

ЯЗЫК ПОЭЗИИ СКВОЗЬ ПРИЗМУ БАЗЫ ДАННЫХ / POETRY LANGUAGE THROUGH DATABASE LENS

Преимущества использования баз данных в лингвистических исследованиях

А. А. Лебедев

Advantages of Using Databases in Linguistic Research

A. A. Lebedev

Александр Александрович Лебедев – кандидат филологических наук, доцент; Петрозаводский государственный университет, Петрозаводск, Российская Федерация

E-mail: perevodchik88@yandex.ru

Статья поступила: 15.09.2024. Принята к печати: 20.10.2024.

В статье рассматриваются особые возможности, которые предоставляют базы данных для лингвистических исследований. Анализируется роль и значимость баз данных на примере четырех конкретных проектов, которые были разработаны или находятся в процессе разработки при активном участии сотрудников кафедры русского языка Петрозаводского государственного университета. Представленные базы данных иллюстрируют, как междисциплинарный подход, включающий тесное взаимодействие лингвистики с информационными технологиями и другими научными дисциплинами, способен обогатить и упростить процесс исследования. В статье подчёркиваются преимущества такого подхода в контексте сбора, систематизации и анализа лингвистического материала, а также в использовании полученных результатов для дальнейших научных изысканий. Одним из ключевых аспектов обсуждения является тот факт, что базы данных позволяют хранить, обрабатывать и анализировать большие объёмы информации, что затруднительно при традиционных методах работы: базы данных упрощают доступ к большому спектру лингвистических данных, позволяют быстро и точно находить необходимые сведения, а также проводить многофакторные анализы, способно увеличить глубину исследований. Приводятся примеры дополнительных возможностей, таких как интеграция данных из разных источников, полуавтоматизированный анализ и возможность быстрого обновления информации, которые практически недоступны при использовании традиционных подходов работы с материалом. Делается вывод о том, что базы данных не только упрощают работу лингвиста, но и открывают новые горизонты для проведения междисциплинарных исследований, предоставляя

Aleksandr A. Lebedev – Candidate of Philological Sciences, Associate Professor; Petrozavodsk State University, Petrozavodsk, Russian Federation

ORCID: 0000-0001-9939-9389

Received: 15.09.2024. Accepted for publication: 20.10.2024.

The paper examines special opportunities that databases provide for linguistic research. The role and significance of databases are analyzed using four specific projects that have been developed or are in the process of development with the active participation of the staff of the Russian Language Department of Petrozavodsk State University. The presented databases illustrate how an interdisciplinary approach, including close interaction of linguistics with information technology and other scientific disciplines, can enrich and simplify the research process. The paper emphasizes the advantages of this approach in collecting, systematizing and analyzing linguistic material, as well as in using the results obtained for further scientific research. One of the key aspects of the discussion is the fact that databases allow storing, processing and analyzing large amounts of information, which is difficult with traditional methods of work: databases simplify access to a wide range of linguistic data, allow to quickly and accurately find the necessary information, as well as to conduct multivariate analyses, which can increase the depth of research. Examples of additional capabilities, such as integration of data from different sources, semi-automated analysis and the ability to quickly update information, which are practically unavailable when using traditional approaches to working with material, are given. It is concluded that databases not only simplify the work of a linguist, but also open up new horizons for interdisciplinary research, providing innovative tools for solving complex problems in the field of linguistics.

инновационные инструменты для решения сложных задач в области лингвистики.

Ключевые слова: база данных, поэтический синтаксис, диахроническая риторика, Феофан Прокопович

УДК 81`322:81`367:808

Keywords: database, poetic syntax, diachronic rhetoric, Feofan Prokopovich

OECD: 6.020Y

V

Постановка проблемы. Современная лингвистика активно использует компьютерные технологии для анализа языкового материала. Одним из ключевых инструментов, способных обеспечить высокую эффективность подобных исследований, являются базы данных. Их применение позволяет существенно расширить возможности

языкового анализа, автоматизировать и алгоритмизировать некоторые рутинные процессы обработки данных, а также обеспечить высокую точность обработки лингвистической информации. При этом преимущества подобного рода работы с материалом известны не всем лингвистам, что и предопределило задачу данной статьи – указать конкретные преимущества использования баз данных на материале тех программных продуктов, которые были разработаны при участии сотрудников кафедры русского языка Петрозаводского государственного университета.

Анализ любых языковых явлений требует обязательного системного подхода, который сложно реализовать без использования современных технологий. Многие современные лингвистические исследования включают в себя необходимость разносторонней обработки больших объемов данных, что значительно усложняет традиционные методы анализа. В связи с этим возникает необходимость детального рассмотрения преимуществ применения баз данных в лингвистических исследованиях на конкретном языковом материале.

История вопроса. В зарубежной лингвистике вопрос использования баз данных в языковых исследованиях был поставлен сравнительно давно (в частности, следует отметить работы [Linguisticdatabases, 1998; Theuseofdatabases, 2009]). В России одним из первых крупных специалистов в этом направлении стали А. С. Герд и А. Н. Баранов. Их работы, как научные [Баранов, 1998; Герд, 2002], так и учебные [Баранов, 2001; Герд, 2005] заложили основы современных представлений о прикладной и компьютерной лингвистике, которые воплощаются в том числе и в современных междисциплинарных исследованиях на стыке лингвистики и информационных технологий. Петрозаводский государственный университет также поддержал еще в начале XXI века эту тенденцию к использованию баз данных в филологических исследованиях; более подробно об этом – в [Захаров, 2006].

Зачастую исследователи описывают использование баз данных в отдельно взятых сферах лингвистической деятельности. Вопрос использования баз данных активно рассматривается в рамках такого направления, как корпусная лингвистика (к числу работ, системно описывающих данный аспект следует отнести [Потапова,

2012]); также подробно исследование аннотированных баз данных в работе корпусного лингвиста представлено в [Пахолкова, 2012].

Помимо корпусных исследований, базы данных играют важную роль в работе по созданию словарей. В статье Н. А. Мишанкиной «Базы данных в лингвистических исследованиях» [Мишанкина, 2013] описаны преимущества применения технологий баз данных в лексикографической практике, к числу которых отнесены эффективность и быстрота получения информации, возможность ее последовательного структурирования, а также перспективы использования полученных материалов как в исследовательской работе, так и в учебном процессе. Также конкретный пример реализации базы данных прагматически маркированной лексики русского языка представлен в статье [Булыгина, 2016].

Методология и методика анализа. При создании любых баз данных (в том числе и предназначенных для лингвистических исследований) применяется ряд принципов и установок, направленных на обеспечение эффективности таких программных продуктов и удобства их использования. Важнейшим принципом выступает систематизация данных: все исследуемые языковые единицы должны быть структурированы таким образом, чтобы обеспечить быстрый доступ к информации и гарантировать возможность выполнения сложных запросов. Достичь реализации данного принципа можно путем создания четкой и непротиворечивой классификации обрабатываемых данных (причем вне зависимости от уровня языка, будь то лексика, морфология, синтаксис и т.п.), а также разработки и последовательного описания метаданных для каждой такой вводимой в систему единицы.

Еще одним важным аспектом выступает обеспечение возможности масштабируемости разрабатываемой базы данных, что должно позволить легко расширить ее по мере поступления новых данных. Если предполагается, что база данных может быть включена в другие проекты, имеет смысл разработать требования к стандартизации данных, что позволяет интегрировать базы между разными проектами, а также обеспечивать их совместимость. Немаловажную роль играет и принцип автоматизации, позволяющий минимизировать вероятность человеческой ошибки в ходе ввода и обработки данных, а также существенно ускорить процесс анализа полученной информации.

Кафедра русского языка Петрозаводского государственного университета активно использует разные типы баз данных в рамках традиционных и современных лингвистических исследований поэтических и прозаических текстов. В работе мы рассмотрим наиболее востребованные из разработанных за последние годы баз данных, связанных с лингвистическими исследованиями, а также опишем те преимущества, которые исследователь получает при работе с такими базами данных в сравнении с традиционным представлением лингвистического материала.

Анализ материала. В ходе работы над «Синтаксическим словарем русской поэзии» в рамках реализации грантов РГНФ 15-04-00180 а – «Синтаксический словарь

языка русской поэзии XVIII – первой половины XIX века» и РФФИ 17-04-00168-ОГН – «Синтаксический словарь русской поэзии XIX века» (руководитель – Н. В. Патроева) с 2015 по 2019 год формировалась база данных синтаксических конструкций, методом сплошной выборки извлеченных из поэтического творчества авторов указанных временных периодов. Немаловажным преимуществом такой базы данных стал принципиально новаторский подход к работе с языковыми единицами. Более подробно методология создания словаря описана в статье [Патроева, 2015], а результатом обработки данных стали два печатных тома синтаксического словаря [Патроева, 2017] и [Патроева, 2019].

Совместный математико-литературоведческо-лингвистический коллектив в ходе работы над проектом РФФИ 18-012-90026 – «Проблема атрибуции анонимных и псевдонимных статей в журналах «Время», «Эпоха» и еженедельнике «Гражданин» (руководитель – А. А. Рогов) сформировал и наполнил базу данных «Размеченный корпус текстов XIX века в дореволюционной и современной графике для решения задач установления авторства» (Свидетельство №2021621353 от 22.06.2021), которая нашла воплощение в электронном ресурсе СМАЛТ (URL: <http://smalt.karelia.ru/>). Преимуществом такой реализации стала общедоступность полученных данных – наличие интернет-версии аналитического блока информационной системы позволяет пользователям работать с данными удаленно. Подробная информация о базе данных представлена в монографии проектного коллектива [Рогов, 2021].

В процессе работы над грантом РФФИ 22-28-00991 «Поэтический синтаксис русского языка XVIII века в риторическом аспекте» (руководитель Н. В. Патроева) была создана и зарегистрирована база данных «Риторический и медитативный вопрос как фигуры речи в лирике XVIII века» (Свидетельство №2023624528 от 11.12.2023). В эту базу данных включены синтаксические конструкции, представляющие собой риторические вопросы, которые были выбраны из поэтического творчества авторов XVIII века – А. Д. Кантемира, М. В. Ломоносова, А. П. Сумарокова, В. К. Тредиаковского.

На момент написания данной статьи научный коллектив под руководством Н. В. Патроевой в ходе работы над проектом «Диахроническая риторика: язык и слог стихотворных произведений Феофана Прокоповича в аспекте тропо- и фигуροобразования» создает базу данных средств речевой выразительности в творчестве Феофана Прокоповича. Предполагается, что сбор и систематизация лингвистических данных обусловит создание прочного фундамента для формирования русской диахронической риторики в дополнение к уже активно и давно развивающимся областям истории риторической традиции в России.

Использование баз данных оказывается оптимальным решением для многих современных лингвистических исследований по целому ряду причин. К примеру, базы данных позволяют эффективно хранить информацию и управлять большими ее объемами. Лингвистические исследования зачастую предполагают работу с большими объемами данных (тексты стихотворений, прозаические тексты,

словарные данные, особым образом размеченные тексты и т.п.). Традиционные методы хранения информации в текстовых файлах становятся неэффективными при работе с большими объемами данных, особенно в тех случаях, когда исследователь формирует сложные запросы. Базы данных обеспечивают эффективное управление этими данными, позволяя легко масштабировать системы и обрабатывать большие массивы информации за небольшое время. Такое хранение информации, к примеру, существенно упростило подготовку к печати материалов двух томов «Синтаксического словаря русской поэзии (объем первого тома составляет 36 печатных листов, объем второго тома – 38 печатных листов). В качестве примера табличного представления данных рассмотрим отрывок из базы данных «Синтаксического словаря русской поэзии» (Таблица 1). В таблице приняты следующие обозначения столбцов:

Столбец 1 — Репрезентации;

Столбец 2 — Паспортизация контекста (заглавие, страница);

Столбец 3 — Жанр;

Столбец 4 — Метр (размер);

Столбец 5 — Количество слов (знаменательных и незнаменательных);

Столбец 6 — Число строк, которые занимает предложение;

Столбец 7 — Тип композиции (тождественнострофическая – С / астрофическая (без графически выделенного разделения на строфы) – А / вольная композиция – ВК);

Столбец 8 — Тип строфы (для тождественнострофических произведений);

Столбец 9 — Наличие переносов из строки в строку (цифра указывает количество случаев несовпадения стихового и синтаксического членения, о чем сигнализирует отсутствие знака препинания в конце строки);

Столбец 10 — Риторические приемы;

Столбец 11 — Структурная схема;

Столбец 12 — Тип предложения с точки зрения полноты / неполноты состава;

Столбец 13 — Тип расширителя модели (однородного ряда);

Столбец 14 — Тип осложнителя модели;

Столбец 15 — Коммуникативный тип предложения (по целеустановке);

Столбец 16 — Тип предложения по эмоциональной окраске (интонации);

Столбец 17 — Тип предложения по общему модальному значению.

Таблица 1. Осложненное простое двусоставное предложение в поэзии М. В. Ломоносова (фрагмент базы данных). Используемые сокращения и обозначения расшифрованы в [Патроева, 2019, с. 30–38].

1	Европа, утомленна в брани, Из пламени подняв главу, К тебе свои простерла длани Сквозь дым, курение и мглу.	Там равной ревностью пылают Сердца, как стогны(,) все сияют В исполненной утех ночи.	Петрополь, небу подражая, Подобны испустил лучи.	Небесну пищу я вкушаю, На верх Олимпа вознесен!	Но каждую обозревши часть, С веселием сие вещает: [ПР]	Во храм, сияющий металлом, Пред трон, украшенный кристаллом, Поспешно простирает ход; Венцем зеленым уязенной И в виск, вещает, облеченной Владычице российских вод.
2	18, 123	18, 126	18, 126	19, 127	19, 128	19, 128
3	ода похвальная	ода похвальная	ода похвальная	ода похвальная	ода похвальная	ода похвальная
4	Я4	Я4	Я4	Я4	Я4	Я4
5	18	13	6	8	8	22
6	4	3	1	2	2	6
7	С	С	С	С	С	С
8	10-стишие	10-стишие	10-стишие	10-стишие	10-стишие	10-стишие
9	1	2				2
10	инв., олиц.	инв., мт., син.	инв., мт.	инв., рит.вос., мт.	инв., эл.	инв., эл., прф., мт.
11	[сущ.+гл.].	[гл.+сущ.+гл.].	[сущ.+гл.].	[мест.+гл.].	[<сущ.>+гл.]:	[<сущ.>+гл.+гл.].
12	пол	пол	пол	пол	неп	неп
13	о/об					о/ск, о/об
14	пр/об (подл.вн.об.), д/об	срв/об	д/об	пр/об	д/об	пр/об
15	пов	пов	пов	пов	пов	пов
16	Невос	Невос	Невос	Вос	Невос	Невос
17	утв	утв	утв	утв	утв	утв

Использование особых языков запросов позволяет исследователям быстро и точно извлекать необходимую информацию из обширных корпусов данных. К примеру, можно получить сведения о статистике частотности слов в определенной группе текстов, извлечь контексты использования определенных лексических единиц или построить сложные запросы для анализа морфологических и синтаксических структур. Эта гибкость значительно повышает продуктивность исследований, существенно сокращая время, необходимое на предварительную обработку данных. Подобные инструменты, к примеру, встроены в базу данных СМАЛТ, и у пользователя есть возможность формировать группы текстов и применять к ним различные статистические методы для исследования (с этой функцией можно ознакомиться

непосредственно в интернет-версии базы данных, доступной по ссылке URL: <http://smalt.karelia.ru/shower/>, блок «Исследования»).

К примеру, в основе метрики Г. Хетсо, используемой для анализа каждого представленного в базе данных «СМАЛТ» текста, учитывается 15 параметров (подробнее изложено в [Хетсо, 1986]):

1. Общее распределение частей речи в первых двух и в последних трех позициях предложения.
2. Распределение частей речи в первой позиции предложения.
3. Распределение частей речи во второй позиции предложения.
4. Сочетание частей речи в первых двух позициях предложения.
5. Распределение частей речи в третьей с конца позиции предложения.
6. Распределение частей речи в предпоследней позиции предложения.
7. Распределение частей речи в последней позиции предложения.
8. Сочетание частей речи в последних трех позициях предложения.
9. Средняя длина слова в буквах, вычисляемая на основании выборок размером в 500 текстовых слов.
10. Общее распределение длины слова.
11. Средняя длина предложения в словах, вычисляемая на основании выборок размером в 30 предложений.
12. Общее распределение длины предложения.
13. Лексический спектр текста на уровне словаря.
14. Лексический спектр текста на уровне текста.
15. Индекс разнообразия лексики.

Ручной подсчет каждого из этих параметров чрезвычайно затратен по времени, в то время как автоматическая обработка текстов позволяет получить нужные данные за небольшой промежуток времени, минимизируя вероятность возникновения ошибок, связанных с подсчетами количественных характеристик (примеры подобных автоматических формируемых данных представлены в Таблицах 2 и 3). Подобные статистические данные могут быть использованы в решении вопросов, связанных с определением авторства текстов, чему посвящена монография [Рогов, 2021].

Таблица 2. Общее распределение частей речи в первых двух и последних трех позициях предложения в тексте «Ряд статей о русской литературе. Введение» Ф. М. Достоевского.

Глава	1	2	3	4	5
0: Существительное	57	83	94	76	379
1: Прилагательное	15	43	44	96	39
2: Числительное	9	5	1	9	1
3: Местоимение	176	133	128	163	65
4: Глагол	57	125	101	63	112
5: Причастие	1	5	11	6	6
6: Деепричастие	2	2	1	0	1
7: Наречие	37	58	57	39	35
8: Категория состояния	5	7	1	5	3
9: Частица	93	95	41	55	6
10: Предлог	58	57	115	86	0
11: Союз	108	35	56	47	1
12: Модальное слово	18	3	2	3	0
13: Междометие	2	0	0	0	0
15: Иностранное слово	0	0	3	4	4
16: Цитата	4	4	5	5	5
17: Вводное слово	26	14	5	1	0
20: Неязыковой символ	0	0	0	1	1
21: Сокращенное слово	1	0	3	7	8
22: Часть многочленного названия	1	1	2	4	4

Таблица 3. Общее распределение длины слова в тексте «Ряд статей о русской литературе. Введение» Ф. М. Достоевского

Длина слова (в буквах)	Количество
1	944
2	1853
3	1226
4	1612
5	1036
6	1496
7	1101
8	951
9	721
10	572
11	409
12	261
13	152
14	78
15	50
16	28
17	9
18	2
19	4
20	1
21	1
22	1

Некоторые лингвистические задачи подразумевают работу с иерархически организованными данными или сложными сетями связей. Современные базы данных обеспечивают эффективное представление и управление такими структурами. Это позволяет исследователям моделировать сложные лингвистические феномены с учетом всех межуровневых связей и зависимостей. В качестве иллюстрации можно рассмотреть структуру базы данных «Риторический и медитативный вопрос как фигуры речи в лирике XVIII века» (Рис. 1).

Многотабличная структура базы данных позволяет формировать сложные запросы, включающие в себя несколько параметров (например, можно получить список произведений, в которых встречается два и более риторических вопроса, выраженных неосложненными простыми двусоставными предложениями или список риторических вопросов определенного автора, которые имеют нужную синтаксическую структуру, причем только из тех стихотворений, которые принадлежит определенному жанру – см. пример в Таблице 4).

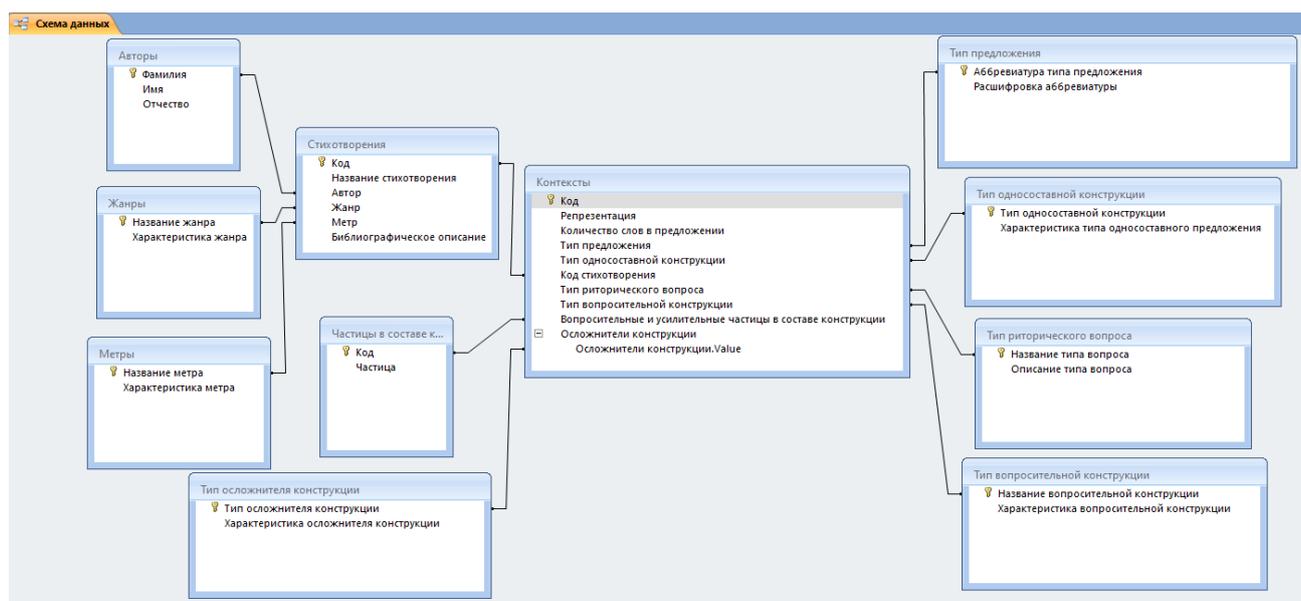


Рис. 1. Структура базы данных «Риторический и медитативный вопрос как фигуры речи в лирике XVIII века»

Таблица 4. Выборка из базы данных «Риторический и медитативный вопрос как фигуры речи в лирике XVIII века», в которой представлены риторические и медитативные вопросы, присутствующие в баснях А. П. Сумарокова и выраженные многокомпонентными сложными предложениями.

Репрезентация	Автор	Аббревиатура типа предложения	Название жанра
От водки голова болит, Но водка сердце веселит, Молошное питье не диво, Его хмельняй и пиво; Какое ж им питье и пить, Коль водки не купить? [Сумароков, 1957, с. 221]	Сумароков	МСП	Басня
О мальчик! узнавать ты был людей искусен, Но знаешь ли теперь, что ты парнасский рак? [Сумароков, 1957, с. 246]	Сумароков	МСП	Басня
Так я тебе скажу об этом поученье: О чем ты сетуешь напрасно, человек, Что твой недолог век И скоро наших тел со духом разлученье? [Сумароков, 1957, с. 222]	Сумароков	МСП	Басня
Не будет никогда чертями там вонять, То правда, и стихи такие пахнут худо, Однако запах сей и истреблять не чудо, Почаще надобно курить, А черт от курева престанет ли дурить? [Сумароков, 1957, с. 237]	Сумароков	МСП	Басня

Использование баз данных способствует автоматизации многих рутинных процессов, что особенно важно при проведении масштабных исследований. Например, можно частично автоматизировать процесс обработки морфологических данных, предлагая в ходе заполнения базы данных наиболее вероятный вариант разбора, что также было реализовано в системе СМАЛТ. Это не только снижает трудозатраты, но и обеспечивает воспроизводимость результатов при всех последующих разборах, что является важным критерием научной достоверности. Возможность хранить и повторно использовать сложные запросы и процедуры в базах данных делает процесс исследования более прозрачным и систематичным. Помимо этого, возможность извлекать статистическую информацию из обработанных данных позволяет получать сведения, которые могут быть использованы для сопоставительного анализа. Примером таких сведений могут стать полученные сравнительные данные об активности риторических приемов в поэзии А. Д. Кантемира, В. К. Третьяковского, М. В. Ломоносова и А. П. Сумарокова (Таблица 5).

Таблица 5. Сравнительные данные, свидетельствующие об активности риторических приемов в поэзии А. Д. Кантемира, В. К. Третьяковского, М. В. Ломоносова и А. П. Сумарокова.

Троп или фигура речи	Количество репрезентаций в стихотворениях				Средняя активность риторического приема на 1 предложение / 1 текст автора			
	Кантемира	Третьяковского	Ломоносова	Сумарокова	Кантемира	Третьяковского	Ломоносова	Сумарокова
Лексический повтор	158	83	276	639	0,2 / 3,6	0,14 / 2,41	0,13 / 2,34	0,21 / 2,49
Метафора	77	141	1275	1025	0,1 / 1,8	0,24 / 4,15	0,6 / 10,81	0,34 / 3,99
Метонимия	39	49	384	207	0,05 / 0,89	0,08 / 1,44	0,18 / 3,25	0,07 / 0,81
Олицетворение	25	36	252	155	0,03 / 0,57	0,06 / 1,06	0,12 / 2,14	0,05 / 0,6
Перифраза	14	14	81	82	0,02 / 0,32	0,02 / 0,41	0,04 / 0,69	0,03 / 0,32
Сравнение	35	21	116	99	0,04 / 0,8	0,04 / 0,62	0,01 / 0,98	0,03 / 0,39
Риторическое восклицание	30	121	267	325	0,04 / 0,68	0,2 / 3,56	0,12 / 2,26	0,11 / 1,26
Риторический вопрос	99	106	227	264	0,12 / 2,25	0,18 / 3,12	0,1 / 1,9	0,09 / 1,03
Риторическое обращение	34	87	304	316	0,04 / 0,79	0,15 / 2,56	0,14 / 2,58	0,1 / 1,23
Гипербола	5	42	28	41	0,01 / 0,11	0,07 / 1,24	0,01 / 0,24	0,01 / 0,16

Еще одно важное преимущество состоит в том, что базы данных позволяют интегрировать лингвистические данные с данными из других дисциплин. Это открывает новые возможности для различных междисциплинарных исследований, где анализ лингвистических данных может быть дополнен другими факторами. В частности, предполагается, что разрабатываемая на данный момент база данных средств речевой выразительности в творчестве Феофана Прокоповича поможет внести вклад в изучение диахронического синтаксиса, исторической лексикологии, лингвопоэтики и стилистики, а также стать основой для лексикографического представления языка Феофана Прокоповича (Таблица 6).

Таблица 6. Фрагмент базы данных «Средства речевой выразительности в творчестве Феофана Прокоповича» (база данных на момент написания статьи находится в процессе разработки)

Автор	Феофан Прокопович
Произведение, год создания	«К Петру Второму», 1728 (?)
Жанр	надпись
Метрика	силлабический 13-сложник
Композиция, строфика	4-стишие
Контекст, выявляющий тропообразование	<i>Бог и Петру Второму вручил стада многа...</i> [Прокопович, 1961, с. 216]
Тип тропа или фигуры	глагольная метафора, узуальная; «фигура смысловая», способствующая «услаждению речи» [Прокопович, 2020, с. 251-259].
Данные словарей русского языка XVIII в.	«Вручаю<...> 1) Отдаю кого или что кому; ввѣрю чьему попеченію, смотрѣнію. <...> 2) Въ руки отдаю...» [Словарь Академии Российской. – СПб.: При Имп. Акад. наук, 1789–1794. Ч. 5. От Р до Т. 1794.] «ВРУЧИТЬ (-ти) <...>. 1. что. Передать из рук в руки, отдать лично кому-л. 2. что. Вверить управление, распоряжение чем-л., поручить. <...> Перен. О духовной пастве» [Словарь русского языка XVIII века / АН СССР. Ин-т рус. яз.; Гл. ред.: Ю. С. Сорокин. Л.: Наука. Ленингр. отд-ние, 1984 (продолж. изд.). Вып. 4. (Воздух — Выпись). Л.: Наука. Ленингр. отд-ние, 1988. 256 с.]

Выводы. Таким образом, следует отметить, что именно использование баз данных в лингвистических исследованиях открывает широкие перспективы для повышения эффективности и качества научных исследований. Они обеспечивают удобное и масштабируемое хранение данных, мощные инструменты для их анализа, поддерживают сложные структуры и связи, способствуют автоматизации и воспроизводимости, а также поддерживают междисциплинарный подход. В будущем, с развитием технологий, базы данных будут играть все более важную роль в лингвистике, открывая новые горизонты для научных исследований.

К числу наиболее значимых преимуществ, которые предлагают базы данных, следует отнести возможность эффективно работать с большими объемами данных, извлекать нужную информацию по запросу пользователя, моделировать сложные многоуровневые лингвистические структуры, автоматизировать некоторые процессы, связанные с обработкой данных, а также интегрировать полученные сведения в рамках междисциплинарных исследований, упрощая совместную работу научных коллективов. Все перечисленные выше возможности нашли свое воплощение в конкретных примерах реализованных баз данных.

Благодарности: исследование выполнено за счет гранта Российского научного фонда № 24-28-00696, <https://rscf.ru/project/24-28-00696/>

Литература

Баранов, А. Н. (1998). Автоматизация лингвистических исследований: корпус текстов как лингвистическая проблема. *Русистика сегодня: язык: система и ее функционирование: сборник статей*. Москва: Наука. 1–2, 179–191.

Булыгина, Е. Ю., Трипольская, Т. А. (2016). База данных прагматически маркированной лексики русского языка: материал, принципы описания, возможности использования. *Вестник Новосибирского государственного педагогического университета*, 6(34), 70–85. DOI: 10.15293/2226-3365.1606.06.

Герд, А. С. (2002). Базы данных и прикладная лингвистика. *Доклады научной конференции «Корпусная лингвистика и лингвистические базы данных», 5–7 марта 2002 г., Санкт-Петербург*. Санкт-Петербург: Изд-во СПбГУ. 3–6.

Герд, А. С. (2005). Прикладная лингвистика. Санкт-Петербург, Изд-во Санкт-Петербургского университета.

Захаров, Н. В. (2006). Филологические науки и информационные технологии. *Знание. Понимание. Умение: фундаментальные и прикладные исследования в области гуманитарных наук*, 2, 166–172.

Мишанкина, Н. А. (2013). Базы данных в лингвистических исследованиях. *Вопросы лексикографии*, 1(3), 25–33.

Патроева, Н. В., Лебедев, А. А. (2015). Проект синтаксического словаря языка русской поэзии XVIII – первой половины XIX века. *Ученые записки Петрозаводского государственного университета. Серия. Общественные гуманитарные науки*, 3(148), 53–55.

Пахолкова, И. А. (2012). Применение методов корпусной лингвистики в традиционном языкознании. *Вестник Московского государственного лингвистического университета*. 646. 125–135.

Потапова, Р. К. (2012). Дискурсивная составляющая современной корпусной лингвистики (применительно к устно-речевым базам данных). *Вестник Московского государственного лингвистического университета*. 639. 157–167.

Прокопович, Ф. (1961) Сочинения / под редакцией И. П. Еремина. Москва; Ленинград: Издательство Академии наук СССР.

Прокопович, Ф. (2020) Об искусстве риторическом десять книг / перевод Г. А. Стратановского; ответственный редактор С. И. Николаев. Москва; Санкт-Петербург: Альянс-Архео.

Рогов, А. А., Абрамов, Р. В., Бучнева, Д. Д., Захарова, О. В., Кулаков, К. А., Лебедев, А. А., Москин, Н. Д., Отливанчик, А. В., Савинов, Е. Д., Сидоров, Ю. В. (2021). *Проблема атрибуции в журналах «Время», «Эпоха» и еженедельнике «Гражданин»: коллективная монография*. Петрозаводск: Острова.

Патроева, Н. В. (редактор) (2017). Синтаксический словарь русской поэзии XVIII века: в 4 т. Т. 1: Кантемир, Третьяковский. Санкт-Петербург: Дмитрий Буланин.

Патроева, Н. В. (редактор) (2019). Синтаксический

References

Baranov, A. N. (1998). Automation of linguistic research: a corpus of texts as a linguistic problem. *Rusistika segodnya [Russistics today: language: system and its functioning: collection of papers]*, 1–2, 179–191. (In Russian).

Bulygina, E. Yu., Tripolskaya, T. A. (2016). Database of pragmatically marked lexical items of the Russian language: the content, principles of description, and possibilities of using. *Vestnik Novosibirskogo gosudarstvennogo pedagogicheskogo universiteta*, 6(34), 70–85, 10.15293/2226-3365.1606.06. (In Russian).

Gerd, A. S. (2002). Databases and applied linguistics. *Proceedings of the Scientific Conference "Corpus Linguistics and Linguistic Databases", March 5–7, 2002, St. Petersburg*. St. Petersburg: St. Petersburg State University Publ., 3–6. (In Russian).

Gerd, A. S. (2005). *Applied linguistics*. St. Petersburg: St. Petersburg State University Publ. (In Russian).

Khetso, G. (1986). *Attribution to Dostoevsky: On the Question of the Attribution to F. M. Dostoevsky of Anonymous Articles in the Magazines "Vremya" and "Epokha"*. Oslo: Solum Forlag A. S. 82 (In Russian).

Linguistic databases (1998). Edited by John Nerbonne. Stanford, Calif.: CSLI Publications.

Mishankina, N. A. (2013). Databases in linguistic studies. *Voprosy leksikografii*, 2013, 1(3), 25–33. (In Russian).

Patroeva, N. V. (ed.) (2017). *Syntactic Dictionary of Russian Poetry of the 18th Century: in 4 volumes. Vol. 1: Kantemir, Trediakovskiy*. Saint Petersburg: Dmitrii Bulanin Publ. (In Russian).

Patroeva, N. V. (ed.) (2019). *Syntactic Dictionary of Russian Poetry of the 18th Century: in 4 volumes. Vol. 2: Lomonosov*. Saint Petersburg: Dmitrii Bulanin Publ. (In Russian).

Patroeva, N. V., Lebedev, A. A. (2015). Syntactic dictionary of Russian poetry of XVIII — first half of XIX century. *Proceedings of Petrozavodsk state university. Social sciences & Humanities*, 3(148), 53–55. (In Russian).

Pakholkova, I. A. (2012). Application of the methods of corpus linguistics in traditional linguistics. *Vestnik of Moscow State Linguistic University*, 646, 125–135. (In Russian).

Potapova, R. K. (2012). Discourse component of modern corpus linguistics (as applied to oral speech databases). *Vestnik of Moscow State Linguistic University*, 639, 157–167. (In Russian).

Prokopovich F. (1961). *Works*. Edited by I. P. Eremin. Moscow; Leningrad: USSR Academy of Sciences Publ. (In Russian).

Prokopovich F. (2020). *Ten books on the art of rhetoric*. Translated by G.A. Stratanovsky; executive editor S.I. Nikolaev. Moscow; Sankt-Peterburg: Al'yans-Arkheo Publ. (In Russian).

Rogov, A. A., Abramov, R. V., Buchneva, D. D., Zakharova O. V., Kulakov, K. A., Lebedev, A. A., Moskin, N. D., Otlivanchik, A. V., Savinov, E. D., Sidorov, Iu. V. (2021). *The*

словарь русской поэзии XVIII века: в 4 т. Т. 2: Ломоносов. Дмитрий Буланин.

Сумароков, А. П. (1957). Избранные произведения. Ленинград: Советский писатель.

Хетсо, Г. (1986). Принадлежность Достоевскому: к вопросу об атрибуции Ф. М. Достоевскому анонимных статей в журналах *Время* и *Эпоха*. Oslo: Solum Forlag A. S.

Nerbonne, J. (edited) (1998). *Linguistic databases*. Stanford (California): CSLI Publications.

Everaert, M., Musgrave, S., Dimitriadis, A. (edited) (2009). *The use of databases in cross-linguistic studies*. Berlin; New York: Mouton de Gruyter.

Problem of Attribution in the Magazines "Vremya", "Epokha" and the weekly "Grazhdanin". Petrozavodsk: Ostrova Publ. (In Russian).

Sumarokov, A. P. (1957). *Selected Works*. Leningrad: Sovetskiy pisatel' Publ., 608. (In Russian).

The use of databases in cross-linguistic studies (2009). Edited by Martin Everaert, Simon Musgrave, Alexis Dimitriadis. Berlin; New York: Mouton de Gruyter Publ.

Zakharov, N. V. (2006). Philological sciences and information technologies. *Knowledge. Understanding. Skill*. 2006, 2, 166–172. (In Russian).

Для цитирования статьи:

Лебедев, А. А. (2024). Преимущества использования баз данных в лингвистических исследованиях. *VERBA. Северо-Западный лингвистический журнал*, 3(13), 36–49. DOI: 10.34680/VERBA-2024-3(13)-36-49

For citation:

Lebedev, A. A. (2024). Advantages of Using Databases in Linguistic Research. *VERBA. North-West linguistic journal*, 3(13), 36–49. (In Russian). DOI: 10.34680/VERBA-2024-3(13)-36-49