

**Труды  
Института русского языка  
им. В. В. Виноградова**

**XXI**

**Национальный корпус русского языка:  
исследования и разработки**

МОСКВА  
2019

**Труды Института русского языка им. В.В. Виноградова. Вып. 21.** Национальный корпус русского языка: исследования и разработки. — М., 2019. — 328 с.  
ISSN 2311–150X

Издание основано в 2013 г.

Главный редактор А. М. Молдаван

**РЕДАКЦИОННЫЙ СОВЕТ**

- А. Е. Аникин, д. ф. н., академик РАН (Новосибирск);  
Ю. Д. Апресян, академик РАН, профессор (Москва, Россия);  
С. Вакарелийска, PhD, профессор (Орегон, США);  
Ж.Ж. Варбот, доктор филол. наук, профессор (Москва, Россия).  
Бьёрн Вимер, доктор филологии, профессор (Майнц, Германия);  
М. Гардзанити, PhD, профессор (Флоренция, Италия);  
А. А. Гиппиус, член-корреспондент РАН, профессор (Москва, Россия);  
О. Йокояма, PhD, профессор (Лос-Анджелес, США);  
Я. Какридис, Dr. habil., профессор (Берн, Швейцария);  
М. Л. Каленчук, доктор филол. наук, профессор (Москва, Россия);  
А. П. Майоров, д. ф. н., профессор (Улан-Удэ, Россия);  
О. Младенова, PhD, профессор (Калгари, Канада);  
Туре Нессет, доктор филологии, профессор (Тромсё, Норвегия);  
В. А. Плунгян, академик РАН, профессор (Москва, Россия);  
А. Рабус, Dr. habil., профессор (Фрайбург, Германия);  
Ф. Б. Успенский, д. ф. н., член-корреспондент РАН, профессор (Москва, Россия);  
М. Флайер, PhD, профессор (Кембридж, США);  
Вацлав Чермак, доктор филологии (Прага, Чехия);  
А. Д. Шмелев, доктор филол. наук, профессор (Москва, Россия).

**ОТВЕТСТВЕННЫЙ РЕДАКТОР ВЫПУСКА**

В. А. Плунгян, академик РАН, профессор (Москва, Россия).

**РЕДАКЦИОННАЯ КОЛЛЕГИЯ ВЫПУСКА**

Е. В. Рахилина, д. филол. наук (Москва, Россия)  
С. О. Савчук, к. филол. наук (Москва, Россия).

Выходит 4 раза в год

Адрес редакции:

119019, Москва, ул. Волхонка, д. 18/2

E-mail: [ruslang@ruslang.ru](mailto:ruslang@ruslang.ru)

Издательство зарегистрировано Федеральной службой по надзору  
в сфере связи, информационных технологий и массовых коммуникаций.

Свидетельство о регистрации ПИ № ФС 77-57258

© Институт русского языка им. В. В. Виноградова РАН, 2019  
© Авторы, 2019

## СОДЕРЖАНИЕ

Предисловие (В. А. Плулгян) . . . . .	11
---------------------------------------	----

### ***I. Развитие Национального корпуса русского языка и других корпусов***

Е. С. Иншакова, Л. Л. Иомдин, Л. Г. Митюшин, В. Г. Сизов, Т. И. Фролова, Л. Л. Цинман (Москва) СинТагРус сегодня . . . . .	14
Д. В. Сичинава (Москва) Параллельные тексты в составе Национального корпуса русского языка: новые языки и новые задачи . . . . .	41
Б. В. Орехов, С. О. Савчук (Москва) Акцентологический корпус как инструмент для исследования русского ударения . . . . .	61
А. Е. Поляков (Москва) Орфография Острожской Библии в контексте церковнославянского корпуса . . . . .	83
Н. В. Богданова-Бегларян, О. В. Блинова, Г. Я. Мартыненко, Т. Ю. Шерстинова (Санкт-Петербург) Корпус русского языка повседневного общения «Один речевой день» (ОРД): текущее состояние и перспективы . . . . .	100
Н. В. Богданова-Бегларян, О. В. Блинова, К. Д. Зайдес, Т. Ю. Шерстинова (Санкт-Петербург) Корпус «Сбалансированная аннотированная текстотека» (САТ): изучение специфики русской монологической речи . . . . .	110

### ***II. Национальный корпус русского языка в научных исследованиях***

Г. И. Кустова (Москва) Относительные наречия в контексте глаголов и прилагательных (по данным Национального корпуса русского языка) . . . . .	126
С. Ю. Жукова, Б. В. Орехов, Е. В. Рахилина (Москва) Дискурсивные формулы русского языка: диахронический подход . . . . .	142
В. П. Захаров, А. Ц. Масевич (Санкт-Петербург) Категоризация прилагательных цвета в русских поэтических текстах (корпусное исследование) . . . . .	164
К. М. Корчагин (Москва) Цезура в русском 5-стопном ямбе и ее роль в ритмической эволюции этого размера . . . . .	181
Д. О. Добровольский (Москва) «Показатели безразличия» в русско-немецком корпусе параллельных текстов . . . . .	202

### ***III. Слово и жест: вопросы мультимодальной коммуникации***

Т. Е. Янко (Москва) Просодия и жестикация . . . . .	213
О. В. Федорова (Москва) О коммуникативной функции взгляда . . . . .	222

**<sup>1</sup>Б.В. Орехов, <sup>2</sup>С.О. Савчук**  
<sup>1</sup>НИУ «Высшая школа экономики»,  
<sup>2</sup>Институт русского языка им. В. В. Виноградова РАН  
(Россия, Москва)  
<sup>1</sup>nevmenandr@gmail.com, <sup>2</sup>savsvelana@mail.ru

## АКЦЕНТОЛОГИЧЕСКИЙ КОРПУС КАК ИНСТРУМЕНТ ДЛЯ ИССЛЕДОВАНИЯ РУССКОГО УДАРЕНИЯ\*

В настоящей статье рассмотрено несколько вопросов, связанных с разработкой и использованием акцентологического корпуса в качестве инструмента для исследования ударения: состав и структура корпуса, текущее состояние, перспективы развития, пополнение новым материалом. Особое внимание уделено подкорпусу наивной поэзии в составе акцентологического корпуса как источнику акцентологических данных. Возможности этого ресурса, его эффективное использование проверены на нескольких участках акцентологической системы.

Было проведено корпусное исследование акцентных вариантов форм единственного и множественного числа кратких прилагательных, а также падежных форм имен существительных. В ходе изучения падежных форм существительных было обследовано несколько зон активной конкуренции акцентных вариантов: существительных женского рода на *-а* (*стена, доска*), существительных женского рода с основой на мягкий согласный (*кисть, челюсть*) и существительных мужского рода с твердой основой (*ветер, шарф*).

Результаты корпусного исследования подтвердили предположение о том, что материал акцентологических корпусов может быть использован как достоверный источник получения акцентологических данных. Увеличение объема корпуса делает эти данные статистически достоверными, а также расширяет круг исследуемых форм и способствует обнаружению новых точек вариативности.

*Ключевые слова.* Акцентологический корпус, Национальный корпус русского языка, «наивная поэзия», русское словесное ударение, акцентные варианты.

Русское ударение, его сложная система и состояние неустойчивого равновесия привлекают внимание как отечественных, так и зарубежных лингвистов

---

\* Работа выполнена при поддержке РФФИ (грант № 17-29-09154).

к изучению его истории и тех процессов, которые наблюдаются в области словоизменения и словообразования. Обычно изучение основывается на использовании источников материала, среди которых наиболее известными являются:

— данные словарей и справочников, содержащие сведения об ударении в разные исторические периоды и на разных территориях [Грам.; Еськова 1967; Зализняк 1985];

— акцентуированные тексты — памятники письменности [Зализняк 1985], а также учебные тексты;

— поэтические тексты [Воронцова 1979];

— транскрипты звучащей речи с проставленными ударениями;

— экспериментальный материал разного рода, полученный в ходе направленных интервью, анкетирования, опросов и др. [Куракина 2011; Marklund Sharapova 2000; Ukiiah 2002; Каленчук и др. 2017];

— корпуса текстов [Савчук 2010; Гришина 2012; Пожарицкая и др. 2017].

Можно предположить, что объединение разных источников в один корпус текстов многократно повышает их информативность и производительность. Идея создания корпуса для изучения русского ударения в составе НКРЯ, принадлежащая Е. А. Гришиной, была реализована в 2008 году, когда Акцентологический корпус был открыт для всеобщего доступа. За десять лет существования корпус многократно увеличился в объеме и продолжает пополняться. В настоящей статье будет рассмотрено несколько вопросов, связанных с разработкой и использованием акцентологического корпуса в качестве инструмента для исследования ударения: текущее состояние, состав и структура корпуса, перспективы развития, пополнение новым материалом. Особое внимание будет уделено подкорпусу наивной поэзии в составе акцентологического корпуса как источнику акцентологических данных. Возможности этого ресурса, его эффективное использование будут проверены на нескольких участках акцентологической системы.

## 1. Акцентологический корпус: состав и структура, перспективы роста

Акцентологический корпус первоначально разрабатывался как инструмент для изучения истории русского ударения [Гришина 2009]. Его основу составили поэтические тексты с размеченными долями (иктами), которые в подавляющем большинстве случаев позволяли вычислить место ударения в слове<sup>1</sup>.

---

<sup>1</sup> Правила вычисления места ударения по стихотворной разметке подробно описаны в [Гришина 2009, 153-155], там же рассмотрены случаи, когда однозначно определить место ударения не помогает ни метрическая организованность, ни рифма [там же, 155-157]. Как показывает практика применения акцентологического корпуса, таких контекстов в целом немного, не более 10%. Стихovedы утверждают, что «в размерах русской силлаботоники нет таких слабых позиций, которые вообще не могли бы быть заняты ударными слогами, и нет таких сильных позиций, которые не могли бы быть заняты безударными слогами» [Илюшин 2004, 71], но разработанная Е. А. Гришиной методика акцентологической работы позволяет получать нужные результаты.

Поэтическая часть акцентологического корпуса позволяет изучать изменения, которые произошли в системе ударения на протяжении XVIII-XXI вв. Однако при использовании корпуса следует иметь в виду, что информация об этих изменениях извлекается из текстов, относящихся к особому роду словесности — поэзии, которая обладает специфическими свойствами как в отношении лексического состава, так и в отношении к языковым нормам, в частности, акцентологическим. В научной литературе неоднократно отмечалось, что пользуясь поэтическим материалом, необходимо учитывать его «особость» — «большую вариантность, возможность использования ударения в стилистических целях, намеренную в некоторых случаях его архаизацию и др.» [Воронцова 1979, 6]. Поэтому дальнейшее развитие акцентологического корпуса требовало расширения его текстового состава за счет включения в него устных текстов, функционирующих в различных речевых сферах, что и стало задачей создания прозаической части подкорпуса [Savchuk 2009, Гришина и др. 2009].

Прозаический подкорпус содержит транскрипты устной речи, в которых представлены ударения в соответствии с реальным произношением и которые представляют устную русскую речь в ее функциональных разновидностях. В подкорпусе собраны образцы спонтанной бытовой речи, публичной устной речи разной степени спонтанности (устной научной речи, устной политической речи, радио- и телепублицистики, современной деловой речи, религиозных наставлений и др.). Значительную долю материала составляют транскрипты кинофильмов, театральных постановок, образцы художественного и авторского чтения. Прозаическая часть корпуса предназначена для изучения современного состояния акцентологических норм и их варьирования в зависимости от сфер устной речи.

В настоящее время объем текстов акцентологического корпуса, доступных на сайте [ruscogroa.ru](http://ruscogroa.ru), составляет более 17 млн словоупотреблений (без учета пилотного корпуса наивной поэзии, о котором речь пойдет дальше). Распределение текстов по двум зонам представлено в таблице 1.

Примерное распределение по периодам представлено в таблице 2.

Ценность акцентологического корпуса состоит в предоставляемой исследователю возможности получения ответов на как можно большее число вопросов.

Таблица 1

**Распределение текстов акцентологического корпуса по функциональным разновидностям**

Зона		Объем, млн	%
Поэзия		11	63 %
Проза		6,4	37 %
В т. ч.	Речь кино и театра	3,5	20 %
	Публичная речь	2,4	14 %
	Непубличная речь	0,4	2 %
	Художественное и авторское чтение	0,1	0,5 %

Таблица 2

**Распределение текстов акцентологического корпуса по периодам**

Период	Поэзия		Проза	
	Объем, млн	%	Объем, млн	%
1700–1800	0,92	8,1 %	–	–
1801–1910	4,47	39,4 %	0,02	0,3 %
1911–1950	4,43	39,2 %	0,4	6,2 %
1951–1980	1,26	11,1 %	2,14	33,1 %
1981–2000	0,24	2,1 %	1	15,4 %
2001–2017	0,02	0,1 %	2,9	45 %

Причем ответов не единичных, а статистически значимых, позволяющих строить гипотезы о тенденциях в развитии акцентологических норм. Поэтому ответ на главный вопрос о том, в каком направлении следует развивать акцентологический корпус, напрашивается сам собой: нужно увеличивать его объем, включая новые тексты. Но тогда возникает другой вопрос: какими именно текстами следует пополнять корпус? Дело в том, что специфика корпуса предъявляет определенные требования к отбору текстов. Во-первых, акцентологические изменения затрагивают далеко не каждое слово русского языка, большинство слов (по оценкам А. А. Зализняка, около 80% лексического состава) не меняют свой просодический профиль на протяжении столетий, сохраняя стабильное поведение при словоизменении и словообразовании: *август, дорога, жизнь, книга, кошка, нога, подруга, поле, привет, строитель, хвост, яблоко* и др. Поэтому было бы желательным, если бы корпус расширялся за счет ускоренного увеличения доли интересного в акцентологическом отношении материала, таких нестабильных участков, как глаголы с подвижным ударением, существительные 2-го склонения, заимствованные слова и др.

Во-вторых, технология создания акцентологического корпуса, особенно подготовка прозаических текстов, отличается большой трудоемкостью. Поэтому увеличение темпов и объемов пополнения корпуса ставит задачу снижения трудозатрат на подготовку текстов, причем не в ущерб качеству подготовки, прежде всего точности акцентологической разметки. Значительно сократить трудозатраты помогают специальные программы, позволяющие автоматизировать некоторые процессы подготовки, в частности, используемое при обработке транскриптов рабочее место лингвиста *Scripter* (автор Л. Д. Алексеевский). Еще больший эффект имело бы применение надежных программ автоматической акцентуации текстов, которые способны за короткое время разметить большие текстовые массивы. В последнее время на основе разных подходов активно ведутся разработки таких программ-акцентуаторов (см. [Зеленков и др. 2014]; [Ponomareva и др. 2017]; [Reynolds, Tyers 2015]; [Хомицевич и др. 2008]). Программа Ю. Зеленкова предназначена для разметки стихотворного текста, ее алгоритм позволяет обрабатывать русские словоформы с точностью 0.95–0.97, что вполне приемлемо для использования автоматически размеченного стихотворного материала в исследованиях ударения, поскольку в случае ошибочного предсказания программой места ударения есть возможность определить его, опираясь на метрическую разметку. Подробнее о принципах работы программы см. [Гришина и др. 2015, 260–263].

## **2. Подкорпус «наивной поэзии» как источник акцентологических данных**

Исходя из этих соображений было принято решение использовать в качестве источника для пополнения акцентологического корпуса коллекцию текстов «наивной поэзии», размещенной на сайте *stihi.ru*. Пилотный корпус «наивной поэзии» объемом около 13 млн словоупотреблений был сформирован и прошел апробацию, результаты которой изложены в работе [Гришина и др. 2015, 257–268].

Понятие наивной поэзии поначалу появилось в среде фольклористов [«Наивная литература» 2001]. В самом деле, культурный статус этих текстов сильно отличается от того, который закреплен за поэтической классикой. Если последняя имеет репутацию золотого фонда языка, непререкаемых образцов, тематически разнообразных и формально совершенных текстов, то тексты, которые обычно публикуются в Интернете непрофессиональными писателями, не выдерживают эстетической критики, построены формульно, однообразно и ограничены небольшим тематическим набором. Такое описание в гораздо большей степени подходит под шаблонно выстроенный фольклорный текст (с определенной оговоркой относительно эстетических достоинств фольклорных произведений, которые в любом случае второстепенны для системы устного народного творчества).

Типичность этих текстов позволяет оценить их количественно. В рамках концепции *distant reading* [Moretti 2013] перед нами открываются перспективы компьютерного анализа больших корпусов художественных текстов и литературоведческих обобщений. Такие оценки были сделаны и для наивной поэзии, размещенной авторами на сайте *stihi.ru* [Бонч-Осмоловская, Орехов 2014, 20–36]. Стихи «наивных» авторов в основном посвящены внутренним переживаниям, моментальным эмоциям и психологическому пейзажу.

Объемы текстов на ресурсе *stihi.ru* поистине циклопические: каждые 7 секунд на нем появляется новое стихотворение, поэтому размер корпуса *stihi.ru* превышает не только 100 млн словоупотреблений, которыми мы собираемся расширить акцентологический корпус, но и — в несколько раз — объем основного корпуса НКРЯ. Большинство авторов при этом формально стараются повторить известные им поэтические образцы и поэтому следуют силлабо-тоническому шаблону, что, в свою очередь, облегчает работу с акцентологическим материалом.

Тексты собраны с сайта *stihi.ru* в автоматическом режиме. С помощью программы Ю. Зеленкова для каждой строки был установлен метр и расставлены икты. В дальнейшем тексты, в которых для более чем 15% строк не удалось определить силлабо-тонический метр, отсеивались. Согласно правилам разметки акцентологического корпуса, в текстах была выделена зона рифмовки, которая дает дополнительную информацию о позиции ударения. Для достижения порога в 100 млн словоупотреблений потребовалось около 1 млн стихотворений «наивных» поэтов.

В настоящее время подготовлен и акцентологически размечен подкорпус «наивной» поэзии в объеме 100 млн словоупотреблений. Предпринимая предварительный анализ этого корпуса, мы ставили задачу получить ответы на следующие вопросы. Достаточен ли объем корпуса для принципиального повышения статистических показателей в акцентологически «проблемных» точках? Достаточно ли лексическое разнообразие текстов для того, чтобы охватить наибольшее количество «проблемных» точек, а не только увеличить его объем? Отражает ли корпус наивной поэзии живое русское произношение и позволяет ли проследить динамику изменений акцентологических предпочтений? Коррелируют ли данные, полученные в корпусных исследованиях, с результатами, полученными экспериментальным путем?



### 3. Анализ корпуса наивной поэзии

С целью проверки возможностей 100-миллионного корпуса было проведено исследование акцентных вариантов в области словоизменения существительных и прилагательных. Для исследования были отобраны лексемы существительных и прилагательных, которые относятся к проблемным точкам акцентологической системы, то есть при словоизменении имеют вариативное ударение во всех или в отдельных формах. Для каждой леммы был установлен состав вариантов и их частотные показатели в каждом из трех описанных ниже корпусов.

Первый корпус составили тексты основного акцентологического корпуса за период до 2000 года. Его объем — 14 468 529 словоупотреблений. В дальнейшем будем обозначать его **Акц\_2000**.

Второй корпус включает тексты после 2000 года, в основном это записи устной речи и пилотный корпус наивной поэзии. Объем корпуса — 17 262 123 словоупотреблений, условное наименование — **Акц\_2000+**.

Третий корпус составлен из текстов сайта stih1.ru, его объем около 100 млн словоупотреблений, условное наименование **Стихи.ру-100**.

По результатам анализа составлены базы данных, фрагменты которых показаны в таблицах 3 и 4<sup>2</sup>.

Таблица 3

Варианты форм кратких прилагательных

Лексема	Вариант словоформы	Акц_2000		Акц_2000+		Стихи.ру-100	
		Вариант 1	Вариант2	Вариант 1	Вариант2	Вариант 1	Вариант2
верный	ве'рны'	54	59	21	135	182	843
глупый	глу'пы'	41	5	37	98	499	302
строгий	стро'ги'	127	5	51	14	516	81
честны	че'стны'	12	8	9	32	103	250
быстрый	бы'стры'	81	11	18	10	164	125
жалкий	жа'лки'	61	1	19	1	147	26
властный	вла'стны'	83	8	78	2	536	4
вкусный	вку'сны'	14	4	8	2	90	59
вредный	вре'дны'	25	10	2	7	20	71
жадный	жа'дны'	23	2	3	0	51	13
ловкий	ло'вки'	8	11	2	4	76	36
тусклый	ту'склы'	21	4	5	0	88	11
храбрый	хра'бры'	22	5	2	5	16	24
звучный	звуч'ны'	21	4	5	1	15	2
душный	ду'шны'	6	0	2	0	14	0

<sup>2</sup> Во второй графе таблиц представлены акцентные варианты словоформ, в следующих графах вариант с ударением на основе обозначен как вариант 1, с ударением на окончании — как вариант 2. В некоторых трех- и четырехсложных формах возможны три варианта ударения (*бо'ро'зда'м*, *по'ло'са'х* и пр.), при этом обе формы с ударением на основе приводятся в графе варианта 1 через вертикальную черту.

Лексема	Вариант словоформы	Акц_2000		Акц_2000+		Стихи.ру-100	
		Вариант 1	Вариант2	Вариант 1	Вариант2	Вариант 1	Вариант2
пресный	пре'сны'	1	0	4	0	47	11
спелый	спе'лы'	8	1	0	0	23	1
знатный	зна'тны'	7	0	1	0	12	0
смуглый	сму'глы'	8	7	0	0	0	4

Таблица 4

## Акцентологические варианты падежных форм существительных

Лексема	Вариант словоформы	Акц_2000		Акц_2000+		Стихи.ру-100	
		Вариант 1	Вариант2	Вариант 1	Вариант2	Вариант 1	Вариант2
Стена	сте'ну'	182	27	113	5	3074	153
	сте'на'м	69	159	104	29	1142	245
	сте'на'ми	32	124	57	35	505	288
	сте'на'х	97	339	225	139	2048	1000
Волна	во'лна'м	52	229	83	184	687	1474
	во'лна'ми	52	410	107	172	1065	1322
	во'лна'х	76	385	107	210	1020	1468
Доска	до'ску'	59	14	66	10	417	122
	до'со'к	30	50	23	13	198	122
	до'ска'м	26	9	4	2	62	10
	до'ска'ми	10	4	6	3	72	23
	до'ска'х	26	16	9	7	50	59
Щека	щё'ку'	67	27	90	25	415	184
	щё'ка'м	6	86	14	469	35	2359
	щё'ка'ми	0	23	0	9	3	110
	щё'ка'х	5	207	7	286	23	1710
Полоса	по'лосу'	26	29	18	34	32	294
	по'ло'с	4	28	0	0	102	560
	по'лоса'м	1	1	1	3	3 2	32
	по'лоса'ми	3	41	2	12	1	0
	по'лоса'х	3	5	2	9	3 2	0
Борозда	бо'розду'	3	16	12	2	102	0
	бо'ро'зд	7	11	1	11	3	38
	бо'ро'зда'м	1	7	1	0	0 1	10
	бо'розда'ми	0	3	2	0	0	19
	бо'ро'зда'х	1 3	6	1	0	1 2	12
Сковорода	ско'вороду'	2	12	0	12	3	57
	ско'воро'д	1	4	0	0	0	4
	ско'ворода'м	0	0	0	0	0	0
	ско'ворода'ми	0	0	0	0	0	0
	ско'ворода'х	0	0	0	0	0	5
Изба	и'збу'	75	81	20	20	168	287
	и'зба'м	23	5	3	1	32	4

Лексема	Вариант словоформы	Акц_2000		Акц_2000+		Стихи.ру-100	
		Вариант 1	Вариант2	Вариант 1	Вариант2	Вариант 1	Вариант2
Изба	и'зба'ми	7	0	4	1	28	1
	и'зба'х	31	4	9	0	87	6
Челюсть	че'люсте'й	8	5	2	3	49	6
	че'люстя'м	0	1	0	0	9	0
	че'люстя'ми	3	7	0	8		
	че'люстя'х	6	3	0	1	12	1
Кисть	ки'сте'й	2	32	4	16	21	174
	ки'стя'м	0	2	1	1	4	15
	ки'стя'ми	3	42	4	10	40	206
	ки'стя'х	2	11	4	3	30	39
Ветер	ве'тры'/ветра'	513	109	310	480	3793	5324
	ве'тро'в	114	194	57	272	581	3641
	ве'тра'м	31	78	24	150	231	1797
	ве'тра'ми	29	115	33	163	468	2125
	ве'тра'х	4	36	2	55	36	828
Шарф	ша'рфа'	13	0	7	1	61	17
	ша'рфу'	1	0	2	0	6	3
	ша'рфо'м	17	1	21	16	192	89
	ша'рфе'	6	0	9	6	60	30
	ша'рфы'	14	3	8	7	68	60
	ша'рфо'в	2	0	1	0	10	19
	ша'рфа'м	0	0	0	0	2	1
	ша'рфа'ми	2	1	0	3	15	18
ша'рфа'х	3	0	0	1	12	11	

Первый вопрос, поставленный в исследовании, касался количественных параметров корпуса и состоял в выяснении того, насколько 100-миллионный корпус наивной поэзии обеспечивает пополнение акцентологического корпуса в «проблемных» точках. Подобный анализ проводился ранее на пилотном корпусе [Гришина и др. 2015, 263–268], в настоящей работе мы прежде всего сосредоточились на лексемах, которые не обнаружили заметного увеличения частоты в пилотном корпусе.

Как видно из таблицы 3, пополнение акцентологического корпуса текстами наивной поэзии обеспечивает существенный прирост материала, значимого для изучения акцентной вариативности. Так, например, в современном корпусе для вариантов *ло'вки* — *ловки'*, *че'стны* — *честны'*, *вку'сны* — *вкусны'*, *вре'дны* — *вредны'* не было достоверных данных, на основании которых можно было бы сделать вывод о соотношении вариантов, а 100-миллионный корпус предоставил десятки примеров, позволяющих определять это соотношение и строить гипотезы о конкуренции вариантов и о тенденциях развития вариативности.

Однако в отдельных случаях корпус Стихи.ру-100 не дал ожидаемого прироста вариантных форм. Подобный отрицательный результат для отдельных лексем был получен ранее и на материале пилотного корпуса: в частности, для форм *знатны*,

смуглы, спелы, тучны, серьгам не зафиксировано ни одного вхождения в исследуемом корпусе, формы храбры, жадны, душины, звучны, тусклы получили 1–2 новых вхождения [Гришина и др. 2015, 265–268]. Но тогда высказывалось предположение, что благодаря 100-млн корпусу десятикратное увеличение объема поможет преодолеть большую часть проблем с полнотой представления акцентологически значимых форм. Как выяснилось, это не совсем так. То же можно сказать и о вариантах форм существительных, представленных в таблице 4. Для многосложных форм существительных *сковорода*, *борозда*, *ведомость* и под. обращение к 100-млн корпусу не дает существенного увеличения данных, и количественные показатели находятся на уровне основного акцентологического корпуса. Возможно, это объясняется «непоэтичностью» подобных слов, которые нетипичны для стихотворного текста как в плане содержания, так и в плане выражения: в силу своей многосложности они с трудом вписываются в стихотворную строку. Видимо, исследовать редкие варианты придется другими методами, в частности, на основе экспериментов.

Содержательный анализ данных ставил своей целью выяснение возможностей корпуса для исследования вариативного ударения. Для этого в исследуемых корпусах был собран материал: в каждом подкорпусе для каждой лексемы определена акцентная парадигма (а.п.) и для каждой словоформы установлено количественное соотношение акцентологических вариантов. Результаты представлены в таблице 5<sup>3</sup>.

Таблица 5

**Количественное соотношение акцентологических вариантов падежных форм существительных**

Слово	а.п.	Варианты	Соотношение вариантов		
			Акц 2000	Акц 2000+	Стихи.ру-100
Стена	d//f	стену	87:13	96:4	95:5
		стенам	30:70	78:22	82:18
		стенами	20:80	62:38	64:36
		стенах	22:78	62:38	67:33
Волна	f//d	волну	2:98	0,8:99,2	–
		волнам	19:81	31:69	32:68
		волнами	11:89	38:62	45:55
		волнах	17:83	34:66	41:59
Доска	f//d	доску	80:20	87:13	77:23
		досок	38:62	64:36	62:38
		доскам	74:36	100:0	86:14
		досками	71:39	67:33	76:24
		досках	62:38	56:44	46:54

<sup>3</sup> Ударные гласные в вариантах словоформ в графе «Варианты» обозначены полужирным шрифтом. Соотношение вариантов складывается из отношения количества каждого из акцентных вариантов словоформы к общему количеству акцентуированных вхождений словоформы. В случае вариативного наосновного ударения соотношение с флексийными ударными формами рассчитано отдельно, и оба варианта приводятся в соответствующей графе через /. Соотношения вариантов, рассчитанные на основе единичных данных (см. табл. 4), обозначены !! Прочерк обозначает отсутствие данных.

Слово	а.п.	Варианты	Соотношение вариантов		
			Акц_2000	Акц_2000+	Стихи.ру-100
Щека	f//f	щёку	71:29	78:22	69:31
		щёками	7:93	3:97	1:99
		щёками	0:100	0:100	3:97
		щёках	2:98	2:98	1:99
Полоса	f//f	полосу	47:53	29:71	10:90
		полос	13:87	0	15:85
		полосам	50:50!	25:75!!	9:91/6:94
		полосами	7:93	8:92	100:0!!
Борозда	f//f	полосах	38:62!!	18:82!	100:0
		борозду	16:84	86:14	100:0
		борозд	39:61	8:92	7:93
		бороздам	12:88	100:0!	0:100/10:90
Сковорода	f//f	бороздами	0:100	100:0!	0:100
		бороздах	14:43:43	100:0!	8:92/14:86
		сковороду	14:86	0:100	5:95
		сковород	20:80	–	0:100
Изба	d//d'	сковородам	–	–	–
		сковородами	–	–	–
		сковородах	–	–	0:100
		избу	48:52	50:50	37:63
Ведомость	e//a	избам	82:28	75:25	89:11
		избами	100:0	80:20	97:3
		избах	89:11	100:0	94:6
		ведомостей	–	–	–
Челюсть	a//e	ведомостям	–	–	–
		ведомостями	–	–	–
		ведомостях	–	67:33	–
		челюстей	61:39	40:60	89:11
Кисть	e	челюстям	0:100	–	100:0
		челюстями	30:70	0:100	–
		челюстях	67:33	0:100!!	92:8
		кистей	6:94	20:80	11:89
Ветер	e//a//c	кистям	0:100!!	50:50!!	21:79
		кистями	7:93	29:71	16:84
		кистях	15:85	57:43!!	43:57
		ветры/ветра	82:18	39:61	42:58
Шарф	a//b	ветров	37:63	17:83	14:86
		ветрам	28:72	14:86	11:89
		ветрами	20:80	17:83	18:82
		ветрах	10:90	4:96	4:96
Шарф	a//b	шарфа	100:0	88:12	78:22
		шарфу	100:0	100:0	67:33
		шарфом	94:6	57:13	68:32
		шарфе	100:0	60:40	67:33
		шарфы	82:18	53:47	53:47
		шарфов	100:0	100:0	34:66
		шарфам	–	–	67:33
		шарфами	67:33!!	–	45:55
шарфах	100:0	0:100	52:48		

Задачей сравнительного анализа данных, представляющих соотношение вариантов в трех корпусах, было проверить, отражают ли корпусные данные реальное

соотношение вариантов в живой речи или в какой-то ее разновидности. Отражают ли полученные соотношения те процессы, которые происходят в акцентологической системе и описаны в лингвистической литературе?

Было обследовано несколько зон активной конкуренции акцентных вариантов падежных форм имен существительных: существительных женского рода на *-а*, существительных женского рода с основой на мягкий согласный и существительных мужского рода с твердой основой.

Напомним, что ударение в падежных формах имен существительных может быть описано с помощью 6 основных и 4 второстепенных схем (Грамм. 2003, 31–32)<sup>4</sup>.

Таблица 6

Основные схемы ударения в субстантивном склонении

Формы	a	b	c	d	e	f
Ед.ч.	■○	□●	■○	□●	■○	□●
И. мн.	■○	□●	□●	■○	■○	■○
Р. Д. Т. П. мн.	■○	□●	□●	■○	□●	□●
	<i>карта, спор</i>	<i>стол, очко</i>	<i>сад, море</i>	<i>лист, вино</i>	<i>зуб, вещь</i>	<i>губа, конь</i>

Таблица 7

Второстепенные схемы ударения в субстантивном склонении

Формы		b'		d'	f'	f''
И. Р. Д. П. ед.		□●		□●	□●	□●
В. ед.		□●		■○	■○	□●
Т. ед.		■○		□●	□●	■○
И. мн.		□●		■○	■○	■○
Р. Д. Т. П. мн.		□●		■○	□●	□●
		<i>вошь</i>		<i>спина</i>	<i>рука</i>	<i>грудь (f''/e)</i>

Для имен существительных активность акцентных вариантов определяется направлением эволюции к формированию акцентной оппозиции единственного и множественного числа: ед. *на'рус* — мн. *паруса'*, ед. *высота'* — мн. *высо'ты*. Исторически процесс складывается из двух компонентов: 1) акцентная поляризация И. ед. и И. мн.; 2) выравнивание каждой из двух подпарадигм (Зализняк 1985, 373).

У существительных женского рода на *-а* акцентная вариантность наблюдается в формах В. ед. (*сте'ну'*, *ре'ку'*) и в формах Д., Т., П. мн. (*сте'на'м*, *сте'на'ми*, *сте'на'х*)<sup>5</sup>.

Сравнение акцентных вариантов форм Д., Т., П. мн. по корпусу Акц\_2000 (напомним, что тексты представляют поэзию XVIII–XX вв., кино, театр и художественное

<sup>4</sup> В схемах ударная основа обозначена черным квадратом, безударная — белым, соответственно ударное окончание — черным кружком, безударное — белым.

<sup>5</sup> В Грамм. это колебание в выборе схемы ударения при словоизменении обозначается как принадлежность лексемы к одной из двух акцентных парадигм f//d, f''/d', f''/f, d''/d.

слово) и корпусу Акц\_2000+ (в основном это записи устной речи и наивная поэзия) показало, что соотношение вариантов изменилось более или менее существенно. Зависит ли степень изменений от лингвистических характеристик слов или экстралингвистических факторов, станет понятнее при рассмотрении конкретных лексем.

**Стена** (d'//f'). У существительного *стена* преобладание форм с ударением на основе (*сте'нам, сте'нами, сте'нах*, соответствующих схеме d') в корпусе Акц\_2000+ отражает рекомендацию Грам., в котором эта модель рассматривается как современная, а модель с ударением на окончаниях (*стена'м, стена'ми, стена'х*, соответствующая схеме f') — как устаревшая. Предпочтение этой устаревшей схемы в текстах корпуса Акц\_2000 объясняется высокой долей в нем поэзии XIX в. и высокой частотностью этого слова в поэтических текстах, отличающихся, как уже говорилось, бо́льшим консерватизмом, чем проза. Закрепление ударения на основе в формах Д., Т., П. мн. приводит к выравниванию внутри парадигмы множественного числа и к акцентному противопоставлению парадигм единственного и множественного числа. Корпус Стихи.ру-100, на порядок увеличивая количество вхождений, демонстрирует то же соотношение вариантов, что и современный корпус Акц\_2000+.

**Волна** (f//d). Соотношение в пользу вариантов Д., Т., П. мн. с флексийным ударением (*волна'м, волна'ми, волна'х*) сохранилось в корпусе современных текстов, однако доля вариантов с наосновным ударением (*во'лнам, во'лнами, во'лнах*) выросла в 2–3 раза, в чем можно усмотреть действие тенденции к выравниванию внутри парадигмы множественного числа. Интересно, что в 100-млн корпусе на фоне многократного увеличения абсолютной частоты вхождений словоформ Д., Т. и П. мн. увеличивается доля вариантов Т. и П. мн. с ударением на основе, но сохраняется неизменным (на уровне Акц\_2000+) соотношение вариантов Д. мн., — возможно, здесь проявляется сдерживающее влияние частотной формулы «по моря'м, по волна'м».

**Доска** (f//d). Все три корпуса показывают преобладание варианта с ударением на основе (*до'скам, до'сками, до'сках*) над вариантом с окончательным ударением (*доска'м, доска'ми, доска'х*), причем в количественном отношении это соотношение также стабильно. Таким образом, вопреки нормативным рекомендациям, для лексемы *доска* отдающим предпочтение схеме ударения f, в современном узусе, представленном в акцентологическом корпусе, приоритетной является схема ударения d'. Кроме того, следует отметить, что в корпусных данных зафиксировано и зеркальное изменение соотношения вариантов *до'сок* — *досо'к* в формах Р. мн.: 38% : 62% в Акц\_2000 и 64% : 36% в Акц\_2000+, 62% и 38% в Стихи.ру-100.

**Щека** (f'//f). В формах Д., Т., П. мн. во всех корпусах преобладание варианта с флексийным ударением *щека'м, щека'ми, щека'х* близко к 100%, что свидетельствует о закреплении в этих формах традиционного ударения<sup>6</sup>.

<sup>6</sup> Слово *щека* входит в семантическую группу названий частей тела (наряду с *голова, борода, нога, рука, спина, коса*), устойчиво сохраняющую схему ударения d' или f' (см. об этом [Зализняк

**Полоса, борозда, сковорода** (f/f'). Конкуренция вариантов в формах Д., Т., П. мн., видимо, еще продолжается и по-разному протекает в разных падежных формах. С одной стороны, в Д. и П. мн. выросла доля форм с конечным ударением (*полоса'м, полоса'х*), с другой стороны, увеличилась доля форм Т. мн. с наосновным ударением в соотношении *по'лосами* и *полоса'ми* — от 7% до 14%. То же можно сказать о формах *бо'роздах, бо'роздами, бо'роздах*, которые представлены в корпусе Акц\_2000+. В корпусе Стихи.ру-100 преобладают варианты с конечным ударением: *полоса'м, борозда'м, борозда'ми, борозда'х, сковорода'х*. Впрочем, эти лексемы имеют невысокую частоту как в современном акцентологическом корпусе, так и в корпусе наивной поэзии, так что для выводов по всем интересующим формам требуется прибегнуть к другим акцентологическим источникам.

В целом можно сказать, что поведение акцентных вариантов в рассматриваемых лексемах постепенно ведет к выравниванию ударения внутри парадигмы множественного числа и способствует вовлечению данных лексем в группу существительных со сформированной акцентной оппозицией чисел.

Более сложную картину дает исследование соотношения форм Вин. ед. рассматриваемых существительных женского рода.

С одной стороны, у существительных **волна, гряда, сковорода, полоса, изба** предпочтительной оказывается форма с конечным ударением: *волну', гряду', сковороду', полосу', избу'*. При этом в первых трех словах конечное ударение в формах Вин. ед. доминирует; форма *полосу'* показывает в современном корпусе Акц\_2000+ значительное перераспределение в свою пользу по отношению к варианту *по'лосу*, а в корпусе Стихи.ру-100 оно доминирует. Сохраняющееся колебание вариантов *и'збу* — *избу'* (примерно 50% : 50%, хотя в 100-млн корпусе все-таки преобладает вариант *избу'*) может объясняться тем, что ударение *и'збу* поддерживается употреблением в поэтических текстах (они составляют 55 из 75 всех вхождений формы в корпусе Акц\_2000), в том числе хрестоматийных: *Прибежали в и'збу дети, В торопях зовут отца; Дурачина ты, прямой простофиля! Выпросил, простофиля, и'збу!* (А. С. Пушкин); *Коня на скаку остановит, В горящую и'збу войдет!* (Н. А. Некрасов); *И'збу освещает Огонек светца; Зимний вечер длится, Длится без конца...* (И. З. Суриков); *Метель мысли путает, Метель в и'збу ломится.* (М. И. Цветаева). В современной живой речи слово практически вышло из употребления и усваивается, включая и фонетический облик, по известным литературным цитатам. Достаточно сказать, что четверть вхождений формы *и'збу* в корпусе наивной поэзии встречается в перифразах цитаты из поэмы Н. А. Некрасова «в горящую избу войдет».

Преобладание формы с флексийным ударением в Вин. ед. отражает тенденцию к выравниванию подпарадигмы единственного числа и акцентному противопоставлению единственного и множественного числа (Зализняк 1985, 374).

Наряду с этим у существительных **стена, доска, борозда, щека, борода** в конкуренции акцентных вариантов формы Вин. ед. побеждает вариант с наосновным

1985, 29], то есть в сохранении стабильности соотношения вариантов можно предположить влияние семантического фактора.



ударением: *сте'ну, до'ску, щё'ку, бо'роду, бо'розду*. При этом в корпусе Акц\_2000+ доля победившего варианта выше.<sup>7</sup>

В преобладании форм Вин. ед. с наосновным ударением проявляется локальная тенденция к акцентной оппозиции Им. ед. и Вин. ед. (*трава'* — *тра'ву*), которую А.А. Зализняк относит к более поздним инновациям [Зализняк 1985, 374]. Рост доли этих форм в корпусах Акц\_2000+ и Стихи.ру-100, отражающих по преимуществу речь молодого поколения, согласуется с этим положением.

Вместе с тем, оба разнонаправленных процесса, обуславливающих тяготение вариантов Вин. ед. существительных ж. р. на *-а* к одной из акцентологических схем (с ударением на основе или на окончании), является проявлением глобальной тенденции к грамматикализации ударения, что выражается в закреплении его за определенными грамматическими формами.

Еще одной зоной вариативности, возникающей под воздействием грамматического фактора, являются существительные ж. р. с основой на мягкий согласный. Тенденция к акцентному противопоставлению форм ед. и мн. числа (*ветвь* — *ве'тви, ве'тви* — *ветве'й*) проявляется в существовании вариантов с конечным ударением в формах Р., Д., Т., П. мн., которые конкурируют с вариантами, имеющими наосновное ударение, то есть схемы е//а: **ве'домость** — *ве'домосте'й, ве'домостя'м, ве'домостя'ми, ве'домостя'х*; **че'люсть** — *че'люсте'й, че'люстя'м, че'люстя'ми, че'люстя'х*; **кре'пость** — *кре'посте'й, кре'постя'м, кре'постя'ми, кре'постя'х*; **кисть** — *ки'сте'й, ки'стя'м, ки'стя'ми, ки'стя'х* и др.

Акцентная оппозиция чисел ограничена в этой группе морфонологическими правилами, по которым форма Им. мн. имеет ударение только на основе. Вследствие этого вступает в действие противоположная тенденция — акцентное выравнивание парадигмы множественного числа по форме Им. п., имеющей наосновное ударение [Зализняк 1985, 375]. Перечисленные лексемы являются малочастотными в обследованных корпусах, хотя для некоторых форм можно отметить наличие вариантов с флексийным ударением (*челюсте'й, челюстя'ми, челюстя'х*) как проявление тенденции к числовому противопоставлению, а для других — рост вариантов с наосновным ударением (*ки'стей, ки'стями, ки'стях*) как проявление противоположной тенденции.

В группе существительных мужского рода акцентная перестройка коснулась слов с наосновным ударением в единственном числе, в результате которой возникли варианты Им.-Вин. мн. (*но'рты* — *порты', ве'кторы* — *вектора'*) и варианты косвенных падежей мн. ч. (*ве'кторо'в, ве'ктора'м, ве'ктора'ми, ве'ктора'х*).

**Ветер** (е//а//с). В косвенных падежах мн. ч. во всех корпусах преобладают варианты с ударением на флексии, при этом в формах Р. мн. и Д. мн. доля флексийных вариантов выросла в современных корпусах. В соотношении вариантов Им. мн.

<sup>7</sup> Исключение составляет форма *бо'розду'*, для которой разные корпуса фиксируют зеркальное соотношение вариантов: в Акц\_2000 — 16:84% в пользу *борозду'*, в Акц\_2000+ — 86:14% в пользу *бо'розду*. Корпус Стихи.ру-100 дает безальтернативный ответ в пользу формы *бо'розду* (102 вхождения). Возможно, это связано с низкой частотностью этого слова в корпусе и в речи вообще, что требует проверки на альтернативном материале.

*ве'тры* — *ветра'* произошла смена доминанты: если в корпусе Акц\_2000 преобладал вариант с наосновным ударением, то в корпусах Акц\_2000+ и Стихи.ру-100 его место занял конкурирующий вариант *ветра'*. Все это отражает тенденцию к акцентному противопоставлению форм единственного и множественного числа в парадигме склонения.

Модель Им.-Вин. мн. с ударным окончанием *-а* является очень продуктивной, в нее вовлекаются новые существительные, что особенно наглядно на примере заимствованных слов.

**Шарф** (a/b). В корпусе Акц\_2000 преобладают формы с неподвижным ударением на основе (что соответствует а. п. а, приписанной этой лексеме в Грам.), варианты отдельных форм с флексийным ударением единичны (*шарфы'*, *шарфо'м*, *шарфа'ми*). Однако по данным корпуса Акц\_2000+ и особенно Стихи.ру-100, доля вариантов с ударением на окончании значительно выросла, хотя и неравномерно в разных падежных формах. Особенно заметен рост вариантов с флексийным ударением в формах единственного числа и в формах *шарфы'* (до 47%), *шарфо'в* (до 66%), *шарфа'ми* (до 55%).

Большую роль в появлении вариантов с конечным ударением играет прагматический фактор: у освоенных слов растет доля вариантов с подвижным ударением, в то время как неосвоенная лексика (книжные, редкие, новые слова) сохраняет тривиальные схемы словоизменения [Зализняк 1985, 376].

#### 4. Заключение и выводы

Подводя итоги анализа акцентных вариантов в области именного словоизменения, можно сказать, что результаты корпусного исследования подтверждают выводы, сформулированные в лингвистической литературе на основе изучения словарных материалов, акцентуированных источников, поэтических текстов. Следовательно, материал акцентологических корпусов может быть использован как достоверный источник получения акцентологических данных. Увеличение объема корпуса делает эти данные статистически достоверными, а также расширяет круг исследуемых форм и способствует обнаружению новых точек вариативности.

Надежность результатов, полученных с помощью корпусов, подтверждается их количественным совпадением с результатами экспериментальных исследований акцентных вариантов. В частности, описанные выше соотношения вариантов *до'ску* — *доску'*, *по'лосу* — *полосу'*, *ско'вороду* — *сковороду'*, *щё'ку* — *щеку'*, *сте'нам* — *стена'м*, *ша'рфы* — *шарфы'* совпадают с данными, приведенными в диссертационном исследовании Е.Б. Куракиной, посвященном изучению места ударения в вариативных акцентных формах в речи молодых москвичей и построенном на экспериментальном материале [Куракина 2011].

При этом корпус наивной поэзии, как показало проведенное исследование, отражает тенденции в современном состоянии русской акцентной системы: соотношение словоизменяемых вариантов в текстах этого корпуса в целом совпадает с результатами, полученными на материале современных текстов основного

акцентологического корпуса. Объем корпуса в 100 млн, видимо, следует признать достаточным для обеспечения статистической надежности результатов: количество большинства исследуемых вариантов выросло многократно, те же малочастотные лексемы, которые не обнаружили роста в 100-миллионном корпусе, вряд ли увеличат частоту при увеличении общего объема корпуса.

Однако идея корпуса наивной поэзии не всеми принимается безоговорочно. Возражения, в частности, вызывает сомнительный в содержательном плане материал. Далее, обращают внимание на специфичность материала: стиховая форма может накладывать ограничения на выбор лексики и акцентных вариантов. Можно предположить, что в корпусе будут слабо представлены или вообще отсутствовать «непоэтические» слова (в частности неологизмы, термины и под.) из интересующего акцентолога списка лексем и словоформ, что подтвердилось в процессе нашего исследования, когда ряд «непоэтических» слов не обнаружил никакого роста на фоне десятикратного увеличения объема корпуса. Вызывает замечания и анонимность материала: отсутствие социологических данных об авторах текстов, в отличие основного акцентологического корпуса или экспериментального материала (за исключением допущения о возрастной принадлежности авторов к молодому поколению).

Отдельно следует остановиться на ошибках в акцентологической разметке. На фоне экспериментальных данных или размеченного вручную основного акцентологического корпуса, в котором количество ошибок стремится к нулю, погрешности в акцентологической разметке корпуса наивной поэзии кажутся значительными. Ввиду ее автоматического характера ошибки неизбежны, и задача разработчиков корпуса — добиться их приемлемого уровня. Анализ пилотного корпуса показал, что в целом погрешности невелики и на большом объеме материала не будут иметь статистического эффекта. Однако в отдельных случаях доля ошибок в разметке могла достигать 30%. Пожалуй, это наиболее серьезный аргумент, который требует более подробного пояснения.

Причины ошибок могут быть различными. Программа автоматической разметки, как об этом говорится в статье [Гришина и др. 2015], расставляет ударения в несколько этапов. Сначала акцентуация каждого слова определяется в соответствии с наиболее вероятным вариантом, выводимым на основе модели машинного обучения, построенной на окончаниях русских словоформ. Затем, исходя из этой информации, фиксируется наиболее вероятный метр строки, который верифицируется данными о соседних строках. Чаще всего дальнейшие этапы анализа позволяют исправить ошибки первого шага, когда ударения расставляются в соответствии со статистической языковой моделью, а не опираются на реальное положение в строке. Так, ударение слова «вьюга» с точки зрения машинной модели могло бы приходиться на последний слог по аналогии с «юга́» или «города́». Но в разметке достаточное число случаев правильных вхождений формы «вью́га», подсказанных метрической схемой: «С *опа́ской смóтрим дру́г на ду́га* <так в тексте! — Б.О., С.С.>, / Но ме́жду нами сно́ва вью́га»; «Когда́ унима́ется вью́га / стано́вится сне́г безмяте́жным...» В то же время наивные поэты часто

не справляются с силлабо-тонической версификацией, пропускают или добавляют слоги, осуществляют наращеня и усечения, подобные цезурным, но не в позиции цезуры, фактически превращая силлабо-тонический стих в дольник. Нарушения регулярного стиха приводят к ошибкам в автоматической расстановке ударений. Так, следующее стихотворение наивного автора начинается как правильный трех-стопный анапест: «Он дарил ей цветы и духи, / а она поцелуй дарила / сочинял для неё он стихи, / а его она боготворила / они летом любили бродить / в сосняке, и в лесу где берёзы, / с родника воду чистую пить / и вдыхать аромат белой розы». Однако в следующем стихе метр нарушается и в третьей стопе недостает одного безударного слога: «а когда начинала петь / за окошком седая вьюга». Так как в данном случае это нарушение регулярное (то есть наблюдается в нескольких строках подряд), его можно отнести на счет сознательно примененного приема: переход на дольник совпадает с поворотным моментом в лирическом сюжете. Однако программа продолжает считать стих силлабо-тоническим и пытается расставить ударения в нем так, чтобы они соответствовали метрической схеме. В результате мы получаем неверное решение: «вьюга».

Случай другого типа, приводящий к ошибкам в разметке, — разница между письменным и произносимым вариантом. Формально из четырех строк «Возможно, за какие-то заслуги, / И более желанный ждёт обмен, / Я буду, слушать завывание вьюги, / И не бояться больше перемен» три строки (первая, вторая и четвертая) написаны пятистопным ямбом, а третья — снова нарушает версификацию, наращивая лишний безударный слог перед пятой стопой. Эта формальная ошибка снова приводит к неправильной разметке ударения в слове «вьюги». На деле же третья строка точно так же представляет собой правильный пятистопный ямб, но слово, произнесенное как «завыванье» написано иначе: «завывание». Возможно, это результат коррекции орфографии, встроенной в офисный текстовый процессор. Добавление лишней графемы, обозначающей гласный, на месте отсутствующего слога, негативно влияет на точность работы программы разметки. Такого рода «графическая» ошибка не единичный случай, вот аналогичное ему нарушение метра (здесь — четырехстопный ямб) в письменном представлении при очевидно правильном произносительном варианте: «Уж не быть в твоёй судьбе / Любимой девушки, подруги! / А колыбельная тебе — / Ночное завывание вьюги...».

Как представляется, однако, сильным аргументом в пользу включения материала наивной поэзии в состав акцентологического корпуса, который подтвердился в ходе предпринятого исследования, являются достоинства корпуса. К их числу относится, во-первых, большой объем, который, как уже говорилось, позволяет на порядок увеличить данные об акцентных вариантах. Во-вторых, репрезентативность материала: в текстах корпуса представлено достаточно большое количество словоформ, значимых для акцентологических исследований. В-третьих, относительная возрастная однородность авторов, которые в большинстве своем относятся к молодому поколению носителей языка, поэтому в текстах в основном представлена «младшая норма», что позволяет более уверенно делать выводы

о направлении развития акцентологических норм. Дальнейшие исследования с использованием корпуса наивной поэзии, проведенные на других участках колебания акцентологической системы, возможно, подтвердят или уточнят полученные результаты и характеристики материала.

### Литература

*Бонч-Осмоловская А. А., Орехов Б. В.* Корпусно-статистические подходы к наивной поэзии // Корпусный анализ русского стиха: Сборник научных статей. Вып. 2. / Отв ред. В. А. Плунгян, Л. Л. Шестакова. М. : Издательский центр «Азбуковник», 2014. С. 20–36.

*Воронцова В. Л.* Русское литературное ударение XVIII–XX вв. Формы словоизменения. М. : Наука, 1979.

*Грам.* — *Зализняк А. А.* Грамматический словарь русского языка. М. : Русские словари, 2003.

*Гришина Е. А.* Корпус «История русского ударения» // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб. : Нестор-История, 2009. С. 150–174.

*Гришина Е. А.* Микроизменения в акцентологической системе русских прилагательных по материалам Национального корпуса русского языка // Вопросы культуры речи. Вып. 11. М., 2012

*Гришина Е. А., Корчагин К. М., Плунгян В. А., Сичинава Д. В.* Поэтический корпус в рамках НКРЯ: общая структура и перспективы использования // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб. : Нестор-История, 2009. С. 71–113.

*Гришина Е. А., Зеленков Ю. Г., Орехов Б. В.* Наивная поэзия в акцентологическом корпусе // Труды Института русского языка им. В. В. Виноградова. 2015, № 6. С. 257–271.

*Гришина Е. А., Савчук С. О.* Устный корпус в Национальном корпусе русского языка: состав и структура // Национальный корпус русского языка 2006–2008: Новые результаты и перспективы, СПб. : Нестор-История, 2009, 71–113.

*Еськова Н. А.* Картоoteca по современному русскому ударению // Лингвистические источники. Фонды Института русского языка. М.: Наука, 1967.

*Зализняк А. А.* От праславянской акцентуации к русской. М. : Наука, 1985.

*Зеленков Ю. Г., Зобнин А. И., Маслов М. Ю., Титов В. А.* Илья Сегалович и развитие идей компьютерной лингвистики в Яндексе // Труды международного семинара «Диалог14» по компьютерной лингвистике и ее приложениям (электронный документ). URL: <http://www.dialog-21.ru/digests/dialog2014/materials/pdf/ZelenkovYuG.pdf>. 2014.

*Илюшин А. А.* Русское стихосложение. М. : Высшая школа, 2004. 240 с.

*Каленчук М. Л., Савинов Д. М., Скачедубова Е. С.* Активные процессы в просодической системе русского языка: акцентуация прилагательных. // Русский язык в научном освещении. № 2 (34). 2017. С. 9–28.

Куракина Е. Б. Тенденции развития акцентологических норм в речи современных молодых москвичей. АКД Дисс. канд. филол. н. М., 2011.

«Наивная литература»: исследования и тексты / Сост. С. Ю. Неклюдов. М., 2001.

Пожарщкая С. К., Добрушина Е. Р. Орфоэпический взгляд на некоторые варианты явления русского литературного языка в эпоху корпусной лингвистики // Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог 2017». М., 2017. URL: <http://www.dialog-21.ru/media/3940/pozharickajaskdobrushinaer.pdf>

Савчук С. О. Активные процессы в системе русского словоизменения: опыт корпусного исследования акцентологических норм // Труды II Международной конференции «Русский язык и литература в международном образовательном пространстве: современное состояние и перспективы». Т. 2. Гранада, 2010. С. 1549–1554.

Хомицевич О. Г., Рыбин С. В., Таланов А. О., Опарин И. В. Автоматическое определение места ударения в незнакомых словах в системе синтеза речи // Материалы XXXVI Международной филологической конференции. СПб., 2008.

Bonch-Osmolovskaya A., Orekhov B. Distant reading of naive poetry: corpora comparison as research methodology // Digital humanities Lausanne — Switzerland '14. URL: <http://dharchive.org/paper/DH2014/Paper-777.xml>

Grishina, E., Polyakov, A., Savchuk, S. Design and data collection for the Accentological corpus of Russian // LREC 2010 Proceedings, p. 618–622. URL: <http://www.lrec-conf.org/proceedings/lrec2010/>

Lagerberg, Robert. Variation and Frequency in Russian Word Stress. München : Verlag Otto Sagner, 2011. 184 p. URL: [https://www.researchgate.net/publication/45172817\\_Variation\\_and\\_frequency\\_in\\_Russian\\_word\\_stress](https://www.researchgate.net/publication/45172817_Variation_and_frequency_in_Russian_word_stress).

Marklund Sharapova, E. Implicit and Explicit Norm in Contemporary Russian Verbal Stress (Acta Universitatis Upsaliensis. Studia Slavica Upsaliensia 40). Uppsala: Uppsala University, 2000. URL: <https://uu.diva-portal.org/smash/get/diva2:160762/FULLTEXT01.pdf>

Moretti F. Distant reading. London, New-York: Verso, 2013.

Ponomareva, M., Milintsevich, K., Chernyak, E., Starostin, A. Automated Word Stress Detection in Russian. In Proceedings of the First Workshop on Subword and Character Level Models in NLP, pp. 31–35, Copenhagen, Denmark, September 7, 2017. Available at: <http://aclweb.org/anthology/W17-4104>.

Reynolds, R., Tyers, R. (2015) Automatic word stress annotation of russian unrestricted text. In Proceedings of the 20<sup>th</sup> Nordic Conference of Computational Linguistics (NO-DALIDA 2015). Linköping University Electronic Press, Sweden, Vilnius, Lithuania, pp. 173–180. URL: <http://www.aclweb.org/anthology/W15-1822>.

Savchuk, S. Spoken texts representation in the Russian National Corpus: Spoken and Accentologic sub-corpora // “NLP, Corpus Linguistics, Corpus Based Grammar Research”, 5th International conference SLOVKO Smolenice, Slovakia, 25–27 November 2009 Brno, Tribun EU, 2009, pp. 310–320.

Ukiah, N.: 2002, ‘The stress of Russian nouns in -a and -я of Zaliznjak’s pattern f (rybá type)’, Australian Slavonic and East European Studies, 16/1–2, 1–39.

<sup>1</sup>*B.V. Orekhov*, <sup>2</sup>*S.O. Savchuk*

<sup>1</sup>*National Research University Higher School of Economics*

<sup>2</sup>*Vinogradov Russian Language Institute of the Russian Academy of Sciences  
(Russia, Moscow)*

<sup>1</sup>*nevmenandr@gmail.com*, <sup>2</sup>*savsvetlana@mail.ru*

## ACCENTOLOGICAL CORPUS AS A TOOL FOR STUDYING RUSSIAN WORD STRESS

The article addresses several issues related to the development and use of the Accentological corpus as a tool for studying word stress. The current state of the corpus, its composition and structure, development prospects, replenishment with new material are described in this article. Particular attention is paid to the subcorpus of “naive” poetry within the Accentological corpus as a source of accentological data. The capabilities of this resource, its effective use are demonstrated in the study of variants of nominal inflection.

A corpus-based study of inflectional stress variation of singular and plural forms of short adjectives and case forms of nouns has been conducted. While studying case forms of nouns several areas of active competition of accent variants have been examined: feminine nouns with –a ending (*doska* ‘desk’, *stena* ‘wall’), feminine nouns with a soft stem (*kist* ‘brush’, *chel’ust* ‘jaw’) and masculine nouns with a hard stem (*veter* ‘wind’, *sharf* ‘scarf’).

The results of the corpus based study have shown that the material of accentological corpora can be used as a reliable source of accentological data. Increasing the volume of the corpus makes these data statistically reliable, as well as expanding the range of the studied forms and contributes to the detection of new points of variability.

*Key words:* Accentological corpus, the Russian National Corpus, “naive poetry”, Russian word stress, accent variants.

### References

Bonch-Osmolovskaya A. A., Orekhov B. V. [Corpus-statistical approaches to naive poetry] // *Korpusnyi analiz russkogo stikha: Sbornik nauchnykh statei* [Corpus analysis of Russian verse: Collection of articles]. Iss. 2. / V. A. Plungyan, L. L. Shestakova (eds.). Moscow, Azbukovnik Publ., 2014, pp. 20–36. (In Russ.)

Bonch-Osmolovskaya A., Orekhov B. Distant reading of naive poetry: corpora comparison as research methodology // *Digital humanities Lausanne—Switzerland’14*. Available at: <http://dharchive.org/paper/DH2014/Paper-777.xml>

Es’kova N. A. [Card index on modern Russian accent] // *Lingvisticheskie istochniki. Fondy Instituta russkogo yazyka* [Linguistic sources. The funds of the Institute of the Russian language]. Moscow, Nauka Publ., 1967. (In Russ.)

Grishina E. A. [Corpus “The History of Russian word stress”] // *Natsional’nyi korpus russkogo yazyka: 2006–2008. Novye rezul’taty i perspektivy* [The Russian National

Corpus: 2006–2008. New results and prospects]. Saint-Petersburg, Nestor-Istoriya Publ., 2009, pp. 150–174. (In Russ.)

Grishina E. A. [Micro-changes in the accentological system of Russian adjectives based on the materials of the Russian National Corpus] // *Voprosy kul'tury rechi* [Issues on culture of speech]. Iss. 11. Moscow, 2012. (In Russ.)

Grishina E. A., Korchagin K. M., Plungyan V. A., Sichinava D. V. [The poetic corpus within the RNC: General structure and future use] // *Natsional'nyi korpus russkogo yazyka: 2006–2008. Novye rezul'taty i perspektivy* [The Russian National Corpus: 2006–2008. New results and prospects]. Saint-Petersburg, Nestor-Istoriya Publ., 2009, pp. 71–113.

Grishina E. A., Zelenkov Yu. G., Orekhov B. V. [Naive poetry within the Russian accentological corpus] // *Trudy Instituta russkogo yazyka im. V. V. Vinogradova* [Proceedings of the V. V. Vinogradov Russian Language Institute]. 2015, no. 6, pp. 257–271. (In Russ.)

Grishina E. A., Savchuk S. O. [Oral Speech corpus within the Russian National Corpus: composition and structure] // *Natsional'nyi korpus russkogo yazyka 2006–2008: Novye rezul'taty i perspektivy* [The Russian National Corpus: 2006–2008. New results and prospects]. Saint-Petersburg, Nestor-Istoriya Publ., 2009, pp. 71–113. (In Russ.)

Grishina E., Polyakov A., Savchuk S. Design and data collection for the Accentological corpus of Russian // *LREC 2010 Proceedings*, p. 618–622. Available at: <http://www.lrec-conf.org/proceedings/lrec2010/> (accessed 30.10.2018).

Ilyushin A. A. *Russkoe stikhoslozhenie* [Russian poetry]. Moscow, Vysshaya shkola Publ., 2004. 240 p. (In Russ.)

Kalenchuk M. L., Savinov D. M., Skachedubova E. S. [Active processes in the prosodic system of the Russian language: accentuation of adjectives] // *Russkii yazyk v nauchnom osveshchenii* [Russian Language and Linguistic Theory]. 2017, no. 2 (34), pp. 9–28. (In Russ.)

Kurakina E. B. *Tendentsii razvitiya aktsentologicheskikh norm v rechi sovremennykh molodykh moskvichei* [Trends in the development of accentological norms in speech of contemporary young moscovites]. Abstract of Ph. D. thesis. Moscow, 2011. (In Russ.)

Khomitsevich O. G., Rybin S. V., Talanov A. O., Oparin I. V. [Automatic determination of the place of stress in unknown words in the speech synthesis system] // *Materialy XXXVI Mezhdunarodnoi filologicheskoi konferentsii* [Proceedings of XXXVI<sup>th</sup> International philological conference]. St. Petersburg, 2008. (In Russ.)

Lagerberg R. *Variation and Frequency in Russian Word Stress*. München: Verlag Otto Sagner, 2011. 184 p. Available at: [https://www.researchgate.net/publication/45172817\\_Variation\\_and\\_frequency\\_in\\_Russian\\_word\\_stress](https://www.researchgate.net/publication/45172817_Variation_and_frequency_in_Russian_word_stress).

Marklund Sharapova, E. *Implicit and Explicit Norm in Contemporary Russian Verbal Stress*, Uppsala: Uppsala University (Acta Universitatis Upsaliensis. Studia Slavica Upsaliensia). 2000. Available at: <https://uu.diva-portal.org/smash/get/diva2:160762/FULLTEXT01.pdf>

Moretti F. *Distant reading*. London, New-York: Verso, 2013.



“*Naivnaya literatura*”: issledovaniya i teksty [“Naive literature”: studies and texts] / S. Yu. Neklyudov (comp.) Moscow, 2001. Available at: <http://www.ruthenia.ru/folklore/luriem43.pdf> (accessed on 14.09.2018). (In Russ.)

Ponomareva, M., Milintsevich, K., Chernyak, E., Starostin, A. Automated Word Stress Detection in Russian. In *Proceedings of the First Workshop on Subword and Character Level Models in NLP*, pp. 31–35, Copenhagen, Denmark, September 7, 2017. Available at: <http://aclweb.org/anthology/W17-4104>. (In Russ.)

Pozharitskaya S.K., Dobrushina E.R. [Orthoepy in an era of corpus linguistics: some case studies in variation in standard Russian] // *Komp'yuternaya lingvistika i intellektual'nye tekhnologii. Po materialam ezhegodnoi Mezhdunarodnoi konferentsii “Dialog”* [Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue” (2017). Moscow. Available at: <http://www.dialog-21.ru/media/3940/pozharickajaskdobrushinaer.pdf> (In Russ.)

Reynolds, R., Tyers, R. Automatic word stress annotation of russian unrestricted text. In *Proceedings of the 20<sup>th</sup> Nordic Conference of Computational Linguistics (NODALIDA 2015)*. Linköping University Electronic Press, Sweden, Vilnius, Lithuania, pp. 173–180. Available at: <http://www.aclweb.org/anthology/W15-1822>.

Savchuk, S. Spoken texts representation in the Russian National Corpus: Spoken and Accentologic sub-corpora // “*NLP, Corpus Linguistics, Corpus Based Grammar Research*”, 5th International conference SLOVKO Smolenice, Slovakia, 25–27 November 2009 Brno, Tribun EU, 2009, pp. 310–320.

Savchuk S.O. [Active processes in the system of Russian inflection: A corpus-based research of accentological norms] // *Trudy II Mezhdunarodnoi konferentsii “Russkii yazyk i literatura v mezhdunarodnom obrazovatel'nom prostranstve: sovremennoe sostoyanie i perspektivy”* [Proceedings of the II<sup>nd</sup> International conference “Russian language and literature in the international educational space: current state and prospects”]. V. 2. Granada, 2010, pp. 1549–1554. (In Russ.)

Ukiah, N. ‘The stress of Russian nouns in -a and -я of Zaliznyak’s pattern f (rybá type)’, *Australian Slavonic and East European Studies*, 2002, 16/1–2, 1–39.

Vorontsova V.L. *Russkoe literaturnoe udarenie XVIII-XX v. Formy slovoizmeneniya* [Russian literary word stress of the XVIII-XX centuries, Forms of inflection]. Moscow, Nauka Publ., 1979. (In Russ.)

Zaliznyak A. A. *Ot praslavyanskoi aktsentuatsii k russkoi* [From proto-Slavic accentuation to Russian]. Moscow, Nauka Publ., 1985. (In Russ.)

Zaliznyak A. A. *Grammaticheskii slovar' russkogo yazyka* [Grammatical dictionary of the Russian language]. Moscow, Russkie slovari Publ., 2003. (In Russ.)

Zelenkov Yu.G., Zobnina A. I., Maslov M. Yu., Titov V. A. [Ilya Segalovich and development of ideas of computational linguistics to Yandex] // *Trudy mezhdunarodnogo seminarina «Dialog14» po komp'yuternoi lingvistike i ee prilozheniyam* [Proceedings of the international seminar “Dialog14” on computational linguistics and its applications (electronic document)]. Available at: <http://www.dialog-21.ru/digests/dialog2014/materials/pdf/ZelenkovYuG.pdf>. (In Russ.)