/551671/discussion\\_platform](https://interactive-science.media/ru/article/551671/discussion_platform)

Ивлиева И.В. Роль имплицитных х-компонентов значения в синтезе русских глагольных модификаций (на примере глаголов механической фонации) // Интерактивная наука. 2019. – № 6 (40). С. 106-113. – ISSN 2414-9411. Access mode: [https://interactive-science.media/en/article/497213/discussion\\_platform](https://interactive-science.media/en/article/497213/discussion_platform)

### **ЭЛЕКТРОННЫЕ ЛЕКСИКОГРАФИЧЕСКИЕ РЕСУРСЫ**

Грамматический словарь А.А. Зализняка. The grammatical dictionary of the Russian language A.A. Zalizniak. Access mode: <https://morfologija.ru/>

Национальный корпус русского языка. The Russian National Corpus. Access mode: <https://ruscorpora.ru/new/index.html>

Национальный корпус русского языка (параллельный перевод, 02.21.2022). Access mode: <https://ruscorpora.ru/new/search-multi.html>