



DOI: <https://doi.org/10.15688/jvolsu2.2022.6.4>

UDC 811.161.1'04
LBC 81.411.2-03



Submitted: 31.05.2022
Accepted: 19.09.2022

**CORPUS OF THE ARCHIVAL DOCUMENTS
OF THE DON COSSACK ARMY:
PROBLEMS OF MORPHOLOGICAL ANALYSIS**

Oksana A. Gorban

Volgograd State University, Volgograd, Russia

Marina V. Kosova

Volgograd State University, Volgograd, Russia

Elena M. Sheptukhina

Volgograd State University, Volgograd, Russia

Andrey V. Svetlov

Volgograd State University, Volgograd, Russia

Abstract. The article presents the results of the collective project aimed at comprising a special annotated diachronic corpus of documents of the 18th – 19th cen. from the “Mikhailovsky Stanitsa Ataman” Archive Fund (State Archive of Volgograd Region, Russia). In the course of the work, linguistic, technical and software tasks related to meta-marking, morphological tagging and representation of marked texts in an electronic search environment were solved. The texts are written in cursive script of the 18th cen. with the use of the old Cyrillic letters, which have spelling specificity. To work correctly with them, an add-on to the stemming tool MyStem by I. Segalovich was created. This application adds to the MyStem the following capabilities: the option to work with the old Cyrillic symbols, a convenient graphical interface; it provides the opportunity to remove homonymy manually, enables marked text exporting to an external data storage and processing system. Morphological analysis of some texts revealed the presence of nominal case form variants, which were not noted in the “Russian Grammar” by M.V. Lomonosov, in modern studies of literary texts of the 18th century. These findings point to effectiveness of automatic tagging which allows word form correction. The research results substantiated text tagging software tools adjustment for the extension of homonymous forms grammatical analysis options, aimed at identification and manual removal of homonymy. A quantitative analysis of these variants will allow the authors to evaluate their significance for the regional administrative language. The information obtained confirms the importance of the corpus creation for studying the history of the Russian language.

Key words: history of the Russian language, regional business writing, linguistic corpus, morphological markup, variants of case forms, grammatical homonymy.

Citation. Gorban O.A., Kosova M.V., Sheptukhina E.M., Svetlov A.V. Corpus of the Archival Documents of the Don Cossack Army: Problems of Morphological Analysis. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], 2022, vol. 21, no. 6, pp. 47-56. (in Russian). DOI: <https://doi.org/10.15688/jvolsu2.2022.6.4>

**КОРПУС АРХИВНЫХ ДОКУМЕНТОВ ВОЙСКА ДОНСКОГО:
ПРОБЛЕМЫ МОРФОЛОГИЧЕСКОГО АНАЛИЗА****Оксана Анатольевна Горбань**

Волгоградский государственный университет, г. Волгоград, Россия

Марина Владимировна Косова

Волгоградский государственный университет, г. Волгоград, Россия

Елена Михайловна Шептухина

Волгоградский государственный университет, г. Волгоград, Россия

Андрей Владимирович Светлов

Волгоградский государственный университет, г. Волгоград, Россия

Аннотация. В статье обобщены некоторые результаты исследования, нацеленного на создание специального аннотированного диахронического корпуса документов архивного фонда Михайловского станичного атамана Государственного архива Волгоградской области. Предложены решения лингвистических и технических задач, связанных с разметкой текстов и представлением их в электронно-поисковой среде. Для корректной работы с текстами документов, написанными скорописью XVIII в. и первой половины XIX в., группой специалистов под руководством А.В. Светлова создана надстройка над утилитой для стемминга MyStem И.В. Сегаловича. Она добавляет к возможностям утилиты выводить всю грамматическую информацию и количественные данные опцию работы с устаревшей кириллицей, удобный графический интерфейс, имеет функционал для снятия омонимии вручную, поддерживает экспорт размеченного текста во внешнюю систему хранения и обработки данных. В результате морфологического анализа части текстов О.А. Горбань и Е.М. Шептухиной установлены варианты падежных форм существительных, не отмеченные в «Российской грамматике» М.В. Ломоносова, в современных исследованиях литературных текстов XVIII в., и обоснована необходимость корректировки программных инструментов разметки текстов в плане расширения возможных вариантов грамматического разбора омонимичных форм для последующего снятия омонимии вручную. Количественный анализ выявленных фактов, проведенный М.В. Косовой, и их лингвистическая интерпретация показали в сфере грамматики, наряду с общими тенденциями развития и функционирования русского языка, особенности, характерные для деловой письменной речи в ее региональном варианте.

Ключевые слова: история русского языка, региональная деловая письменность, лингвистический корпус, морфологическая разметка, варианты падежных форм, грамматическая омонимия.

Цитирование. Горбань О. А., Косова М. В., Шептухина Е. М., Светлов А. В. Корпус архивных документов Войска Донского: проблемы морфологического анализа // Вестник Волгоградского государственного университета. Серия 2, Языкознание. – 2022. – Т. 21, № 6. – С. 47–56. – DOI: <https://doi.org/10.15688/jvolsu2.2022.6.4>

Введение

Одна из проблем развития русского исторического языкознания заключается в малодоступности значительной части источников для широкого круга исследователей. Особенно это касается массива памятников деловой письменности, хранящихся в многочисленных российских региональных архивах. Введение такого материала в научный обо-

рот даст возможность более полно представить картину исторического развития русского языка во всем ее многообразии, с учетом местных вариантов. Традиционно эта задача решается путем публикации письменных памятников. Новые возможности открываются благодаря современным информационным технологиям и созданию лингвистических корпусов. В данной статье представлены некоторые результаты исследования, осуществ-

ленного в Волгоградском государственном университете в рамках коллективного проекта, нацеленного на создание специального аннотированного диахронического корпуса документов Войска Донского XVIII–XIX вв. (участники проекта – О.А. Горбань, М.В. Косова, Е.М. Шептухина, А.В. Светлов, И.С. Герасимова, Н.И. Тихонова). В 2019–2021 гг. работа по составлению этого корпуса выполнялась при финансовой поддержке РФФИ (проект № 19-012-00246 «Параметризация текстов как основа создания лингвистического корпуса архивных документов XVIII–XIX вв. канцелярий Войска Донского»).

Земля Войска Донского представляла собой специфичное административно-территориальное образование в составе Российской империи. Лишившись при Петре I государственной автономии, оно сохраняло автономию областную, что обуславливало его социальное и этнокультурное своеобразие, несмотря на все большее подчинение на протяжении двух столетий общегосударственной системе управления. В рамках норм этой системы складывалась и функционировала документация, обеспечивавшая управление Войском со стороны верховной власти и управление внутри Войска.

В языковом отношении это территория формирования и распространения донских говоров южнорусского наречия. Памятники деловой письменности, созданные в этом регионе, отражают взаимодействие общерусских и региональных языковых черт и представляют собой ценный источник для изучения истории русского языка на Юге России в XVII – начале XX века. Однако, как отмечают исследователи, обращение к донским документальным источникам затруднено тем, что значительная их часть утрачена, а оставшаяся рассредоточена по различным центральным и региональным хранилищам, из которых основное – Государственный архив Ростовской области [Сень, 2014, с. 484]. Часть документов хранится в Государственном архиве Волгоградской области (далее – ГАВО).

Материал и методы

Создаваемый лингвистический корпус основан на документах 1734–1837 гг., отложившихся в фонде Михайловского станично-

го атамана ГАВО (фонд 332, опись 1, дела 1–158; далее при цитировании источников указываются только дело и лист / оборот листа, примеры приводятся в упрощенной графике с использованием современного алфавита). Фонд включает оригиналы и копии документов, поступивших в канцелярию (станичную избу) станицы Михайловской на Хопре из войсковой канцелярии, из учреждений Волжского войска, соседних губерний, в частности Воронежской, от должностных лиц и рядовых казаков; черновики и копии документов, отправленных из станичной канцелярии в другие учреждения или отдельным должностным лицам; копии некоторых высочайших указов, касавшихся Войска Донского. Часть документов опубликована (см.: [Горбань и др., 2020]). В фонде представлена большая коллекция оригиналов войсковых грамот; документы этого уникального жанра характерны именно для Войска Донского. Войсковые грамоты, содержащие распоряжения Войска для исполнения в донских станицах, копировались в минимальном количестве, грамота от Черкасского – столицы Войска – направлялась по всем станицам, расположенным вдоль одной реки, в каждой из них читалась и копировалась, затем передавалась дальше и оставалась в последней. Такой станицей в верховьях реки Хопёр и была Михайловская.

Документы написаны скорописью первой и второй половины XVIII в. и первой половины XIX века. Для них характерно: использование вышедших из употребления букв кириллицы (ѣ, і, ѣ, ѡ, ѣ, ѓ, ѣ, ѣ, ѣ), выносных букв; запись слов под титлом; деление текста на слова, но слитное написание союзов, предлогов и частиц с последующей или предыдущей словоформой; отсутствие прописных букв как маркеров имен собственных; редкое и непоследовательное употребление знаков препинания, затрудняющее выделение синтаксических единиц, в частности предложений; написание чисел арабскими цифрами.

Разработка параметров и инструментов метаразметки [Филимонов и др., 2020], принципов представления текстов в электронно-поисковой среде, подготовка их для автоматической машинной обработки потребовали решения ряда лингвистических и технических задач.

С целью расширения возможностей исторических лингвистических исследований тексты даны в корпусе с сохранением орфографии оригиналов и подвергнуты лишь частичной адаптации: выносные буквы даны в строке, диграф передан как сочетание двух букв в строке; титла раскрыты; союзы, предлоги и частицы даны отдельно с последующими или предыдущими словоформами; имена собственные (личные имена и географические названия) – с прописной буквы. Такое представление материала в основном соответствует принципам презентации текстов в Национальном корпусе русского языка [Савчук, Сичинава, 2009, с. 57], облегчает работу пользователей и обеспечивает адекватную работу программ по морфологической разметке текстов. Однако при этом вышедшие из употребления кириллические буквы сохранены, восстановленные при раскрытии титл буквы приведены в круглых скобках, поскольку они даются в современной орфографии и должны быть маркированы как неаутентичные. Сохранена также пунктуация оригиналов, поскольку невозможно однозначно осуществить синтаксическое членение текстов и оформить его по современным правилам. Проведена структурная разметка, предусматривающая деление текстов на сегменты (не всегда совпадающие с предложениями), необходимые и достаточные для контекстуального представления пользователю результатов поиска в корпусе (подробно об этом см.: [Горбань, Косова, Шептухина, 2021]).

В данной статье рассмотрим некоторые вопросы, связанные с морфологической разметкой, правила которой во многом совпадают с принятыми в Национальном корпусе русского языка (см. о них: [Ляшевская, Плунгян, Сичинава, 2005]).

Результаты и обсуждение

Для морфологической разметки текстов используется морфологический анализатор MyStem, разработанный И.В. Сегаловичем [Segalovich, 2003]. MyStem корректно обрабатывает устаревшую лексику, если она передается в программу с использованием современной кириллицы. Поэтому для работы с устаревшей графикой IT-специалистами – участниками проекта было создано приложение-

надстройка к MyStem, которое буквы *i, ѣ, ѡ, ѣ, љ* заменяет соответственно на *и, е, о, ф, кс*, затем передает текст для анализа в MyStem и после этого производит уже в обработанном документе обратную замену символов. Приложение добавляет к возможностям утилиты удобный графический интерфейс, имеет функционал для снятия омонимии вручную [Комендантов, Матвеев, Светлов, 2019].

При помощи созданного приложения была проведена пробная морфологическая разметка документов 1734–1753 гг., которая выявила случаи ошибочного определения форм программой и вызвала необходимость ее корректировки. Эти случаи обусловлены, с одной стороны, выбранным подходом к орфографической передаче текстов, с другой – грамматическими особенностями самих текстов.

Так, слова с раскрытыми титлами не были прочитаны как целостные единицы, потому что круглые скобки воспринимались как знак препинания. Данная проблема устранена путем игнорирования круглых скобок; такая возможность объясняется тем, что скобки в функции знаков препинания в размечаемых документах не используются и не имеют значения для передачи их синтаксиса.

Трудности связаны также с правописанием некоторых грамматических формантов, например падежных окончаний адъективного и местоименного склонений. Как известно, флексия родительного падежа единственного числа мужского и среднего родов прилагательных, причастий, ряда местоимений могла передаваться на письме как *-аго/-яго* и *-ого/-его* (это объясняется историей местоименных форм в церковнославянском и восточнославянских языках; отношения данных вариантов – предмет отдельного рассмотрения). Такое вариативное написание широко представлено и в документах Войска Донского, однако программа корректно определяет оба варианта как формы прилагательных и причастий.

Наряду с указанными, в текстах часто встречаются написания окончаний *-ова/-ева*, местоимения 3-го лица *ево* (вместо *его*), которые отражают живое произношение соответствующих форм: *пьянова* (д. 8, л. 4), *хорчевова* (д. 9, л. 14), *коровева, понова* (д. 9, л. 14 об.), *какова* (д. 8, л. 30), *никакова* (д. 9, л. 9) и др.; они же в современной орфографии:

пьяного, харчевого, коровьего, постного, какаго, никакого. Формы на *-ова* в своем графическом облике омонимичны формам широко распространенных русских фамилий на *-ов* и последовательно определены программой как имена существительные (отметим, что в рассмотренных документах фамилия графа Г.П. Толстого в родительном падеже пишется только как *Толстова*, в других падежах, по аналогии, *Толстову*, *Толстовым* наряду с правильными *Толстой*, *Толстому*). В связи с этим требуется корректировка программы в плане расширения количества возможных грамматических разборов подобных словоформ для последующего снятия омонимии.

Такое расширение количества омонимов и, следовательно, вариантов разбора у одной морфологической формы может быть связано не только с орфографией, но и с собственно грамматическими особенностями текстов. Морфологический анализ обнаружил формы, не сохранившиеся в современном русском языке, некоторые из них не отмечаются уже грамматиками XVIII века. Соответственно, омонимичные формы не получают всего набора морфологических признаков, что обуславливает возможность ошибки в их разборе. Известные же и сегодня варианты формы могут иметь функциональную специфику. В связи с этим перед исследовательским коллективом стоит задача определить, насколько частотны такие явления и требуют ли они существенной корректировки программы. Однако уже имеющиеся результаты позволяют оценить значимость документов создаваемого корпуса как источников для истории русского языка и получить представление о нормах делового языка рассматриваемого исторического периода в среде донского казачества.

В качестве иллюстрации рассмотрим падежные формы имен существительных родительного и предложного падежей единственного числа (далее – Р. ед., П. ед.), дательного, творительного и предложного падежей множественного числа (далее – Д. мн., Т. мн., П. мн.).

Родительный и предложный падежи единственного числа. Наличие у существительных мужского рода вариантных окончаний *-а/-я*, *-у/-ю* в Р. ед. и окончаний *-е*, *-у/-ю* в П. ед. является хорошо известной и давней

чертой русского языка. Они широко представлены в памятниках деловой письменности XVII и XVIII вв., что отмечается многими учеными (сошлемся здесь на книгу Н.И. Тарабасовой, в которой учтены наблюдения А.А. Шахматова, С.И. Коткова, В.М. Маркова и других исследователей: [Тарабасова, 1986, с. 62–80]). Варианты Р. ед. и П. ед. подробно описаны в «Российской грамматике» М.В. Ломоносова, «Российской грамматике» А.А. Барсова с указанием лексических, словообразовательных, морфонологических, стилистических признаков их дифференциации [Ломоносов, 1755, с. 81–85; Барсов, 1981, с. 117–122, 437–442]. Н.А. Еськовой показано употребление этих форм в литературных произведениях XVIII в., отличающееся от современной нормы и свидетельствующее об их конкуренции [Еськова, 2008, с. 545–621]. Так называемые второй родительный и второй предложный с окончанием *-у* отражены в Национальном корпусе русского языка (далее – НКРЯ), где отмечается омонимия этих форм и формы дательного падежа единственного числа. Традиционно вариант на *-у* связывают с народно-разговорной традицией, однако С.И. Котков считал, что это черта всего русского языка, разговорного и литературного (см.: [Тарабасова, 1986, с. 74]). Материал документов изучаемого фонда подтверждает имеющиеся сведения об употреблении данных вариантов, однако количественный анализ словоформ позволяет дополнить их некоторыми наблюдениями.

Так, в Р. ед. в размеченных текстах у слова *год* существенно преобладает форма на *-у* (86 %). У слов *полк*, *род*, *сыск*, *уезд*, *указ* она является единственной. Сопоставление этих данных с данными НКРЯ (в подкорпусе 1734–1753 гг.) по состоянию на 25.10.2021 показывает в деловых текстах большую частотность этой формы у слов *род*, *сыск*, *полк* и меньшую – у слов *указ*, *уезд*.

Формы П. ед. на *-у* в текстах немногочисленны и зафиксированы у слов *круг*, *год*, *сыск*. Отметим такую особенность, как употребление слова *сыск* только во втором предложном, в том числе с предлогом *о*, причем в одном случае, вероятно, под влиянием стоящей рядом формы *сыску* написана и форма *поиску*: *велено о сыску и поиску бежавших*

ис крепости святые Анны [...] публиковать (д. 9, л. 38). Источники, преимущественно следственные документы, этого же периода в НКРЯ показывают распределение *о сыске*, но *в сыску* (хотя зафиксирован один случай употребления *о сыску* в работе В.Н. Татищева 1733 г.). Словоформа *о сыску* отмечается в памятниках деловой письменности XVII в., где П. ед. на -у, характерный для народно-разговорной речи, использовался довольно широко [Тарабасова, 1986, с. 76, 80]. Документы фонда в целом сохраняют эту особенность.

Дательный, творительный и предложный падежи множественного числа. Результаты морфологического анализа показали наличие вариантов форм Д. мн., Т. мн. и П. мн., нехарактерных для современного русского языка.

Для Д. мн. наряду с узуальным окончанием *-ам/-ям* отмечается также окончание *-ом/-ем*, восходящее к древнему продуктивному типу склонения на **ǫ* и свойственное большинству имен существительных мужского рода с окончанием *-ь/-ь* (позднее – нулевым) в начальной форме. Как известно, оно постепенно вытеснялось формами на *-ам/-ям* под влиянием имен на **ǣ* в результате унификации системы склонений. Однако, по наблюдениям Н.И. Тарабасовой, еще в XVII в. в деловых памятниках окончание *-ом/-ем* в Д. ед. в именах мужского рода преобладало, причем встречалось и под ударением, что позволяет видеть здесь не только орфографическую примету [Тарабасова, 1986, с. 97–107]. Грамматики М.В. Ломоносова и А.А. Барсова данный вариант Д. мн. не отмечают [Ломоносов, 1755, с. 69; Барсов, 1981, с. 128, 447–448]; поиск в подкорпусе НКРЯ результатов также не дал.

Формы на *-ом/-ем* являются омонимичными формам творительного падежа единственного числа, и для соответствующих словоформ при морфологической разметке должны быть предусмотрены варианты грамматического разбора. Наличие Д. мн. на *-ом* характерно для существительных *указ* и *атаман*, что может быть интерпретировано по-разному. В первую очередь необходимо сказать, что в обоих словах окончание Д. мн. является безударным и в «акающих» донских говорах *-ом* и *-ам* фонетически не различаются.

Слово *указ* в размеченных документах встречается только в форме *указом* в сочетании с предлогом *по*; данная конструкция регулярно используется для ссылки на постановляющие документы верховной власти, и ее можно рассматривать как устойчивую и воспроизводимую, в том числе орфографически; встречается она и в копиях высочайших указов.

Словоформа *атаманом* используется, как правило, в устойчивых формулах адресата *атаманом и казакам*, например в начальном протоколе войсковых грамот: *От донских атаманов и казаков [...] станицнымъ атаманомъ и казакамъ объявляемъ* (д. 1, л. 2; и др.). Результаты количественного анализа показывают, что вариантные формы представлены примерно в равных долях, но *атаманом* чаще встречается в ранних документах и со временем уступает место форме на *-ам*. Элементы формуляра войсковых грамот во многом продолжают традиции делопроизводства московских приказов и сохраняют некоторые обороты, устоявшиеся в деловой письменности.

Кроме того, образцами для войсковых и станичных писарей могли служить документы, поступавшие из вышестоящих учреждений. Так, высочайшие грамоты XVII–XIX вв. в адрес Войска Донского, собранные и опубликованные А.А. Лишиным, содержат аналогичную начальную формулу, в которой примерно до середины 1730-х гг. последовательно употребляется выражение *донским атаманом и казаком*, затем наблюдается регулярное *атаманом и казакам*, и примерно с середины 1750-х гг. – *атаманам и казакам* (Акты..., т. 1, с. 241, 246 и др.; т. 2, с. 132 и след.; т. 3, с. 728 и след.). Если форму *казаком* с ударным окончанием еще можно считать архаизмом (возможно, и в произношении) как чертой приказного языка, а при ней так могла восприниматься и *атаманом*, то в сочетании *атаманом и казакам* написание *атаманом*, вероятно, уже чисто орфографическое явление. Добавим также варианты написания окончания Д. мн. прилагательного в одном и том же документе: *станицномъ атаманомъ и казакам* (д. 8, л. 35) и *станицнымъ атаманом и казакамъ* (д. 8, л. 36 об.), где отражена нейтрализация противопоставления /ы/, /о/, /а/ в заударном положении, характерная для донских и некоторых других

южнорусских говоров [Касаткин, 1999, с. 483–485; Осипов, Гейгер, Рогожникова, 1993, с. 21–22]. Таким образом, написание *атаманом*, скорее всего, определяется орфографической нормой, которая поддерживается документами центральных учреждений.

Формы Т. мн. имеют регулярные окончания *-ами/-ями* и *-ми*, что отражает общую норму литературного языка XVIII века. Однако в документах фонда у некоторых имен мужского рода и *pluralia tantum* встречаются также формы с окончанием *-ы/-и*, восходящие к древнему типу склонения на **ѣ*, например: *с пожитки* (д. 5, л. 3), *всеми образы и меры* (д. 5, л. 3 об.), *с тремя человеки* (д. 7, л. 11 об.), *под немалыми штрафы* (д. 8, л. 14). Конкуренция старой и новой форм была широко представлена в деловой письменности XVII в. [Тарабасова, 1986, с. 104–112]. М.В. Ломоносов в своей Грамматике вариант *-ы/-и* у имен данного типа склонения не отмечает [Ломоносов, 1755, с. 68–71, 76], однако А.А. Барсов пишет о ней как о довольно регулярной, хотя и имеющей некоторые лексические ограничения [Барсов, 1981, с. 129–130, 449]. Формы Т. мн. на *-ы/-и* вступают в отношения омонимии с формам родительного падежа единственного числа женского рода и именительного падежа множественного числа мужского и женского родов, предложного падежа имен среднего рода с мягким согласным в исходе основы; окончание *-и* свойственно нескольким падежным формам имен женского рода с нулевым окончанием, что также необходимо учитывать при автоматической разметке.

В размеченных документах окончания *-ы/-и* имеют единичные словоформы Т. мн. Из них *всеми образы* представляет собой устойчивый оборот, употребительный в языке XVIII в., о чем свидетельствуют данные НКРЯ. Окончание *-ы*, как показывает материал, появляется у заимствованного *штраф*, что в целом нехарактерно для языковой нормы. В выражении *всеми образы и меры* видим воздействие формы на *-ы* и на форму существительного женского рода, также наблюдавшееся в деловых памятниках XVII в. [Тарабасова, 1986, с. 105].

Предложный падеж множественного числа регулярно принимает окончания *-ах/-ях*, однако некоторые существительные употреб-

ляются с окончанием *-ех*, исконным (с учетом фонетических изменений гласных) для имен древних типов склонения на **ѣ* и **ї*. Формы на *-ех* в деловых письменных памятниках XVII в. преобладали над словоформами с *-ах* и выступали в качестве узальной нормы [Тарабасова, 1986, с. 115], однако в XVIII в. их соотношение изменилось.

В документах Войска Донского середины XVIII в. окончание *-ех* является единственным для отдельных слов, например *сбор*, *год*, у большинства же существительных отмечается *-ах/-ях*. В источниках того же периода из НКРЯ у слова *год* наблюдаются варианты, причем преобладает форма на *-ах* (56 вхождений против 17), то же обнаруживается у слова *место* (516 случаев употребления форм на *-ах* против 54), у слова *сбор* – только форма на *-ах*. Грамматики М.В. Ломоносова и А.А. Барсова в Т. мн. фиксируют единственное окончание *-ах/-ях*. Таким образом, документы фонда отражают в определенной степени узус деловой письменности предшествующего периода при явной тенденции к утверждению нормы *-ах/-ях*. Форма на *-ех* в парадигме имен существительных является уникальной и не вступает в отношение омонимии с другими формами.

Заключение

Морфологическая разметка части документов Войска Донского XVIII в. из фонда Михайловского станичного атамана Государственного архива Волгоградской области выявила наличие грамматических форм, нехарактерных для современного русского языка и, соответственно, необходимость корректировки программных инструментов разметки текстов в плане расширения возможных вариантов грамматического разбора омонимичных форм для последующего снятия омонимии вручную. Результаты количественного анализа фактов и их лингвистическая интерпретация показали в сфере грамматики, наряду с общими тенденциями развития и функционирования русского языка, особенности, характерные для деловой письменной речи в ее региональном варианте.

На примере падежных форм имен существительных обнаружено, что исследуемые

региональные источники в большей степени сохраняют узус делового языка предшествующего исторического периода, нежели официальные документы центральных учреждений и тексты художественной литературы, публицистики и др. В формах родительного и предложного падежей единственного числа отличие наблюдается в лексическом распределении характерных для общерусского языка вариантов и в их частотности. В дательном, творительном и предложном падежах множественного числа отмечаются архаичные формы, вышедшие из общего употребления либо характеризующиеся меньшей частотностью в иных источниках.

Данные выводы имеют на сегодня предварительный характер и могут быть уточнены в результате расширения обработанных в процессе создания лингвистического корпуса текстов. Пока не представляется возможным сделать вывод о специфике выявленных особенностей для деловой письменной речи именно в среде донского казачества. Несмотря на это, полученные сведения подтверждают значимость документов Войска Донского как источников истории русского языка и актуальность представления их в корпусе для проведения исследований в этой области.

СПИСОК ЛИТЕРАТУРЫ

- Барсов А. А., 1981. Российская грамматика Антона Алексеевича Барсова. М. : Изд-во МГУ. 776 с.
- Горбань О. А., Косова М. В., Шептухина Е. М., 2021. Структурная разметка деловых документов в диахроническом лингвистическом корпусе: проблемы и решения // Вестник Волгоградского государственного университета. Серия 2, Языкознание. Т. 20, № 4. С. 5–18. DOI: <https://doi.org/10.15688/jvolsu2.2021.4.1>
- Горбань О. А., Косова М. В., Шептухина Е. М., Дмитриева Е. Г., Сафонова И. А., 2020. Документы Войска Донского XVIII века: лингвистическое описание и тексты. Волгоград : Изд-во ВолГУ. 464 с.
- Еськова Н. А., 2008. Нормы русского литературного языка XVIII–XIX веков : Ударение. Грамматические формы. Варианты слов. Словарь. Пояснительные статьи. М. : Рукоп. памятники Древ. Руси, 2008. 960 с.
- Касаткин Л. Л., 1999. Современная русская диалектная и литературная фонетика как источник для истории русского языка. М. : Наука : Шк. «Яз. рус. культуры». 526 с.
- Комендантов А. С., Матвеев А. Г., Светлов А. В., 2019. Автоматизация морфологической разметки архивных документов // Математическая физика и компьютерное моделирование. Т. 22, № 4. С. 53–63. DOI: <https://doi.org/10.15688/mpcm.jvolsu.2019.4.4>
- Ломоносов М. В., 1755. Российская грамматика Михайла Ломоносова. СПб. : Тип. Императ. Акад. наук. 210 с.
- Ляшевская О. Н., Плунгян В. А., Сичинава Д. В., 2005. О морфологическом стандарте Национального корпуса русского языка // Национальный корпус русского языка: 2003–2005. Результаты и перспективы. М. : Индрик. С. 111–135.
- Осипов Б. И., Гейгер Р. М., Рогожникова Т. П., 1993. Язык русских деловых памятников XV–XVIII вв. : Фонетический, орфографический и стилистический аспекты. Омск : Ом. ун-т. 96 с.
- Савчук С. О., Сичинава Д. В., 2009. Корпус русских текстов XVIII века в составе Национального корпуса русского языка: проблемы и перспективы // Национальный корпус русского языка: 2006–2009. Новые результаты и перспективы / отв. ред. В. А. Плунгян. СПб. : Нестор-История. С. 52–69.
- Сень Д. В., 2014. Архив Войска Донского и история войскового делопроизводства: актуальные вопросы изучения // Научное наследие профессора А.П. Пронштейна и актуальные проблемы развития исторической науки (к 95-летию со дня рождения выдающегося русского ученого) : материалы Всерос. (с междунар. участием) науч.-практ. конф., г. Ростов-на-Дону, 4–5 апр. 2014 г. / отв. ред. М. Д. Розин, Д. В. Сень, Н. А. Трапш. Ростов н/Д : Фонд науки и образования. С. 484–495.
- Тарабасова Н. И., 1986. Явления вариативности в языке московской деловой письменности XVII в. / отв. ред. С. И. Котков. М. : Наука. 164 с.
- Филимонов Д. Ю., Светлов А. В., Горбань О. А., Косова М. В., Шептухина Е. М., 2020. Автоматизация процесса метаразметки архивных документов // Математическая физика и компьютерное моделирование. Т. 23, № 4. С. 56–68. DOI: <https://doi.org/10.15688/mpcm.jvolsu.2020.4.6>
- Segalovich I., 2003. A Fast Morphological Algorithm with Unknown Word Guessing Induced by a Dictionary for a Web Search Engine // Proceedings of the International Conference on Machine Learning: Models, Technologies and Applications. LasVegas : CSREAPress. P. 273–280.

ИСТОЧНИКИ

Акты, относящиеся к истории Войска Донского, собранные генерал-майором А. А. Лиши-

ным : в 3 т. / изд. Обл. правл. Войска Донского под ред. д. чл. Донского Обл. стат. ком. А. А. Карасева, Х. И. Попова. Новочеркасск : Тип. А. А. Карасева, 1891–1894. Т. 1. 1891. 373 с. ; Т. 2, ч. 1. 1894. XXXII, 368 с. ; Т. 2, ч. 2. 1894. XXII, 369–821, XXVI с. ; Т. 3. 1894. 480 с.

Михайловский станичный атаман // Государственный архив Волгоградской области. Ф. 332. Оп. 1.
Национальный корпус русского языка. 2003–2021.
URL: <http://ruscorpora.ru>

REFERENCES

- Barsov A.A., 1981. *Rossiyskaya grammatika Antona Alekseevicha Barsova* [The Russian Grammar of Anton Barsov]. Moscow, Izd-vo MGU. 776 p.
- Gorban O.A., Kosova M.V., Sheptukhina E.M., 2021. Strukturnaya razmetka delovykh dokumentov v diakhronicheskom lingvisticheskom korpusе: problemy i resheniya [Structural Markup of Official Documents in Diachronic Linguistic Corpus: Problems and Solutions]. *Vestnik Volgogradsko gosudarstvennogo universiteta. Seriya 2. Yazykoznanie* [Science Journal of Volgograd State University. Linguistics], vol. 20, no. 4, pp. 5-18. DOI: <https://doi.org/10.15688/jvolsu2.2021.4.1>
- Gorban O.A., Kosova M.V., Sheptukhina E.M., Dmitrieva E.G., Safonova I.A., 2020. *Dokumenty Voyska Donskogo XVIII veka: lingvisticheskoe opisanie i teksty* [Documents of the Don Army of the 18th Century: Linguistic Description and Texts]. Volgograd, Izd-vo VolGU. 464 p.
- Eskova N.A., 2008. *Normy russkogo literaturnogo yazyka XVIII–XIX vekov: Udarenie. Grammaticheskie formy. Varianty slov. Slovar. Poyasnitelnye statyi* [Norms of the Russian Literary Language of the 18th – 19th Centuries: Stress. Grammatical Forms. Variants of Words. Dictionary. Explanatory Articles]. Moscow, Rukop. pamyatniki Drev. Rusi Publ. 960 p.
- Kasatkin L.L., 1999. *Sovremennaya russkaya dialektnaya i literaturnaya fonetika kak istochnik dlya istorii russkogo yazyka* [Modern Russian Dialect and Literary Phonetics as a Source for the History of the Russian Language]. Moscow, Nauka Publ., Shk. «Yazyki russkoy kultury» Publ. 526 p.
- Komendantov A.S., Matveev A.G., Svetlov A.V., 2019. Avtomatizatsiya morfologicheskoy razmetki arkhivnykh dokumentov [Automation of Morphological Tagging of Archival Documents]. *Matematicheskaya fizika i kompyuternoe modelirovanie* [Mathematical Physics and Computer Simulation], vol. 22, no. 4, pp. 53-63. DOI: <https://doi.org/10.15688/mpcm.jvolsu.2019.4.4>
- Lomonosov M.V., 1755. *Rossiyskaya grammatika Mikhayla Lomonosova* [Mikhail Lomonosov's Russian Grammar]. Saint Petersburg, Tip. Imperat. Akad. nauk. 210 p.
- Lyashevskaya O.N., Plungyan V.A., Sichinava D.V., 2005. O morfologicheskom standarte Natsionalnogo korpusa russkogo yazyka [Russian Language National Corpus Morphological Standard]. *Natsionalnyy korpus russkogo yazyka: 2003–2005. Rezultaty i perspektivy* [National Corpus of the Russian Language: 2003–2005. Results and Prospects]. Moscow, Indrik Publ., pp. 111-135.
- Osipov B.I., Geyger R.M., Rogozhnikova T.P., 1993. *Yazyk russkikh delovykh pamyatnikov XV–XVIII vv.: Foneticheskiy, orfograficheskiy i stilisticheskiy aspekty* [The Language of Russian Business Monuments of the 15th – 18th Centuries: Phonetic, Spelling and Stylistic Aspects]. Omsk, Om. un-t. 96 p.
- Savchuk S.O., Sichinava D.V., 2009. Korpus russkikh tekstov XVIII veka v sostave Natsionalnogo korpusa russkogo yazyka: problemy i perspektivy [Corpus of Russian Texts of the 18th Century as Part of the National Corpus of the Russian Language: Problems and Prospects]. *Natsionalnyy korpus russkogo yazyka: 2006–2009. Novye rezultaty i perspektivy* [National Corpus of the Russian Language: 2006–2009. New Results and Prospects]. Saint Petersburg, Nestor-Istoriya Publ., pp. 52-69.
- Sen D.V., 2014. Arkhiv Voyska Donskogo i istoriya voyskovogo deloproizvodstva: aktualnye voprosy izucheniya [Archive of the Don Army and the History of Military Records Management: Current Issues of Study]. Rozin M.D., Sen D.V., Trapsh N.A., eds. *Nauchnoye nasledie professora A.P. Pronshcheyna i aktualnye problemy razvitiya istoricheskoy nauki (k 95-letiyu so dnya rozhdeniya vydavushchegosya rossiyskogo uchenogo): materialy Vseros. (s mezhdunar. uchastiem) nauch.-prakt. konf., g. Rostov-na-Donu, 4–5 apr. 2014 g.* [The Scientific Legacy of Professor A.P. Pronstein and the Actual Problems of the Development of Historical Science (To the 95th Anniversary of the Birth of an Outstanding Russian Scientist). Materials of All-Russian (with International Participation) Scientific and Practical Conference, Rostov-on-Don, 4–5 April 2014]. Rostov-on-Don, Fond nauki i obrazovaniya, pp. 484-495.
- Tarabasova N.I., 1986. *Yavleniya variativnosti v yazyke moskovskoy delovoy pismennosti XVII v.* [Phenomena of Variability in the Language of

- Moscow Business Writing of the 17th Century]. Moscow, Nauka Publ. 164 p.
- Filimonov D.Yu., Svetlov A.V., Gorban O.A., Kosova M.V., Sheptukhina E.M., 2020. Avtomatizatsiya protsessa metarazmetki arkhivnykh dokumentov [Automation of Archival Documents Meta Tagging]. *Matematicheskaya fizika i kompyuternoe modelirovanie* [Mathematical Physics and Computer Simulation], vol. 23, no. 4, pp. 56-68. DOI: <https://doi.org/10.15688/mpcm.jvolsu.2020.4.6>
- Segalovich I., 2003. A Fast Morphological Algorithm with Unknown Word Guessing Induced by a Dictionary for a Web Search Engine. *Proceedings of the International Conference on Machine Learning: Models, Technologies and Applications*. Las Vegas, CSREA Press, pp. 273-280.

SOURCES

- Karaseva A.A., Popova Kh.I., eds. Akty, odnosyashchiesya k istorii Voyska Donskogo, sobrannye general-mayorom A. A. Lishinym: v 3 t. [Acts Relating to the History of the Don Army, Collected by Major General A.A. Lishin. In 3 Vols.]. Novocherkassk, Tip. A.A. Karaseva, 1891–1894. Vol. 1. 1891. 373 p. ; Vol. 2, part 1. 1894. XXXII, 368 p. ; Vol. 2, part 2. 1894. XXII, 369–821, XXVI p. ; Vol. 3. 1894. 480 p.
- Mikhaylovskiy stanichnyy ataman [Mikhailovsky Stanitsa Ataman]. *Gosudarstvennyy arkhiv Volgogradskoy oblasti* [State Archive of Volgograd Region], f. 332, inv. 1.
- Natsionalnyy korpus russkogo yazyka. 2003–2021* [The Russian National Corpus]. URL:<http://ruscorpora.ru>

Information About the Authors

Oksana A. Gorban, Doctor of Sciences (Philology), Professor, Department of Russian Philology and Journalism, Volgograd State University, Prosp. Universitetsky, 100, 400062 Volgograd, Russia, oa_gorban@volsu.ru, <https://orcid.org/0000-0002-2345-3673>

Marina V. Kosova, Doctor of Sciences (Philology), Professor, Department of Russian Philology and Journalism, Volgograd State University, Prosp. Universitetsky, 100, 400062 Volgograd, Russia, mv_kosova@volsu.ru, <https://orcid.org/0000-0003-2854-8759>

Elena M. Sheptukhina, Doctor of Sciences (Philology), Professor, Department of Russian Philology and Journalism, Volgograd State University, Prosp. Universitetsky, 100, 400062 Volgograd, Russia, em_sheptuhina@volsu.ru, <https://orcid.org/0000-0002-8007-6042>

Andrey V. Svetlov, Candidate of Sciences (Physics and Mathematics), Associate Professor, Department of Mathematical Analysis and Function Theory, Volgograd State University, Prosp. Universitetsky, 100, 400062 Volgograd, Russia, andrew.svetlov@volsu.ru, <https://orcid.org/0000-0002-8764-6132>

Информация об авторах

Оксана Анатольевна Горбань, доктор филологических наук, профессор кафедры русской филологии и журналистики, Волгоградский государственный университет, просп. Университетский, 100, 400062 г. Волгоград, Россия, oa_gorban@volsu.ru, <https://orcid.org/0000-0002-2345-3673>

Марина Владимировна Косова, доктор филологических наук, профессор кафедры русской филологии и журналистики, Волгоградский государственный университет, просп. Университетский, 100, 400062 г. Волгоград, Россия, mv_kosova@volsu.ru, <https://orcid.org/0000-0003-2854-8759>

Елена Михайловна Шептухина, доктор филологических наук, профессор кафедры русской филологии и журналистики, Волгоградский государственный университет, просп. Университетский, 100, 400062 г. Волгоград, Россия, em_sheptuhina@volsu.ru, <https://orcid.org/0000-0002-8007-6042>

Андрей Владимирович Светлов, кандидат физико-математических наук, доцент кафедры математического анализа и теории функций, Волгоградский государственный университет, просп. Университетский, 100, 400062 г. Волгоград, Россия, andrew.svetlov@volsu.ru, <https://orcid.org/0000-0002-8764-6132>