

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ
ГИДРОМЕТЕОРОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»


Кафедра русского языка и литературы

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

На тему: Изучение «Жития Феодосия Печерского» средствами корпусной
лингвистики

Исполнитель Лозовая Анна Андреевна
(фамилия, имя, отчество)

Руководитель кандидат филологических наук
(ученая степень, ученое звание)
Непоклонова Елена Олеговна
(фамилия, имя, отчество)

«К защите допускаю»
Заведующий кафедрой 
(подпись)
кандидат педагогических наук, доцент
(ученая степень, ученое звание)
Кипнес Людмила Владимировна
(фамилия, имя, отчество)

1 июня 2022 г.

САНКТ-ПЕТЕРБУРГ
2022

Содержание

ВВЕДЕНИЕ	3
ГЛАВА 1. Теоретические основы корпусного изучения текстов.	6
1.1. Корпусная лингвистика как научная дисциплина.	6
1.2. Терминологический аппарат корпусной лингвистики.	8
1.3. Национальный корпус русского языка: история создания и применения	11
1.4. Корпусные методы исследования.	26
1.5. Корпусные методы в изучении художественной литературы	30
ГЛАВА 2. Корпусные методы в изучении древнерусской литературы	34
2.1. Специфика древнерусского текста и его изучения	34
2.2. Древнерусская литература в НКРЯ: состав, способ представления, разметка	44
2.3. Структура и семантика «Жития Феодосия Печерского» средствами НКРЯ	46
Заключение	58
БИБЛИОГРАФИЧЕСКИЙ СПИСОК	59

ВВЕДЕНИЕ

Текст является одним из основных источников языкового материала. Он нужен для проведения филологических и лингвистических исследований. Сделать вывод о языковом явлении, которое интересует исследователя, позволяет сделать анализ большого количества текстов. Сегодня огромное количество текстов находятся в свободном электронном доступе благодаря техническому прогрессу. Для работы с большими объемами текстов во всем мире стали создаваться коллекции текстов. Эти коллекции являются специально отобранными, размеченными по различным лингвистическим параметрам и обеспеченными системой поиска. Они называются “лингвистическими корпусами”. Исследованиями в данной области занимается такой раздел прикладной лингвистики, который разрабатывает общие принципы построения и использования лингвистических корпусов, как корпусная лингвистика.

Актуальность исследования. Возможность пользоваться обширными электронными ресурсами значительно облегчила процесс сбора материала в лингвистических исследованиях. Безусловно, необходимость и возможность обработки множества текстов с целью извлечения из них литературоведческих, лингвистических и прочих данных привели к бурному росту электронных ресурсов. К таким ресурсам можно отнести лингвистически аннотированные корпуса текстов, что обуславливает актуальность изучения данной темы. Так, любое лингвистическое исследование в той или иной мере опирается на анализ языкового материала, языковых данных. Развитие современных интеллектуальных программных систем, которые предназначены для обработки текстов на естественном языке, также требует большой экспериментальной лингвистической базы. На сегодняшний день спрос на корпусные данные совпадает с современными техническими возможностями. В современной теоретической лингвистике

корпус является как мощным инструментом исследования языка, так и новой идеологией, которая ориентирует исследователя на текст как главный объект теоретической рефлексии. Так, с помощью корпуса стало возможно быстрее и эффективнее решать известные науке задачи, а также ставить принципиально новые задачи, ранее практически невыполнимые из-за их трудоемкости.

Объектами исследования в рамках данной работы являются средства корпусной лингвистики.

Предмет исследования – структура и семантика «Жития Феодосия Печерского» средствами НКРЯ.

Цель исследования – изучение «Жития Феодосия Печерского» средствами корпусной лингвистики.

Задачи исследования:

1. Исследование корпусных методов изучения художественной литературы;
2. Характеристика методов корпусной лингвистики в изучении художественных текстов;
3. Характеристика корпусных методов в изучении древнерусской литературы;
4. Анализ структуры и семантики «Жития Печерского» средствами НКРЯ.

В работе использованы такие *методы* исследования, как метод компонентного анализа, контекстного анализа, сравнительный, описательный, лингвистический и трансформационный методы.

Материалом исследования послужил электронный онлайн-корпус русских текстов «Национальный корпус русского языка» (НКРЯ).

Теоретической базой исследования послужили труды таких ученых, как В.А. Плунгян, Д.В. Сичинава, Б. В. Орехов, А.А. Пичхадзе и другие.

Теоретическая значимость заключается в углублении понимания и уточнении содержания понятий корпусных методов изучения

художественной и древнерусской литературы на примере анализа «Жития Феодосия Печерского». *Практическая значимость* работы состоит в том, что материалы проведенного исследования могут использоваться в системе профессионального образования направления подготовки «Филология», «Лингвистика», в процессе обучения русскому языку в средних и высших школах, при написании курсовых и выпускных квалификационных работ, а также для дальнейшего исследования данной темы.

Структура работы. ВКР состоит из введения, двух глав, заключения и библиографического списка.

Во введении обосновываются выбор темы ВКР, ее актуальность, теоретическая и практическая значимость, формулируются цель, задачи, определяются объект и предмет исследования, методы анализа языкового материала.

В первой главе рассматриваются история создания и применения русскоязычных корпусов текста; дается характеристика структуры и функциональных возможностей НКРЯ; анализируются методы корпусной лингвистики в изучении художественных текстов.

Во второй главе представлена характеристика специфики древнерусского текста и его изучения; рассматриваются состав, способ представления и разметка древнерусской литературы в НКРЯ; анализируются структура и семантика «Жития Феодосия Печерского» средствами НКРЯ.

Заключение содержит обобщающие выводы по результатам проведенного исследования.

Список использованной литературы составлен в алфавитной последовательности и включает в себя 75 наименований.

ГЛАВА 1. Теоретические основы корпусного изучения текстов

1.1. Корпусная лингвистика как научная дисциплина

Главным материалом любого лингвистического и – шире – филологического исследования является текст. Именно текст является основным источником знаний о языке. Под «текстом» мы понимаем любые речевые произведения, устные или письменные, осуществленные в той или иной коммуникативной ситуации. Долгое время в лингвистической науке основное внимание уделялось не текстам как таковым, а языку, противопоставленному речи, как самоценный объект исследования. Однако развитие компьютерных технологий, позволивших обрабатывать большое количество текстов с целью получения данных, касающихся того или иного языкового феномена, предоставило исследователям новые возможности изучения функционирования языка.

Исследование текстов предполагает отбор материала и его обработку с целью выявления определенных закономерностей. Традиционные лингвистические методы позволяют выполнять эти операции, однако методы компьютерной обработки текстов существенно повышают их эффективность благодаря значительному увеличению объема материала и точности подсчетов. Компьютерный анализ текстов тем самым является не только средством оптимизации исследовательского процесса, но и средством установления языковых закономерностей на основе эмпирических данных, недоступных в случае традиционного сбора данных. Не случайно математический лингвист Р.Г. Пиотровский отмечал, что достоверность исследований языковых закономерностей прямо пропорциональна объему рассмотренных текстов. Кроме того, доступность языковых данных оказала существенное влияние на требования, предъявляемые к доказательной базе лингвистических исследований. Так, если раньше исследователь осуществлял выборку способом перебора источников и осознавал

случайность выборки, ее зависимость от большого количества внешних факторов (доступность источников, их качество и объем и т.д.), то с появлением электронных хранилищ текстов отбор текстов превращается в теоретически осмысленную и практически выверенную процедуру. Доступность большого количества разнородных текстов поставило перед лингвистами новые задачи: как обрабатывать такое количество разнородных текстов.

Для организации работы с корпусными данными, в частности – обеспечения компактности представления данных, организации эффективного поиска материала по заданным критериям, в корпусной лингвистике разрабатываются специфические методы исследования текстов, основанные на соединении количественных методов статистического анализа с организацией поиска материала на основе аннотирования текстов. Так, с одной стороны, рост электронных ресурсов сформировал новые запросы исследователей относительно изучаемых текстов, а с другой, возможность анализа большого массива электронных ресурсов формировала новые подходы к хранению и обработке этих ресурсов, новые требования к их аннотированию в рамках лингвистических корпусов. Новые задачи поставили вопрос о корректности получаемых данных, о принципах отбора текстов для корпусов, способы разметки текстов и т.д. Так, в процессе отбора и аннотирования текстов формировались основные цели и задачи корпусной лингвистики. В настоящее время корпусная лингвистика решает широкий спектр задач: от разработки принципов отбора текстов корпусов, параметров аннотирования текстов до описания спектра задач, которые можно решать средствами корпусной лингвистики в различных областях научного знания.

Таким образом, история разработки и создания корпусов русскоязычных текстов была связана с постановкой в рамках корпусной лингвистике следующих вопросов:

- Какие принципы должны лежать в основе создания корпусов в процессе отбора текстов и организации разметки;

- Как должна быть организована разметка лингвистических параметров текста (жанров, конструкций, морфологических характеристик слов и т.д.), чтобы обеспечивать доказательную базу исследований;
- Какие лингвистические задачи, а также задачи из других областей гуманитарного знания, могут быть решены с помощью корпусов;
- Как достичь быстроты и удобства использования поискового интерфейса к корпусу (т.е. программ, обеспечивающих поиск), полноты предоставляемого материала, полностью соответствующего задачам исследования.

Так, в процессе решения этих вопросов были разработаны интерфейсы и экспертные лингвистические разметки, позволяющие находить данные, соответствующие строго определенным критериям. При этом важно, что любой человек, обратившийся к корпусу, получит те же данные, что и все остальные пользователи; тем самым корпус обеспечивает экспериментальный характер исследования, для которого характерна воспроизводимость и верифицируемость полученных данных.

1.2. Терминологический аппарат корпусной лингвистики.

Поскольку корпусная лингвистика является достаточно молодым направлением в лингвистической науке, ее терминология находится на стадии активной разработки. Основной единицей корпусной лингвистики является понятие лингвистического корпуса. Как отмечалось выше, *корпус* – это определенная выборка, основанная на заданных критериях, связанных с конкретным аспектом лингвистических исследований. Так, лингвистическим (языковой) корпусом текстов является массив языковых данных для решения конкретных лингвистических задач, обладающий следующими признаками: большой и структурированный, в электронном виде, унифицированный и размеченный, а также филологически компетентный

Интересующая лингвиста проблемная область может иметь различную широту и характеристики: от лингвистических особенностей одного текста

до особенностей языка значительного исторического периода. В связи с разнообразием проблемной области актуален вопрос отбора текстов: их типология и количество. В результате возникает противоречие между требованиями полноты, разнообразия языковых явлений и репрезентативности, предполагающей отбор наиболее характерных явлений языка. С одной стороны, лингвисты стремятся собрать максимально возможное количество текстов, соответствующих лингвистическим задачам корпуса, а с другой – сохранить его репрезентативность. Никакой корпус не может охватить все специфические особенности языка, все возможные реализации языковых феноменов. Поэтому корпус является выборкой из проблемной области на основе заданных критериев. Репрезентативность корпуса определяется тем, что любая выборка отражает определенные параметры исследуемого явления в определенной пропорции – в той же, что и в самом языке.

В связи с этим в лингвистике сложились понятия *сбалансированный корпус* и *мониторный корпус*. В первом типе корпусов отбираются тексты, представляющие тексты разных жанров, а также письменные и устные. Следовательно, пополнение корпуса новыми текстами требует тщательной проверки в соответствии с выявленными пропорциями. В отличие от сбалансированных корпусов мониторные корпуса не требуют такого тщательного отбора текстов, они непрерывно пополняются без учета баланса в языке текстов разных типов. Это связано с особой позицией разработчиков мониторных корпусов, полагающих, что объективность статистических данных будет достигнута благодаря большому объему корпуса. К мониторным корпусам можно отнести современные веб-корпусы, большой объем которых создается за счет технологий сбора текстов из сети Интернет.

Важным понятием корпусной лингвистики является *разметка*, или *аннотация* – дополнительная лингвистическая и металингвистическая информация о текстах и словоупотреблениях. Если метаразметка предоставляет пользователю значительную экстралингвистическую

информацию о тексте (жанр, автор, сфера бытования, тематика и др.), то лингвистическая разметка определяется основными уровнями лингвистического описания текста: фонетическим, морфологическим, синтаксическим, дискурсивным и др.

Для осуществления поиска используются специальные программы, которые получили название *корпусных менеджеров*. Это программы, благодаря которым обеспечивается сортировка результатов поиска, становятся возможными статистические данные, выдача списков контекстов, содержащих необходимое языковое явление. При использовании корпусных менеджеров производится поиск словоформ, в том числе по *леммам* (это подразумевает произведение поиска всех форм одной лексемы в тексте, а также непосредственно поиск словосочетаний, в том числе и тех, у которых есть разрыв). Минимальная значимая единица текста, которая подвергается автоматической обработке – это *словоформа*, или *токен*. С помощью созданной в корпусе разметки исследователь можно осуществлять поиск по набору различных морфологических признаков (например, конкретных словосочетаний, включающих определенные части речи, проверять сочетаемость слов разных частей речи) и т.д. С помощью метаразметки исследователь может создать собственный *подкорпус текстов* (например, одного автора, одного произведения, жанра, тематики, времени и т.д.

Результат выдачи в корпусной лингвистике получил название *конкорданс* – т.е. множество найденных текстов, содержащих запрашиваемое языковое явление. Для каждого найденного примера указывается источник, также часто предоставляется статистическая информация: частота по корпусу, распределение по жанрам, частота сочетаемости и др.

1.3. Национальный корпус русского языка: история создания и применения

На сегодняшний день в научный оборот лингвистов активно внедряется такое понятие, как «корпус текстов - унифицированный, структурированный и размеченный массив языковых (речевых) данных в электронном виде, предназначенный для определенных филологических и, более широко, гуманитарных изысканий» [20]. Корпусная лингвистика формируется на основе приведенного понятия. Необходимо также отметить, что корпус текстов на данный период времени изучается рядом исследователей «как сложно организованная онтология речевой деятельности, отражающая в себе все жанровое разнообразие представленного в нем рода словесности» [58].

Представляется важным проанализировать историю создания и применения русскоязычных корпусов текстов. Созданию русскоязычных корпусов предшествовали разработки корпусов текстов в ряде европейских стран. Первым национальным корпусом, ставшим своего рода образцом для создания всех последующих корпусов текстов, стал Британский национальный корпус (BNC). В дальнейшем появились национальные корпуса для различных европейских языков: немецкого, финского, чешского и др. Отечественная корпусная лингвистика возникает в 1980-е гг. и связывается с появлением работ А. П. Ершова и В. М. Андрющенко [2]. Одновременно в Институте русского языка имени В. В. Виноградова был создан РАН Отдел машинного фонда, занимавшийся, в частности, переводом русских литературных текстов в электронную форму. К сожалению, технических и организационных возможностей для того, чтобы создать полноценный машинный фонд текстов русского языка, в тот период не было, поэтому данный проект не привел к созданию первого корпуса русского языка, а лишь способствовал собранию большого количества различных русскоязычных текстов. Вместе с тем, специалист в области корпусной

лингвистики Д. В. Сичинава в дальнейшем отмечал, что «многие материалы, собранные и обработанные в Отделе машинного фонда, особенно по авторам XX века, впоследствии удалось использовать и при работе над Национальным корпусом русского языка» [67].

Поскольку единого авторитетного для всех корпуса русских текстов долгое время не существовало, лингвисты, как правило, опирались на собственные электронные коллекции русских текстов, а также использовали поиск в Интернете, который также содержит большое количество русскоязычных текстов. Однако Интернет не мог выполнять функции корпуса, поскольку тексты представлены в нем случайным образом, их отбор обусловлен предпочтениями активных пользователей, а отсутствие разметки не позволяет осуществлять поиск в соответствии с поставленными лингвистическими задачами.

Существовали также общедоступные корпуса русского языка, созданные зарубежными славистами, например, самый представительный корпус русскоязычных текстов, созданный в Финляндии в университете Уппсалы – Уппсальский корпус. В нем содержалось около 1 млн. слов, и в равных пропорциях были представлены художественные и информативные тексты периода 1960-1988 гг. Также был достаточно известен Тюбингенский корпус, в котором впервые появилась морфологическая разметка. Однако эти корпуса были ограничены в количестве представленных речевых жанров, а отсутствие или недостаточная разработанность части речевой и пр. аннотации существенно ограничивало возможности их использования.

В 2000 г. С.А. Шаров — один из исследователей, которые впоследствии предложили саму идею создания представительного Национального корпуса русского языка, — представил доклад на тему «Введение в проблематику корпусной лингвистики и форматов хранения текстов (в первую очередь, русских литературных текстов)». В своем докладе он впервые сформулировал основные задачи, стоящие перед создателями корпуса русского языка и основные принципы, которые должны лежать в

основе разработки корпуса. В 2001 году по инициативе В. А. Плуменя началось формирование коллектива исследователей, задачей которых являлась разработка единого национального корпуса русского языка. Предполагалось, что основной (если не единственной) его разметкой станет морфологическая, с возможностью снятия грамматической омонимии. Кроме того, был поставлен вопрос о критериях отбора текстов, представляющих корпус: ранее корпус не мыслился как репрезентативный по отношению к генеральной совокупности русских текстов. Первоначально планировалось, что в корпус должны войти прежде всего литературные тексты, культурно значимые и написанные без нарочитого «экспериментаторства». Таким образом, одним из главных вопросов в процессе создания НКРЯ являлся поиск критериев отбора текстов, способных представлять русский язык. Обсуждалось также название корпуса: предлагались такие варианты, как «Стандартный корпус русского языка», «Представительный корпус русского языка», «Корпус ЦЛД—МГУ» [67]. К работе над корпусом привлекались не только профессиональные лингвисты, но и студенты московских и других вузов.

В 2003 г. была открыта программа РАН «Филология и информатика», в рамках которой была окончательно разработана концепция корпуса русского языка. Был учтен опыт корпусной лингвистики других стран, в частности принципы создания Британского национального корпуса. Так, В.Б. Касевич, занимавшийся проблемами корпусной лингвистики, предложил назвать проект, опираясь на мировую практику, «Национальный корпус русского языка». Именно данное название утвердилось в дальнейшем.

Концепция Национального корпуса расширялась в разных направлениях. С конца 2002 года корпус со снятой омонимией стал доступен для поиска. В дальнейшем Национальный корпус привлекал интерес IT-специалистов, лингвистов, преподавателей и т.д. Уже в первые месяцы после официального открытия Корпус стал предметом рецензий в профессиональной компьютерной среде, для которой теоретическая

лингвистика представляла лишь косвенный интерес. Информация о сайте, несмотря на отсутствие открытого доступа, распространилась среди лингвистов достаточно быстро (причем не только в России). Так, уже весной 2003 года он регулярно использовался в преподавании русского языка в Карловом университете Праги. Днем открытия сайта НКРЯ стало 29 апреля 2004 года. Незадолго до этого Корпус был представлен на конференциях, организованных филологическим факультетом МГУ и ИМЛИ РАН [67].

В Национальном корпусе русского языка представлены тексты самых разных сфер, жанров и функциональных стилей. Классификация специальной литературы осуществляется по определенной тематике, а художественной литературы - по авторам [4].

Особенностью отечественной корпусной лингвистики долгое время был чисто теоретический, «проектирующий» характер: в работах подробно обсуждалось то, каким образом должна формироваться архитектура русского электронного корпуса, какая степень филологической тщательности в нем желательна, а какая, напротив, представляется излишней; какую лингвистическую информацию следует в нем указывать; для решения каких задач он будет пригоден и пр. Так первоначально сложилась теоретическая база корпусных исследований. Вслед за ней, несколько медленнее, начала формироваться сфера практических корпусных исследований. Неотъемлемой чертой современных корпусных исследований во всем мире является постоянная опора на конкретную практику создания корпуса, на реальный опыт работы, результат которой общедоступен для любого заинтересованного в этой проблематике специалиста [49]. В настоящее время на сайте НКРЯ есть раздел «Публикации по Национальному корпусу русского языка», где представлены работы, посвященные самым разнообразным направлениям использования НКРЯ: в лингвистике, литературоведении, методике преподавания языков, переводческой деятельности и т.д.

Так, в частности, во многом именно под влиянием бурного развития корпусой лингвистики произошло формирование современных моделей языка. Владимир Александрович Плуноян писал, что «для этих новых моделей характерна в первую очередь «смена теоретических приоритетов — произошел переход от «системы» к «узусу» и от «языка» к «речи»» [50]. В лингвистических теориях на протяжении веков в качестве базового постулата утверждалось принципиальное фундаментальное противопоставление языка и речи. При этом центральным объектом лингвистической науки всегда выступал язык, языковая компетенция, а речь, особенно спонтанная, воспринималась как производное от языка, нечто вторичное и менее информативное для изучения, представляющее собой продукт отклонения от языковой нормы. Однако с развитием корпусной лингвистики по мере накопления данных о функционировании языка в разнообразных сферах его употребления внимание исследователей переориентируется на тексты как главные объекты изучения. Большое количество текстов, быстрота поиска тех или иных контекстов употребления слов переориентировали лингвистов непосредственно на изучение текстов, важно отметить, на естественном языке, «представленных в недостижимом прежде объеме». В.А. Плуноян указывал в одной из своих работ, что «последствия этого неожиданного приобретения», определили как лингвистическую практику, так и лингвистическую теорию современной науки, а также дальнейшие перспективы их развития [50]. Так, в настоящее время не угасают споры, является ли корпусное исследование всего лишь одним из методов или перед нами самостоятельная методология, имеющая свой базис, философию, комплекс методов и приемов исследования и т.д.

Также стоит отметить, что на ранних этапах своего существования корпус пополнялся исключительно текстами, относящимися к сфере художественной литературы: это объяснялось традиционной «литературоцентричностью» русской культуры, предпочтением «хороших» литературных примеров в лингвистических работах, а также большим

стилистическим и лексическим разнообразием художественной литературы по сравнению, например, с газетными статьями того же объема.

В настоящее время принята концепция архитектуры НКРЯ (в формировании и которой большую роль сыграл один из организаторов проекта С.А. Шаров), представляющая собой в общих чертах компромисс между традиционной «литературоцентричностью» русского лингвистического материала и априорным принципом пропорциональной репрезентативности, согласно которому жанры и типы текстов в корпусе должны быть представлены пропорционально генеральной совокупности («популяции») текстов для данного языка [44]. Так, НКРЯ постоянно пополняется не только новыми произведениями литературы и нехудожественными письменными текстами, но и материалами устной речи, такими как записи радио и телепередач, публичных встреч различных специалистов с аудиторией и т.д. Устный подкорпус НКРЯ постоянно пополняется и записями, сделанными волонтерами, которые записывают и расшифровывают свою и чужую устную речь.

В настоящее время художественная литература составляет 41,5% от общего объема Корпуса (в словоупотреблениях), нехудожественные тексты – остальные 58,5% [43]: в дальнейшем планируется изменение этого соотношения в сторону уменьшения процента художественных текстов (но не ниже, чем до 30– 35%). По сравнению с национальными корпусами других языков в НКРЯ представлен самый высокий процент художественной литературы, в корпусах других языков доля художественных текстов составляет не более 20 %, причем в эти 20% входят такие жанры, как мемуары, письма различных эпох и др.

Таким образом, принцип пропорциональности при практической разработке корпусов является универсальным. В современных исследованиях, посвященных вопросам репрезентативности корпуса, подчеркивается, что «ведущим принципом разработки корпусов для большинства исследований является включение в корпус всего спектра

лингвистического варьирования, существующего в данном языке» [74]. Для достижения как репрезентативности, так и пропорциональной сбалансированности корпуса важна задача точного определения «категорий» текстов, используемых в нем.

Точная типология текстов (например, «письмо», «мемуары», «проповедь», «повесть») актуальна в связи с практическими целями, которые стоят перед исследователями, работающими с корпусами текстов. Все чаще предметом исследования становится языковой материал, требующий статистических и сравнительно-статистических данных. Сравнительная статистика становится основанием для аргументации и выводов в различных исследованиях. Так, сравнительная статистика определенных параметров текстов разных типов позволяет выявить связи тех или иных лингвистических параметров с жанровотематическим принципом текста. (так, например, в работах Д. Байбера обсуждаются вопросы типологии англоязычных текстов на основе морфологических и синтаксических параметров [75], а в статьях ряда отечественных лингвистов [66], обнаруживается, например, связь между частотой использования определенных частей речи, морфем, синтаксических конструкций с определенными типами текстов, например, между частотой использования предлогов и частиц и жанровотематическим типом текста.

Помимо Национального корпуса русского языка в отечественной лингвистике разрабатывались и другие корпуса русскоязычных текстов. Например, в настоящее время развиваются веб-корпуса – корпуса текстов, которые во много раз превышают количество словоупотреблений в национальных корпусах, оснащенные автоматической разметкой. Так, при поддержке компании АBBYY разработан Генеральный Интернет-корпус Русского языка (ГИКРЯ). По своему составу этот корпус превышает все известные корпуса русскоязычных текстов, он содержит около 20 миллиардов слов. Также весьма активно развиваются отдельные разновидности корпусов, в частности ориентированные на

специализированные направления лингвистических исследований. Например, для исследования дискурсов создаются корпуса с дискурсивной разметкой. Так, например, существуют корпуса, в которых разметка связана с теорией риторических структур. Также развиваются корпуса, включающие несколько языков, ориентированные на определенный аспект лингвистических исследований. Например, проект корпуса с единой синтаксической разметкой, содержащего тексты на разных языках, в том числе и достаточно редких. Активно развиваются параллельные корпуса, включающие тексты на двух и более языках. С помощью таких корпусов разрабатываются методики эффективного изучения иностранных языков, поскольку подобный корпус дает возможность соотнести два высказывания – на родном и чужом языках, что облегчает поиск эквивалентов. Параллельные корпуса незаменимы и при сборе данных для усовершенствования машинного перевода, поскольку они предоставляют большое количество соотносимых контекстов употребления единиц языка. Среди новых типов корпусов следует отметить корпуса, состоящие из текстов, созданных людьми на неродном языке. С помощью таких корпусов изучают лингвистические вопросы, лежащие на пересечении лингвистики и психологии, культурологии и др. дисциплин. Известен также корпус СинТагРус, разработанный в Лаборатории машинного перевода ИППИ РАН – это корпус, содержащий глубоко проработанную синтаксическую разметку.

Однако основным корпусом, служащим материалом для исследования русского языка, является НКРЯ. Ведущая роль данного корпуса в лингвистических исследованиях связана не только с его большим объемом (более 200 млн. словоупотреблений), но и с удобством разметки – морфологической и семантической. Так, подкорпус, содержащий 6 млн. словоупотреблений, предоставляет разметку со снятой морфологической омонимией. Это означает, что каждая словоформа имеет отдельный морфологический тег – т.е. морфологическую характеристику, которую выбранная словоформа имеет в определенном контексте.

Таким образом, разнообразные корпуса текстов позволяют не только решать различные исследовательские задачи, лежащие в плоскости лингвистической проблематики, но и ставить такие, которые ранее невозможно было решить из-за их трудоемкости. Это касается прежде всего вопросов, связанных с изучением микроэволюции языка, взятой в достаточно больших временных интервалах, от одного до нескольких столетий. До создания корпусов недостаточный объем данных не позволял лингвистам наблюдать малозаметные изменения в эволюции значений слов, их сочетаемости, произношении; невозможно было проследить изменения частоты употребления тех или иных грамматических вариантов слов, проследить судьбу неологизмов, синтаксических конструкций. В настоящее время подобные сведения, полученные благодаря корпусам, ложатся в основу разнообразных исследований. Так, в частности, корпусно-ориентированными становятся современные грамматические и лексикографические описания языков, поскольку любая лингвистическая верификация в настоящее время базируется в том числе и на проверке языкового явления на материале некоторого корпуса данного языка.

Таким образом, доступность данных, предоставляемых корпусом, оказывает влияние на представления о доказательной базе исследования. До создания корпусов для исследования достаточно было предъявить достаточно небольшую репрезентативную выборку данных, представляющую собой, по сути, случайно найденные примеры. Теперь поиск данных становится методологически выверенной процедурой, напрямую связанной с верификацией предъявленных гипотез. Различия между случайной выборкой и процедурами корпусных исследований связаны со способами представления текстов в корпусах, обеспечивающими корректность выводов, полученных на материале представленных текстов. Именно поэтому упоминавшиеся выше дискуссии о принципах отбора текстов и организации разметки были актуальны в процессе разработки и создания корпусов: ведь именно полнота репрезентативности корпуса и

принципы экспертной разметки обеспечивают доказательный характер представленного материала.

Так, в процессе решения этих вопросов были разработаны интерфейсы и экспертные лингвистические разметки, позволяющие находить данные, соответствующие строго определенным критериям. При этом важно, что любой человек, обратившийся к корпусу, получит те же данные, что и все остальные пользователи; тем самым корпус обеспечивает экспериментальный характер исследования, для которого характерна воспроизводимость и верифицируемость полученных данных.

Таким образом, главная задача корпуса – отражение состояния того или иного языка с максимальной полнотой благодаря аннотированию наибольшего количества языковых феноменов и эффективным корпусным менеджерам, способствующим удобству работы с корпусными данными. В связи с этой задачей постепенно сложились универсальные стандарты и технологии, позволяющие создавать представительные корпуса, предоставляющие данные и инструментарий для разнообразных эмпирических исследований. Разнообразие и объем корпусных данных оказались востребованы не только лингвистами, но и представителями самых разнообразных лингвистических дисциплин, таких как литературоведение, культурология, культурная антропология, межкультурная коммуникация, методика преподавания иностранных языков и многие другие.

Работа над созданием НКРЯ [43] проводилась в рамках программы «Филология и информатика» историко-филологического отделения Российской академии наук. В апреле 2004 г. НКРЯ был официально размещен в открытом доступе. Компания «Яндекс» осуществляла техническую поддержку корпуса: он располагался на сервере компании, которая также предоставляла многообразную технологическую поддержку в размещении программных продуктов НКРЯ. В апреле 2004 г. объем Корпуса составлял 19 млн словоупотреблений, через полгода он достиг 35 млн словоупотреблений, а в настоящее время насчитывает 1, 5 млрд слов.

Созданию НКРЯ предшествовал достаточно долгий подготовительный этап многоаспектного проектирования, в которое входили следующие задачи:

- Подбор качественных электронных текстов;
- Обоснование критериев отбора изданий письменных текстов;
- Формат представления текстов;
- Морфологическая, синтаксическая и лексико-семантическая разметка;
- Представление метаинформации о текстах;
- Усовершенствование поиска и др.

Работа над структурой и содержанием Корпуса продолжается до сих пор. В 2001–2002 гг. предложения по дальнейшей разработке НКРЯ постоянно проводились на семинарах Центра лингвистической документации при Московском центре непрерывного математического образования. Постепенно исследовательские задачи, связанные с Корпусом, усложнялись, и в 2003 г оформилась структура проекта: основная организация, осуществляющая координационную и организационную деятельность, связанную с Корпусом, обеспечивающая его финансирование РАН, стал Институт русского языка (Машинный фонд Института). «Также над корпусом работал Отдел лингвистических исследований ВИНТИ РАН, компания «Яндекс» и другие организации» [43]. В дальнейшей разработке Корпуса участвуют сотрудники, студенты и аспиранты различных лингвистических центров Москвы: МГУ им. М.В. Ломоносова, РГГУ, Институт языкознания РАН, Институт славяноведения РАН, Институт востоковедения РАН и др. Создание корпуса потребовало длительной совместной работы представителей самых разных областей языкознания: морфологии, лексической семантики, синтаксиса, стилистики, орфоэпии и др.; необходима была работа специалистов по метаинформации и типологии текстов, по переводоведению, устной речи, диалектологии и множество других. Одновременно были востребованы программисты, разработчики

Интернет-сайтов, без которых было бы невозможно реализовать разработки лингвистов.

Работа над НКРЯ стала одним из глобальных лингвистических проектов, потребовавших слаженной работы специалистов самых разных сфер на регулярной основе. Научно-практические семинары, посвященные разработке НКРЯ, позволяли вырабатывать конкретные решения больших групп специалистов и способствовали их быстрой практической реализации.

□ Также НКРЯ постоянно является предметом обсуждения среди лингвистов как в России, так и непосредственно за рубежом. Среди участников проекта следует выделить таких выдающихся лингвистов, как С.А. Шаров, В.А. Плунгян, Е.В. Рахилина, Т.И. Резникова, Д.В. Сичинава, О.Н. Ляшевская, Т.И. Резникова и др. Проблематика разработки НКРЯ регулярно обсуждается на таких мероприятиях, как Заседания Секции литературы и языка Отделения историко-филологических наук РАН, на международных конференциях «Языкознание *subspecie* русистики» в Институте русского языка им. В.В. Виноградова, «CORPORA-2004» в Петербургском государственном университете, на II Международном конгрессе исследователей русского языка «Русский язык: исторические судьбы и современность, на зарубежных славистических конференциях – в Лувене (Бельгия) и в Гранаде (Испания), во многих университетах России и других стран, например, в Томском и Казанском государственных университетах, в Карловом, Тюбингенском, Лидском и Суррейском университетах. Отдельно следует отметить крупнейшую в России ежегодную конференцию по компьютерной лингвистике под названием «Диалог». В рамках данной конференции регулярно поднимаются вопросы о разработке и усовершенствовании корпусов.

В процессе такого длительного и многоаспектного обсуждения проблематики НКРЯ сформировался большой фонд разработок, наблюдений, замечаний и предложений по усовершенствованию НКРЯ, которые регулярно учитываются в процессе дальнейшей разработки корпуса. Процесс

разработки НКРЯ отражен в публикациях многих известных изданий, например, в журнале «Научно-техническая информация», в котором публиковались различные исследования, посвященные теоретическим и практическим вопросам разработки корпуса [8] и дальнейшим перспективам его использования [21], как в России, так и за рубежом.

Структура НКРЯ

Структура НКРЯ регулярно корректируется, поскольку, как отмечалось ранее, развитие корпусной лингвистики вызывает непрерывную исследовательскую рефлексию и продолжение разработки устройства языковых корпусов. Семинары по компьютерной и, в частности, корпусной лингвистике способствуют появлению новых идей, касающихся как усовершенствования поиска в корпусе, так и состава текстов, способов аннотирования, общей структуры НКРЯ.

В настоящее время Корпус включает ряд подкорпусов с разным составом и типом разметки. В корпус, называемые «основным» включены тексты XVIII – XXI вв., снабженные морфологической и семантической разметкой; «синтаксический» корпус представлен текстами, созданными в XX-XXI вв., имеющими разметку синтаксических структур, «газетный» корпус — это тексты СМИ. Также в НКРЯ есть «поэтический» подкорпус, в котором представлены стихотворения русских поэтов XVIII – XX вв. с разметкой метра, видов рифмовки, а также строфики и т. д. Немного позже был добавлен «мультимедийный» подкорпус, представляющий собой небольшие видео и аудиозаписи фрагментов кинофильмов, снабженные текстовой расшифровкой и комментариями, касающимися сопутствующей тексту жестикуляции. Так же гораздо позже появился и «исторический» подкорпус, содержащий произведения древнерусской литературы (т.е. периода XI— XVII вв.), а также более поздние тексты на церковнославянском языке, имеющие особый орфографический поиск, специфичный морфологический поиск. Также в НКРЯ появился

«параллельный» корпус, представляющий собой тексты оригинала и переводов, соотнесенные друг с другом построчно.

Основной массив текстов – это тексты середины XX – первых лет XXI века. Тексты более раннего периода (XIX– середина XX века) в рамках Программы РАН выделены в особый, по всем параметрам симметричный данному проект, который осуществляется Петербургской группой ИЛИ РАН под руководством чл.-корр. РАН Н. Н. Казанского. Тем не менее, уже представленная в Корпусе диахроническая шкала текстов достаточно широка, что делает его действенным средством для исследования изменений, произошедших в русском языке начиная с 1950-х годов и до нашего времени.

Среди качественных текстов, доступных в электронном виде, естественно, преобладают произведения последних двух десятилетий; в связи с этим ведется масштабная работа по подбору «сырых» (распространяемых в Интернете и в разного рода CD- «библиотеках») электронных версий печатных изданий 1950– 1970-х годов и вычитке их в соответствии с оригинальными изданиями.

С другой стороны, большинство художественных произведений, приобретших к нашему времени значительную известность и культурную значимость, публикуемых сейчас отдельными изданиями и в журналах, созданы авторами старшего поколения, не позднее 1930-1940-х годов рождения; язык их текстов, созданных в 1990-2000-е годы, может быть уже не вполне показательным с точки зрения языковых изменений, происходящих в указанный период в языке большей части русскоговорящего социума. Указанный «перекос» отчасти исправляется тем фактом, что среди авторов современных нехудожественных текстов – статей в газетах, специальных журналах, не говоря уже об Интернет-новостях и форумах – как раз преобладают люди молодого поколения. И тот, и другой факт, на наш взгляд, отражают реальное состояние генеральной совокупности русских текстов и не нарушают пропорциональной репрезентативности Корпуса.

Таким образом, работа над архитектурой Корпуса демонстрирует известную из опыта мирового «строительства корпусов» взаимосвязь трех проблем:

- 1) Собственно подбора качественных текстов в электронном виде;
- 2) Типологической классификации текстов и разработки инвентаря метапризнаков;
- 3) Репрезентативности / пропорциональной балансировки корпуса.

Ни одна из этих трех проблем не может быть решена без учета двух других; на каждом этапе работы над корпусом – по мере накопления нового материала – решение каждой из них постоянно пересматривается и уточняется [53].

1.4. Корпусные методы исследования

В связи с тем, что корпусная лингвистика является достаточно молодой дисциплиной, корпусные методы до сих пор остаются несистематизированными и не описанными в полной мере в научной литературе. Некоторые методы корпусных исследований формируются в самом процессе работы лингвистов с корпусами текстов. Корпусные методы разнообразны и во многом определяются областью исследования. Рассмотрим методы сбора и систематизации лингвистической информации, которые оказываются востребованными, например, при изучении грамматики и лексики того или иного языка. Корпусы текстов позволяют существенно расширить эмпирическую базу исследования. Преимущество поиска в корпусе особенно очевидно при работе с редко встречающимися языковыми явлениями. Оценить их частотность и подробно описать их практически невозможно без поисковых действий в корпусе.

Квантитативные статистические методы, не являясь собственно корпусными, приобретают некоторую специфику, связанную с осуществлением подсчетов сразу по многим критериям. Так, с помощью корпусов можно исследовать частотное распределение различных

лингвистических параметров. Корпусные данные о частотных характеристиках лексем позволяют получить сбалансированные частотные словари, отражающие реальные частотные характеристики лексем в языке.

Полученные частотные данные дают возможность увидеть новые закономерности в жанровой организации текстов, которые ранее не попадали в поле внимания исследователей. Так, на основе корпусных данных составляются новые словари, в которых, по сравнению с привычными словарями, отдельные значения выделяются не на основе интуиции исследователей и их дальнейшего группового обсуждения, а на основе большого количества контекстов употребления. Также значимо, что при создании современных словарей авторы опираются не только на образцы классической художественной литературы, но и на самые разнообразные контексты, связанные как с письменной, так и со спонтанной устной речью. Квантитативные методы позволяют обработать большие объемы контекстов и компактно представить полученные результаты. Поэтому современные корпуса представляют для лингвистов значительно больший интерес, нежели традиционные словари, поскольку содержат актуальную и достаточно полную информацию о функционировании лексем в языке.

На основе корпусных данных отслеживается микроэволюция языка и предсказываются его дальнейшие преобразования. Также с помощью корпусных данных оценивается продуктивность языковых единиц: морфем, морфологических форм и т.д. Корпусные исследования определяют способы описания грамматики языков. Важным методом корпусной лингвистики является выявление сочетаемости языковых единиц и их свойств. Важно отметить преимущества подходов корпусной лингвистики:

- Возможность производить подсчет;
- Составлять объективные и обоснованные статистики.

Это помогает основывать исследование на точных эмпирических данных, а не только на предположениях и гипотезах. Рассмотрим список

- Диверсифицированный поиск;
- Составление статистики, контекстов и конкордансов;
- Подсчет частотности употребления;
- Анализ метаразметки текста и составление собственного подкорпуса текстов .

функционала корпусного анализа текстов [55] (рисунок 1).

Рис. 1. - Функции корпусного анализа текста

Отметим, в чем заключается значение использования корпусов текстов (рисунок 2).

- 1) Данные разного типа находятся в корпусе в своей естественной контекстной форме, что создает возможность их всестороннего и объективного изучения;
- 2) Достаточно большой (репрезентативный) объем корпуса гарантирует типичность данных и обеспечивает полноту представления всего спектра языковых явлений;
- 3) Однажды созданный и подготовленный массив данных может использоваться многократно, многими исследователями и в различных целях.

Рис. 2 - Значение использования корпусов

Среди методов корпусной лингвистики выделяются следующие группы [35], представленные на рисунке 3.

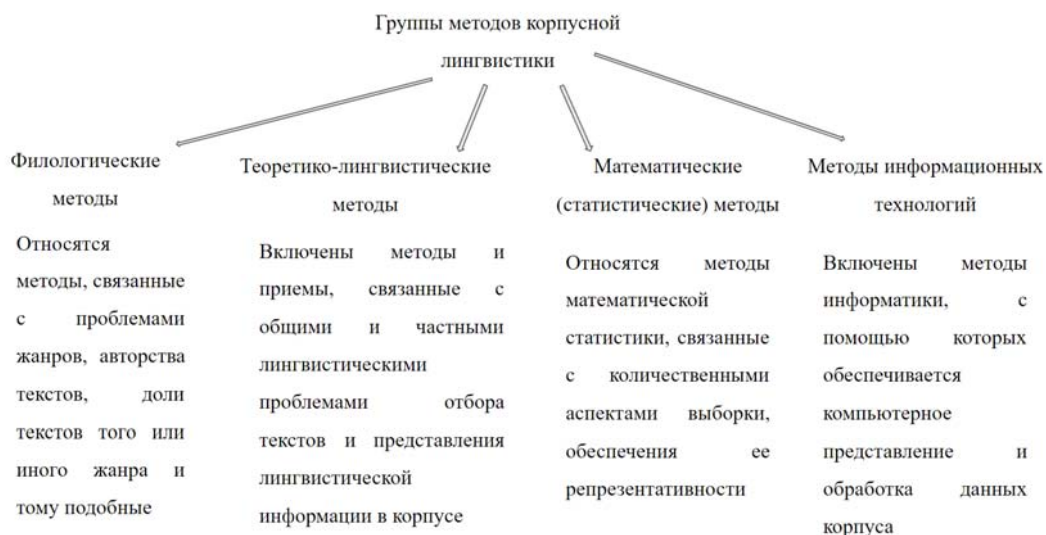


Рис. 3 - Группы методов корпусной лингвистики

Исследование корпусной лингвистики равно как части компьютерной лингвистики приводит к доминированию методов компьютерной лингвистики. Сферы использования корпусов текстов многообразны. Так, они могут быть соединены в лингвистические, педагогические, филологические. То есть в такие дисциплины, как представленные на рисунке 4.



Рис. 4 - Дисциплины, в которых используются корпуса текстов

Сегодня прослеживается расширение области использования корпусной лингвистики.

Необходимо отметить, что прикладная лингвистика изучает методы решения практических задач, связанных с использованием языка. Прикладная лингвистика как научная дисциплина существует с 1950-х годов.

- Фиксация и хранение речевой информации;
- Создание алфавитов и письменностей;
- Транслитерация (главным образом применительно к географическим названиям и собственным именам);
- Практическая транскрипция;
- Орфография; Создание информационных языков;
- Передача речевой информации;
- Создание систем автоматического перевода;
- Теория перевода (главным образом научно-технического);
- Создание систем искусственного интеллекта и т.д.

Основные направления [4] представлены на рисунке 5.

Рис. 5 - Основные направления прикладной лингвистики

Использование методов корпусной лингвистики нужно для того, чтобы решить языковые задачи разного уровня. Особенно важным представляется создания автоматизированных систем для решения задач по обработке информации [10].

1.5. Корпусные методы в изучении художественной литературы

Как известно, корпуса текстов создавались лингвистами и предназначались исключительно для лингвистических исследований. Исключение составлял поэтический подкорпус, имеющий разметку стиховых параметров текстов и, следовательно, способный привлечь внимание исследователей литературы. Вскоре корпуса текстов стали востребованы в самых разных областях научного знания. Национальным корпусом начали интересоваться переводчики, преподаватели русского языка как иностранного.

В настоящее время в литературоведении НКРЯ используется в сопоставительных исследованиях, в частности в сравнительном анализе

художественных произведений различных авторов, а также текста произведения на языке оригинала и в переводе и т.д. Также корпусные методы используются для автоматического выявления особенностей авторского стиля; актуален корпус и для количественного анализа текста, что является вспомогательным средством литературоведческого исследования. В последние годы актуальным становится использование корпуса для автоматического определения авторства. Также следует отметить роль корпуса в стиховедческих исследованиях. Поэтический корпус НКРЯ позволяет задавать параметры поиска стихотворных размеров в подкорпусах разного объема, в настоящее время поэтический корпус стал платформой для стиховедческих исследований.

К сожалению, НКРЯ не предоставляет исследователям художественной литературы различные редакции литературного произведения, отсутствуют справочные сведения текстологического характера. Что касается датировки текстов, атрибуции, орфографического режима, то в этом отношении создатели Корпуса ориентируются на академические издания собраний сочинений представленных авторов. Создатели НКРЯ не пересматривают регулярно библиографических источников. Кроме того, в НКРЯ не представлены полные собрания сочинений авторов (за исключением поэтического подкорпуса), не представлены и все произведения литературы того или иного периода или автора, поскольку, как отмечалось ранее, целью создателей Корпуса была репрезентативность выборки, а не полнота представления всего объема русскоязычных текстов. Тем самым, на основании Корпуса невозможно, например, определить, присутствие или отсутствие того или иного слова в русскоязычных текстах определенного периода.

Между тем, НКРЯ существенно облегчает первичный поиск необходимого литературоведу материала и оценку частотности изучаемого явления по сравнению с традиционным просмотром текстов, составлению картотек и т.д. Так, например, изучая сочетаемость ряда слов, характерных

для жанра элегии первых десятилетий XIX в., с различными эпитетами в русской литературе, можно описать процесс формирования элегических формул русской поэзии и их последующую эволюцию. Также НКРЯ дает возможность проследить использование определенной лексики в диахроническом аспекте. Исследуя, например, жанр дружеского послания в русской поэзии первой трети XIX в., можно выявить наиболее характерные «маркеры» хронотопа, типичного для этого жанра в текстах А.С. Пушкина, К.Н. Батюшкова и др. авторов. При изучении лексики в диахроническом аспекте можно определить, например, время появления того или иного слова в текстах русских писателей. Так, например, В.А. Плунгян проанализировал историю лексемы «жуть» в русскоязычных текстах. НКРЯ показывает, что данная лексема впервые появляется у Н.С. Лескова и воспринимается как неологизм и пишется в кавычках, а через полвека в русском модернизме оно становится одним из маркеров литературного направления, употребляясь и в прозе, и в поэзии.

Также актуальным является поиск лексики того или иного семантического поля средствами НКРЯ. Так, например, в исследовании Е.С. Фоминой, посвященном творчеству И.С. Тургенева с помощью семантической разметки корпуса выявляется специфика образов разных этносов в сочинениях автора. Так, исследователем был составлен подкорпус текстов И.С. Тургенева и в нем был проведен поиск лексики с пометой «этноним». В результате был получен перечень этнонимов и контекстов, позволивших получить точные статистические результаты для обоснования гипотезы исследования. Результаты исследований поэтических произведений средствами НКРЯ отражены в сборниках «Корпусные исследования русского стиха», демонстрирующего различные аспекты применения корпусных методов при анализе поэтического текста. Так, например, с помощью Корпуса рассматривается ритмика свободного стиха, анализируются отдельные частиречные формы, определяется степень оригинальности текста, проводятся такие исследования, как, например, выявление специфики

лексики К.Н. Батюшкова в период душевной болезни на фоне традиционной лексики в поэзии его времени. Чаще всего НКРЯ привлекается для выявления частотности лексики определенной группы текстов (одного автора, направления, периода и т.д.). Так, например, А.А. Бонч-Осмоловская и Б.В. Орехов на базе Корпуса составили частотные словари наивной и высокой поэзии. Благодаря проведенному исследованию они обнаружили, что в наивной поэзии отсутствует значительный пласт «высокой лексики» не востребован в наивной поэзии, предпочитающей «прозаизмы» (*сигарета, проблема и др.*). Тем самым исследователи обнаруживают неактуальность традиционного поэтического инструментария для наивной поэзии, несмотря на декларируемую ориентацию на вершины поэтического искусства, а далее перед ними встает вопрос, каковы механизмы отбора наивными поэтами поэтического материала из арсенала высокой поэзии. Многие исследования, посвященные русской рифме, базируются на данных НКРЯ. Так, Р.Г. Лейбов с помощью статистических методов выявил степень привязанность определенной лексики к рифме. Результаты исследования показали, что частота использования слова в качестве рифмы определяется его концептуальной значимостью, а также количеством его лексем, связанных образующих потенциальные тематико-сюжетные линии. Так, например, Р.Г. Лейбов показывает, что слово Полтава как правило встречается в ряду рифм, имеющих батально-патриотическую тематику (держава – слава и др.) и постепенно втягивает в этот ряд рифм новые слова, в связи с историческими событиями (например, название столицы Польши – Варшава). Стихovedы с помощью НКРЯ рассматривают также различные ритмико-синтаксические клише, задавая в подкорпусе различные частиречные сочетания, связанные с тем или иным размером, например, в трехстопном амфибрахиях можно выбрать последовательность «прилагательное – прилагательное – существительное». Предложенные Корпусом результаты позволяют сравнивать частиречное наполнение этих клише у разных поэтов, определять сближения и различия между ними.

Литературоведы, которые занимаются историей художественного перевода, исследуют, например, лексику, осознаваемую переводчиками как культурно значимая. Так, например, при переводах известного произведения американской литературы слово *hamburger* в разные эпохи переводили по-разному: первоначально его либо пропускали в переводе, либо заменяли на англ. Сэндвич, позже это слово переводили как «котлета», и только с 1990-х годов в переводах появляется слово «гамбургер».

Таким образом, НКРЯ является активно используемым в литературоведении инструментом, предоставляющим вспомогательные методы для изучения литературы и позволяющим ставить новые задачи исследования, которые до появления Корпуса не попадали в поле зрения исследователей.

ГЛАВА 2. Корпусные методы в изучении древнерусской литературы

2.1. Специфика древнерусского текста и его изучения

«Древнерусский текст как объект изучения представляет собой большой интерес для гуманитарных наук, поскольку, являясь культурным и языковым знаком, имеет и универсальные, и специфические национальные черты структурной организации» [22].

Изучение древнерусского текста вызывает множество затруднений у современного исследователя по целому ряду причин. Во-первых, изучение древнерусских текстов базируется на результатах целого корпуса исторических филологических дисциплин, таких как историческая лексикология, историческая семантика (позволяющие уточнять семантику лексических единиц текста определенного периода), историческая стилистика (помогающая установить границы и принципы соотношения народно-литературного, книжно-славянского и делового стилей речи), а также исторической поэтики, помогающей увидеть эволюцию

художественного сознания, образных средств выражения в диахронии и специфику древнерусской литературы в контексте этой эволюции.

Не менее важной проблемой является недостаточная изученность древнерусских рукописей и отсутствие авторского текста. Как известно, исследователям приходится реконструировать архетип авторского текста, для чего необходима объединение данных текстологии, палеографии, кодикологии и других областей.

В конце XVII века Симеон Полоцкий дал концепцию текста как объекта, критически воспринимаемого субъектом. Эта концепция была основана на учении Аристотеля. Считается, что в рамках этой новой концепции развивалось толкование текста при помощи притч. Также в рамках новой концепции доказывалось функционирование в тексте следующих важных категорий:

- Наличие метафор и степеней сравнения;
- Иерархия и движение;
- Меры и “единое централизованное время” [36].

«Существуют разные способы изучения древнерусского текста. Некоторые ученые считают, что основная единица при изучении древнерусских текстов– это предложение, так как книжный текст XI–XVII вв. и содержательно, и структурно достаточно полно расчленяется на такие смыслоразличающие единства, которые уже получили или еще получают статус предложения» [54]. Исследователями разграничиваются три текстовых уровня, представленных в таблице 1.

Таблица 1

Текстовые уровни

1	Микротекст	Он представляет собой такую подсистему макротекста, категориально-оценочные компоненты которой создают информационно-тематический комплекс. Этот комплекс отличается от других подобных структур
---	------------	--

		доминантой в рамках смыслового целого. Необходимо также отметить, что микротексты по объему могут быть схожими с высказываниями. Также их можно обозначить как основные единицы анализа [22].
2	Макротекст	- это смысловое целое, при этом относительно завершенное. Оно отлично, во-первых, своим составом, и, во-вторых, организацией подсистем оценочных компонентов от подобных образований.
3	Мегатекст	Совокупность текстов, которые воспринимаются и исследуются как единое целое, пронизанное общими темами, лейтмотивами, архетипами, символами, ключевыми словами, стилевыми приемами.

Необходимо рассмотреть, как Д.С. Лихачев в литературоведческом аспекте исследует особенности древнерусских текстов. Фактически в древнерусской литературе отсутствовало точное понимание об авторской собственности. Т.е. авторская принадлежность произведений имела значимую ценность только лишь в том случае, когда автору произведения был присущ внелитературный авторитет. К такому относились, как правило, светский авторитет или же церковный.

Если в произведении выявлялась знакомая автору тема, то он непосредственно формировал новое произведение, но как переделку прежнего. Следует отметить, что при этом автор менял в какой-либо мере или структуру, или идею, или стиль произведения, то есть дополнял либо, наоборот, сокращал его. «В результате тексты произведений очень тесно переплетались между собою. Трудно или даже невозможно было различить, какой текст является первоначальным» [31].

«Дмитрий Сергеевич Лихачев рассматривает проблему определения, классификации и взаимодействия жанров. Он отмечает, что древнерусский текст неустойчив, но традиционен, то есть жанры резко отграничены друг от

друга, а произведения, наоборот, слабо отграничены друг от друга, они сохраняют свою устойчивость лишь в некоторых случаях» [32].

Д.С. Лихачевым было выделено 2 вида сборников, каждому из которых дается соответствующая жанровая характеристика:

- Сборники устойчивого состава;
- Сборники неустойчивого состава.

Состав сборников, конечно, может быть абсолютно разным. Однако, важно подчеркнуть, что неизменным при этом остается тип сборников. Лишь отчасти жанр сборника определяется жанрами произведений, включенными в данный сборник.

«Процесс жанрообразования русской литературы XI–XVII вв. протекал весьма активно и своеобразно, в ходе его не только осваивались жанры византийской и болгарской литературы, но и вырабатывались собственные, с самобытными признаками» [24]. Особое внимание уделяется жанровому своеобразию литературы, которая непосредственно носит сакральное содержание. “Которая существовала и развивалась параллельно светской, отчасти обуславливая самый процесс жанрообразования» отмечает Л.П. Клименко при проведении анализа старопечатных памятников семнадцатого-восемнадцатого века.

Отметим, что к древнерусским текстам понятие “гипертекст” применяет Е.С. Курзина при проведении анализа средневековых сборников.

- | | | |
|---|---|--|
| 1) Нелинейность | - | читатель сам выбирает последовательность чтения и тем самым создает свой текст |
| 2) Фрагментарность | - | текст создается на основе множества микротекстов |
| 3) Взаимодействие разных способов воздействия на читателя | | |

Основные черты гипертекста указаны на рисунке 6.

Рис. 6 - Основные черты гипертекста

Можно сделать вывод, что древнерусские сборники являются по своей сути гипертекстуальными. Рассмотрим причины данного утверждения. Во-первых, в древнерусских сборниках имеется синкретичность и символизм, свойственные средневековому мироощущению. Во-вторых, древнерусским сборникам свойственно отсутствие представления о литературном развитии, а также об авторстве. В-третьих, состав древнерусского гипертекста - это тексты меньшего объема, которые соединены в сложные структуры.

Необходимо сделать отсылку на работу В.В. Колесова [23], объектом анализа которой является трансформация древнерусского текста как общности формул-синтагм. Ученый отмечает следующий немаловажный факт: история литературного языка изучает текст как «форму использования системы языка в создании функционально оправданных стилей; в результате возникает норма в исторической последовательности ее проявлений: как образец-текст, затем как узус привычного употребления и только в конце концов как обязательный стандарт. Система языка организует структуру, стиль речи определяет ее функцию, но только норма порождает стандартные правила их совместного и постоянного взаимодействия».

Средневековые писатели использовали для нового текста разные элементы текста-образца. Основными из них являются речевые формулы. Они, к сведению, характеризуются устойчивостью в истории языка в отличие от жанров. При исследовании данной темы Владимир Викторович Колесов приходит к следующему выводу: «так называемого литературного языка есть история постепенного развития и преобразования речевого материала (текста) под влиянием речевой деятельности (творческих потенций своего времени) и обусловленного ею изменения языковой системы» [23].

Однако не все могут дать описание в своих научных трудах, посвященных непосредственно древнерусским текстам. Российский ученый Вера Львовна Ринберг уделяет безусловно особое внимание при анализе связности древнерусских текстов древней цепочечной конструкции, состоящей из нанизывания. Исследователь также обращает значительное

внимание тому, как вышеприведенная конструкция соотносится с системой сложных синтаксических единиц русского языка. К тому же Вера Львовна Ринберг прослеживает семантико-структурные изменения разновидностей данной конструкции, безусловно, отмечая, что “связность древнерусского текста выражается в цепочном строении, но зависимость составных частей в такой структуре особая”. Необходимо дополнить, что исследователь Вера Львовна подчеркивает: “специфика состоит в том, что при всей тематической связности предложения сложной модели сохраняют, в отличие от предикативных частей сложного предложения, независимость строения, с которой согласуются невыраженные между ними смысловые отношения” [57].

Все тексты памятников древнерусской письменности В.Л. Ринберг [57]

1) Произведения, где цепное наизывание предложений составляет синтаксическую доминанту языковой системы документа.

Например различные летописи;

2) Памятники, в которых данная модель представлена в более или менее значительном количестве, но их совокупность не отражает синтаксическую специфику документа.

Например «Изборник 1076 года»;

3) Произведения, в которых отсутствуют цепочные структуры либо обнаруживаются отдельные малопоказательные образцы.

Например «Слово о полку Игореве».

делит на группы, представленные на рисунке 7.

Рис. 7 - Группы текстов памятников древнерусской письменности по В.Р. Ринберг

Языковые средства выражения категории связности летописного текста рассматривает З. Трестерова, отмечая, что с «языковой точки зрения особенности связности текста летописей сводятся к следующим явлениям:

- Особенности согласования;
- Чередование субъектов;
- Особенности в употреблении местоимений;
- Лексико-семантические показатели и корреляты;

- Эллипсис и плеоназмы;
- Употребление конструкций добавления;
- Анаколуфы.» [70].

Особенности связности древнерусских текстов зачастую являются причиной или непонимания текста, или его оценки как «нелогичного». Отдельные компоненты категории концептуальности анализируются различными учеными в данной области. Факторы концептуализации времени и пространства в формировании картины мира древнерусского летописца рассматриваются в исследовании А.А. Селезневой. Исследователь в своем научном труде систематизирует средства реализации концептов времени и пространства в основные типы летописного текста. По мнению автора, «локально-временная информация в летописании реализуется следующими средствами:

- Временной маркер (речевая формула даты);
- Пространственный маркер (различные онимы);
- Лексико-грамматические единицы (глагольные формы с определенной семантикой)» [65].

Особое внимание уделяют анализу модальности в древнерусском письменном тексте такие историки языка, как Отин Е.С. [42], Черепанова О.А. [71], Ваулина С.С. [7], Балута А.А. [3]. Модальность как текстовая категория в диахроническом аспекте рассматривается исследователем А.В. Опаринной. Она указывает этапы эволюции данной категории на пути к авторской модальности. Также она описывает, как модальность формирует и категорию антропоцентричности, которая, в свою очередь, сближает модальность, оценку и экспрессивность [39].

Из других категорий внимание современных исследователей обращено на такие категории, как время, оценка, интенциональность и обусловленность. Однако сегодня историческая лингвистика текста в стадии своего формирования. Остаются актуальными задачи лингвистического изучения текстов, сформулированные еще В.В. Виноградовым, писавшим о

необходимости разработки «исторической типологии композиционных форм речи как систем языковых объединений, которые встречаются в ткани литературно-художественных произведений той или иной эпохи. Изучение этих основных типов естественно предполагает группировку не самих литературных произведений, а отвлеченных от них однородных форм словесной композиции, в общелингвистическом плане – на фоне эволюции композиционных жанров прагматической речи. Это учение не о структуре художественных единств, а о структурных формах речи, которые наблюдаются в организации литературных произведений. И его задача – установить закономерности в их построении как систем языковых отношений» [9].

2.2. Древнерусская литература в НКРЯ: состав, способ представления, разметка

В Национальном корпусе русского языка есть древнерусский подкорпус. Он включает в себя оригинальные древнерусские произведения, памятники южнославянского происхождения, переписанные на Руси, а также переводы с греческого языка, выполненные на Руси. На рисунке 8

Из оригинальных произведений в подкорпус входят все летописи, созданные в древнерусский период:

- Повесть временных лет;
- Киевская, Галицкая, Волынская летописи (все по Ипатьевскому списку);
- Новгородская I летопись (по Синодальному списку);
- Суздальская летопись (по Лаврентьевскому списку);
- Также представлены сочинения древнерусского писателя XII в. Кирилла Туровского «Сказание о черноризском чине», «Притча о душе и теле», «Повесть о беспечном царе и мудром советнике».

Из переведенных в Древней Руси памятников в подкорпусе доступны:

- «История Иудейской войны» Иосифа Флавия (повествование о завоевании Иудеи римлянами в 60-70-е гг. н. э.);
- «Пчела» (сборник фрагментов из Св. Писания, сочинений отцов Церкви и античных философов);
- «Александрия» (исторический роман об Александре Македонском);
- Цикл из восьми «Чудес» Николая Мирликийского, шесть из которых переведены на Русь или представляют собой переработку перевода, а два возникли у южных славян.

указано, какие летописи и памятники есть в открытом доступе в НКРЯ.

Рис. 8 - оригинальные и переведенные в Древней Руси тексты, которые доступны в древнерусском подкорпусе НКРЯ

Из древнеболгарских переводов в древнерусском подкорпусе НКРЯ содержится такой памятник, как Изборник 1076 года. В основе Изборника 1076 г. лежат непосредственно переводы с греческого языка. Сюда включены, во-первых, отрывки из библейской книги Премудрости Иисуса сына Сирахова, во-вторых, отрывки из произведений таких византийских писателей, как Афанасий Александрийский до константинопольского патриарха Фотия (IX в., Иоанн Златоуст, Василий Великий. Для некоторых текстов, к сожалению, не обнаружены аналоги в греческо-византийской литературе. Причиной является тот факт, что эти тексты («Слова некоего отца к сыну», «Слова о чтении книг», «Столбца») являются оригинальными сочинениями неизвестных древнеболгарских авторов.

В древнерусском подкорпусе Национального корпуса русского языка также и по грамматическим признакам можно производить поиск. После извлечения результата поиска, следует отметить, на экране высвечиваются краткие сведения о памятнике. Также появляется такая информация, как указание на издание или рукопись, по которым выполнена разметка. Если нажать на определенную словоформу, то на экране появится всплывающее окно. В данном окне указывается греческое соответствие выбранной словоформы, а также ее адрес: страница издания/лист рукописи и номер строки. Для разграничения омонимов необходимо указывать при поиске конкретной леммы ее частеречную характеристику, а у существительных - род.

Отдельный грамматический поиск по полным и кратким формам прилагательных в древнерусском подкорпусе НКРЯ возможен лишь в Новгородской I летописи. «История Иудейской войны» и «Александрия» размечены по одной и той же акцентуированной рукописи – Архивскому хронографу последней трети XV в. “Знаки акцентуации сохранены при разметке. В тексте «Истории Иудейской войны» они были перепроверены по рукописи, текст «Александрии» не перепроверялся и приводится так, как он издан В. М. Истриным” [38].

2.3. Структура и семантика «Жития Феодосия Печерского» средствами НКРЯ

Современный этап в лингвистике характеризуется возрастающим интересом к тексту как к сложному явлению. Текст изучается исследователями как структурно-семантическое образование, некое сложное единство. Исследование семантической организации текста безусловно включает следующие текстовые универсалии: «человек», «время», «пространство».

Особый интерес возникает при изучении структуры и семантики древнерусского текста. В основе анализа в рамках данной работы лежит

произведение «Житие Феодосия Печерского»—памятник древнерусской литературы о жизни преподобного Феодосия, одного из основателей и игумена Киево-Печерского монастыря. Является древнейшим русским агиографическим произведением.“ Автор: Нестор Летописец.

Произведение подробно повествует про святого Феодосия Печерского. “Житие” содержит описание его рождения, прихода в монастырь, игуменства, смерти. Также произведение включает в себе легендарные житийные мотивы и действительные факты биографии преподобного. Также в преподобническом житие обозначено множество традиционных библейских мотивов. Например, будущий святой был рожден благочестивыми родителями. Феодосий Печерский детства он отличался «прилежанием» к Церкви. Будущий святой шел от радостей и соблазнов мирской жизни. Он стал образцом аскета и подвижника, творил чудеса.

При написании “Жития” Нестор Летописец применял характерные для данного жанра символические тропы:

- Монастырь Феодосия Печерского подобен «небеси, и вънемъблаженныйотецьнашь Феодосий паче солнца въсиявъдобрымидѣлы»;
- “Свѣтило, въвъсьмирѣ видимое и просиявъшеевъсьмьчърноризьцемъ» - Феодосий Печерский.

Также традиционной можно считать эпитетику жития. Она представляет собой оценку автором нравственной сути героев. В “Житие Феодосия Печерского” такая оценка дана через определения и сравнения. Например:

- Разбойники, пришедшие грабить монастырь, “акызвѣрие дивим”;
- “Христолюбивый” князь Изяслав.

Стилю исследуемого жития дают возвышенность сложные слова в виде «медоточные словеса», «богодъхновеный игумен». Также сюда можно отнести и сопряжение в одном контексте слов с конкретным и абстрактным значением. Это, стоит отметить, подчеркивало главную антитезу –

противопоставление материального и духовного начала в жизни человека. Например, «свѣтлая» одежда и «свѣтлая» душа.

Нестор Летописец применяет систему сравнений и аллегорических картин. Данная система происходит из книги Священного Писания. Приведем пример: “Варлаам спешит в пещеру к Антонию как птица, вырвавшаяся из сети, или серна – из западни. Монахи впитывают наставления игумена Феодосия, «яко земля, жажущая воды». Бог, уверяет Нестор, «не попустил» Феодосию стать странником, чтобы он не покинул свою страну: уйдет пастырь, и опустеет пажить, и зарастет тернием и бурьяном, и разбредется стадо”.

Нестор Летописец цитирует евангельский текст. Именно поэтому стиль “Жития Феодосия Печерского” можно отнести к афористичному. Например, призывая монахов устами Феодосия подражать «птицам небесным», которые не думают о завтрашнем дне, полагаясь на то, что Бог их «препитает». Или вспоминает евангельскую притчу о достойном рабе, умножившем талант, данный господином, и недостойном, талант сокрывшем; или рассуждает о том, что «смерть – покой для праведника». Также необходимо отметить тот факт, что «Житие Феодосия Печерского», написанное Нестором Летописцем, заложило некий фундамент для дальнейшего развития жанра преподобнического жития в русской литературе.

«В настоящее время в НКРЯ реализована система поиска по лексико-семантическим признакам, основанная на частичной семантической разметке текстов. При такой разметке большинству слов в тексте приписывается один или несколько семантических и словообразовательных признаков. Используется фасетная классификация, при которой одно слово может попадать в несколько классов. На первом этапе поиск осуществляется по части имеющихся в словаре признаков.

Разметка текстов осуществляется автоматически с помощью программы Semmarkup (автор А. Е. Поляков) в соответствии с

Семантическим словарем Корпуса. Поскольку ручная обработка семантически размеченных текстов очень трудоемка, семантическая омонимия в Корпусе не снимается: многозначным словам приписывается несколько альтернативных наборов семантических признаков.

В основу семантической разметки положена система классификации русской лексики, принятая в базе данных «Лексикограф», которая разрабатывалась с 1992 году в Отделе лингвистических исследований ВИНТИ РАН непосредственно под руководством Елены Викторовны Падучевой и Елены Владимировны Рахилиной».

Словник семантического словаря базируется на морфологическом словаре системы DIALING. Она содержит, к сведению, более 120 тысяч слов. Так, DIALING является расширением Грамматического словаря русского языка Андрея Анатольевича Зализняка. Текущая версия семантического словаря включает как числительные, и местоимения, существительные и прилагательные, так глаголы и наречия.

Лексико-семантическая информация имеет различную структуру для разных частей речи. Кроме того, каждый из разрядов существительных - имена предметные, непередметные и собственные - имеет свою структуру помет. Собственно лексико-семантические пометы сгруппированы по

- Таксономия (тематический класс лексемы) — для имен существительных, прилагательных, глаголов и наречий;
- Топология (топологический статус обозначаемого объекта) — для предметных имен;
- Мереология (указание на отношения «часть — целое», «элемент — множество») — для предметных и непередметных имен;
- Лексико-семантический (таксономический) тип производящего слова (например, наречие, образованное от прилагательного размера);
- Оценка — для предметных и непередметных имен, прилагательных и наречий.
- Служебный статус — для глаголов;
- Словообразовательные характеристики включают несколько типов:
- Морфо-семантические словообразовательные признаки (например, «дими́нтив», «кари́тив», «семельфактив»);
- Каузация — для глаголов;
- Разряд производящего слова (например, отглагольное существительное или ота́дъективное наречие);
- Морфологический тип словообразования (субстантивация, сложное слово).

следующим полям (рисунок 9).

Рис. 9 - Поля, по которым сгруппированы лексико-семантические пометы

В основу метаязыка лексико-семантических помет положена система сокращенных помет на основе англоязычной нотации. В то же время предусмотрена возможность использования при поиске традиционных названий категорий на русском языке.

Для анализа структуры и семантики «Жития Феодосия Печерского» в НКРЯ необходимо обратиться к разделу «Древнерусский корпус» и выбрать подкорпус «Житие Феодосия» (древнейшее восточнославянское житие: описание жизни Киево-Печерского игумена преподобного Феодосия, написанное монахом Нестором. Текст приводится по Успенскому сборнику).

Семантическая основа древнерусской словесности - Священное Писание. Древнерусские тексты являются развертыванием образов-символов и библейских мотивов.

Ситуация с интерпретацией многих древнерусских произведений представляется почти тупиковой. Не располагая непосредственными данными, которые позволяли бы получить прямой ответ на все эти вопросы, исследователь вынужден извлекать возможный ответ косвенным путем, на основании рассмотрения самого текста. Так возникает замкнутый круг: характер изучения текста попадает в зависимость от сведений, которые, в свою очередь, могут быть получены только как косвенный результат изучения этого текста.

Разногласия исследователей в вопросе о времени написания «Жития Феодосия Печерского» существуют по сей день. В НКРЯ «Житие Феодосия Печерского» появилось в 2022 году - в Древнерусском корпусе и подкорпусе «Житие Феодосия». Характеризуется как «древнейшее восточнославянское житие: описание жизни Киево-Печерского игумена преподобного Феодосия, написанное монахом Нестором. Текст приводится по Успенскому сборнику». Содержит 167 предложений, 19571 словоформ. Разновидность языка: церковнославянский.

Существуют различные датировки написания Жития: 1080-е годы (исходя из описания ряда событий произведение было написано не позднее 1088 года) или начало XII века. В НКРЯ дата создания «Жития Феодосия Печерского»: 1080-е. Дата списка: XII – XIII в. По мнению А.А. Шахматова [72], «Житие Феодосия Печерского» было написано до 1088 г. и основанием для такого заключения ему послужили следующие факты: Нестор охотно в своих сочинениях забегают вперед в изложении судеб описываемых им лиц, но не сообщает ни о смерти 23 марта 1088 г. игумена Никона и об избрании нового игумена Иоанна, ни об обретении и перенесении в 1091 г. мощей Феодосия Печерского (для жития Феодосия факт знаменательный); ни об освящении Успенской церкви в 1089 г. Наблюдения А.А.Шахматова можно значительно расширить. Обратим внимание на одно замечание Нестора, которое существенно помогает уточнить время его работы над «Житием Феодосия Печерского»: *«Се бо елико же выше о блаженемъ и велицемъ отци*

нашемъ Феодосии оспытовая слышахъ от древньиихъ мене отецъ, бывъишихъ въ то время, та же въписахъ азъ грешьнии Несторъ. мьнии всьехъ въ монастыри блаженаго и отца всьехъ Феодосия» (рисунок 9).

Рис. 9 – Результаты поиска по слову «мьнии» в тексте «Житие Феодосия Печерского» средствами НКРЯ

Эти слова Нестора можно понимать так, что он, меньший всех в монастыре по возрасту («мьнии всьехъ») монах, спрашивал старших («древньиихъ мене») за себя отцов-иноков Печерского монастыря о Феодосии, и выше изложил услышанное. В словах Нестора отчетливо прослеживается сравнение, точнее – противопоставление возраста его самого, записавшего рассказы об отце Феодосии, и возраста монахов, от которых он эти рассказы слышал. Если бы Нестор хотел самоуничижением подчеркнуть свое «низшее» положение в монастыре (и такое значение имеет иногда слово «мьнии»), то была бы в этой антитезе иная пара слов – «старей-мьнии»: «Аще кто хощеть старей быти, да будетъ всьехъ мьнии и всьемъ

Результаты поиска в древнерусском корпусе

Объём всего корпуса: 201 документ, 639 847 слов.

"мьнии"

Найдено: 1 документ, 4 вхождения.

Поискать в других корпусах: [основном](#), [старорусском](#), [церковнославянском](#), [корпусе берестяных грамот](#).

Страницы: 1

1. Нестор. Житие Феодосия (1080-е) [омонимия снята] [Все примеры \(4\)](#)

сѣдѣть страньныхъ. бѣгънъ же бѣ не попусти имѹ ѿйти отъ страны сѣа. игоже и шреба матерьна и пастѹхѹ быти въ странѣ сѣи бѹгласныиѹхъ овьць назнамена. да не пастѹхѹхѹ оубѹ ѿшѣдъши да опоустѣять пакыиъ юже бѣ благослови. и тырникъ и вѣлащѣ въздрастѣтъ на нѣи. и стадо разидѣтсѣа. по трыхѹ оубѹ дѣиѹхъ оубѣдѣвъши мѣнъ иго яко съ страньнымии шгънде. и авни погъна въ сѣдѣть иго тѣкъмо идиного сѣа своего помѣшии иже вѣ мьнии блженнаго феодосна. таме же яко гънаста поуть мьногъ. ти тако пристигъша баста и. и отъ вѣрости же и гѣбѣа мѣнъ иго нѣмъши и за власы и поворъже и на земли. и своимъа ногъама пѣхашети и. и страньныиъ же много коривъши възвратисѣа въ домъ свои. яко ибѣенкогъ зѣлодѣѣа вѣдоуци съвѣзана. тѹмъи же гѣбѣвъмъ одръжыма яко и въ домъ ни пришѣдъши быти и. дондеже изнеможе. и по снѹ же въвѣдъши [Нестор. Житие Феодосия (1080-е)] [омонимия снята] [-----](#)

Оцъ же нашъ феодосни аще и старѣишностьво примѣтъ не измѣни съмѣренна своего правила на памяти гѣа нѣмѣа рѣсъша. иже аще кто въ васъ хощеть быти старѣи вѹди всѣхѹ оубѹ мьнии и всѣмъ слѹгѹа. тѣмъже съмѣршасѣа мьнии всѣхѹ сѣа творѣа и всѣмъ слѹжѣа. и совою образъ всѣмъ. даи на дѣло преже всѣхѹ исхода. и въ чинѹ сѣма антеѹргна. и ѿтомѣ цѣлѣше и мьножасѣа мѣсто то. правдынка мѣтвое. правдыникъ во рече яко и фунѣкъсъ процвѣтѣтъ и яко и кѣдъръ иже въ ливанѣ оумножитсѣа. оумножаѹхѹ во сѣа ѿтомѣ вратна и цѣлѣше мѣсто то добрыими нрабы. и молитвами нѹхъ и нѣбѣи блгчестивымии нрабы. и мнѹгъмъ отъ вѣлможѣа прихѹдити к нѣмоу бл(с)вленна ради. и ѿ нѣмѣни своѣхъ малѹ нѣкакѹ часть подаѹщи имъ. придѣвныи же оцъ нашъ иже по нестѣгѣ. земьныи [Нестор. Житие Феодосия (1080-е)] [омонимия снята] [-----](#)

нашемъ феодосни. оспытовала самшахъ ѿ добвьниихъ мене оцъ. бывъишихъ въ то вѣрѣма. та же въписахъ азъ грѣшнымъ несторъ. мьнии всѣхѹ въ монастыри блаженаго и оца всѣхѹ феодосна. приидѣтъ же вѹхъ въ мъ придѣвнымъ нѣоумѣмъ стѣфанъмъ. и вѹжеи ѿ того острѣнкѣмъ бывѣтъ. и мьнишьскыи одежа съподѣвлѣнъ. пакы же и на диконьскыи сангъ ѿ него нѣвѣдѣтъ сын. имѹже и не всѣхъ достѣнѣмъ грѹхѹ сын и невѣгла(с). наипаче же и множѣствѣмъ грѹхѹвъ. напѣльниихъ сын ѿ оцности. шѣаже же блжню болю и по мѹвѣни томоу тако сътворѣишью. иже и вратни. икоже мнѹгашѣа ми слышашѹ. дѹврѣе и чистѹкъ житни придѣвнаго и бѣносънаго. [Нестор. Житие Феодосия (1080-е)] [омонимия снята] [-----](#)

слоуга». У Нестора же противопоставлялись слова «древньиихъ-мьнии», т.е. в возрастном значении.

Следует отметить, что Нестор и в «Чтении» охотно использует слово «мьнии» в значении «младший по возрасту», усиливая его другим – «уней»:

«Сии Давыдъ (христианское имя Глеба -А. У.) мний (вар. – меньший») бе въ братьи своеи и уней (вар. – «юнейший») в дому отца своего». То есть, меньший (младший) среди братьев и самый молодой в доме.

В этом смысле Нестора не может быть иной трактовки слова, поскольку очевидно, что он не только говорит о юном возрасте Глеба (*«Бе же Глебъ велми детескъ»*), но и всячески подчеркивает этот юный возраст святого, сравнивая Глеба с юным Давидом: *«Святому же Глебови створе но имя Давыдъ. Видиши ли благодать Божию изъмлада на детици? Створиши бо, рече, имя ему Давыдъ: како или кымъ образомъ? Не им же ли: онъ бе мний въ братьи своей (т.е. младший), тако же и сии святой. Яко же бо и самъ пророкъ свидѣтельствуеть, глаголя: «мний бе въ братьи моеи и уний (вар. – «юнейший») въ дому отца моего»* и прочее; *«также и сии Давыдъ мний (вар. – «меньший») бе въ братьи своеи и уней в дому отца своего».*

Нестор родился около 1056-1057 гг. и вряд ли бы написал после 1108 г., когда ему уже было за 60 лет. Число монахов было более сотни, а он «мънии въсехъ» – моложе всех по возрасту в монастыре. Исходя из вышеизложенного, мы не можем согласиться с мнением С.А.Бугославского и его последователей, что Нестор писал «Житие Феодосия Печерского», а, стало быть, и «Чтение о Борисе и Глебе» только после 1108 г., и поддерживаем точку зрения академика А.А.Шахматова – оба сочинения были написаны Нестором до 1088 г., с небольшим уточнением.

В «Житие Феодосия Печерского», описывая уход Феодосия из материнского дома в Киев, Нестор, составитель Жития, упоминает о купеческом обозе, вслед за которым святой идет к стольному городу: *«И се по приключая Божию бешаидоущепоутьмъ те коупьци на возехъ с бременытяжькы. Оуведевъ же я бл[а]женыи, яко вѣтъ же градъидоуть, прослави Б[ог]а. И идяшетъ въ следъихъ издалеча, не являя сямъ. <...>ЕдиномоуБогоусъблюдающую и».*

В. Н. Топоров так интерпретирует Несторов рассказ о пути Феодосия в Киев: «Нестор описывает его трехнедельное путешествие достаточно кратко. Две особенности этого описания бросаются в глаза — ведомость Феодосия Богом, охраняющим его, и исключительно экономно переданная атмосфера этого путешествия, сочетание конкретных деталей (тяжело груженные возы, ночное становище, юноша, боящийся, что его заметят, и следующий за купцами поодаль, скрываясь от их взглядов), не допускающих сомнений в своей подлинности, с каким-то сверхреальным колоритом, почти мистериальным ожиданием предстоящего и переживанием совершающегося, легко восстанавливаемыми по скупой скорописи жития» [69].

Однако интерпретация выражения «бременытяжкы» как предметного описания представляется небесспорной — прежде всего потому, что предметная детализация чужда агиографической поэтике. Возможное объяснение его семантики дает последующее известие Жития Феодосия Печерского, сообщающее о приходе святого в пещеру Антония: *«Тъгда же бослышавъ о бл[а]женомъ Антонии, живоущимъвъ пещере, и, окрилатевъоумъмъ, оустръмисякъ пещере»*. Метафора «окрилатевъоумъмъ» построена на идентификации святого как «земного ангела» («по истине земльнианг[е]ль и н[е]б[е]сныи ч[е]л[о]в[е]къ») с крылатыми небесными силами.

Так произошло, например, с выражением-оксюмороном «земной ангел и небесный человек», которое встречается во многих древнерусских житиях. Очевидно, в качестве цитат в древнерусской книжности (и, в частности, в агиографии) выступают реминисценции из Библии и из литургических текстов, маркированные своим высоким ценностным статусом, хорошо известные (как предполагается) читателю и потому сохраняющие связь с исконным контекстом; не случайно цитаты из Священного Писания, как правило, сопровождаются указанием источника. Заимствованный характер выражения «окрилатевъоумъмъ» не препятствует тому, что в древнерусском житии оно приобретает дополнительные, контекстуально обусловленные

оттенки значения: в Житии Саввы Освященного оппозиция «устремление, полет святого, уподобленного птице — медленное движение обоза» отсутствовала (обстоятельства прихода Саввы в обитель Евфимия Великого непохожи на путешествие Феодосия в Киев). Святой Савва, «окрилатевъоумомъ, пожадавидети с[вято]го о[ть]ца<...> и пришедъ<...>видевъвеликаго о[ть]цаЕвъфимия».

А в Житии Феодосия Печерского о святом сказано иначе: «окрилатевъ же оумъмъ, оустремисякъ пещере, и пришьдъкъ<...> Антонию, поклонисямоу». Глагол «устремися», выбранный Нестором, может быть применен как к человеку, так и к птице, в нем присутствует значение быстрого движения, подобного полету. Глагол «пожада» из славянского перевода Жития Саввы Освященного такого оттенка значения лишен, он едва ли может быть применен к описанию птицы. Для этого глагола характерны также библейские коннотации, но иные: это ассоциации с серной, жаждущей воды. В церковно-славянском тексте Псалтири (41:1): «Имь же образомъжелаетъ елень на источникыводныа, сице желаетъ д[у]ша моя к Тебе, Боже».

В Песни песней говорится: «Беги, возлюбленный мой; будь подобен серне или молодому оленю на горах бальзамических!» (Песн. 8:14). Согласно толкованию Филона Карпафийского, распространенному в древнерусской книжности, эти стихи Песни песней должно прочитывать так: «И по Бозе человекъ бегаеа греха борзотечением и съизъвещениемпоучениемъ. <...> Бегай оубо от земных и мирьскихъ добродетельным мужем и от тех състааляем, причастникъбоудешиНебесному Царствию о Христе Иусе, Господе нашемъ<...>». Но также она основана и на уподоблении человека птице, восходящем к сравнениям из Священного Писания.

Уподобление бегства человека (души) от опасностей птице, летящей на гору, и птице, вырвавшейся из силков, встречается в Псалтири: «Како речеот[е] д[у]ши моей: „превитаи по горама, яко птица“» (10:1); «д[у]ша наша, яко птица, избавися от сети ловящихъ» (123:7); близкий образ — в

Книге Премудрости Соломона (6:5). Правда, в первом из приведенных примеров говорится о бегстве нежеланном, но сходство в плане выражения этого фрагмента Псалтири и Жития Феодосия Печерского всё же несомненно. Нежеланное бегство, о котором упоминает десятый псалом, противопоставлено желанному и богоугодному удалению святого от мира. Между прочим, в Житии «печерский» локус маркирован как горный. В свете сопоставления со вторым библейским фрагментом бегство Феодосия предстает исходом, освобождением святого/души от тенет, сетей — соблазнов мира.

Еще одна библейская параллель к метафоре Жития Феодосия Печерского — речение Христа: *«Възьрите на птиця н[e]б[e]сьскыя, яко не сеють, ни жьнють, ни събираютьвъжъжитьницю, и О[те]цьвашь н[e]б[e]сьскыипитаетъ я»* (Мф. 26).

Коннотации «тяжесть, тягость», содержащиеся в обеих лексемах этого выражения, контрастируют с коннотациями «легкость, полет, устремленность вверх», присущими выражению «окрилатевъоумъмъ». Создается оппозиция «привязанность к миру и его суетным ценностям (метонимически обозначенными товарами купцов) — стремление к истинному, нетленному благу». Два эти выражения разделены небольшим фрагментом текста в Житии Феодосия Печерского, но, тем не менее, их соотносительность более чем вероятна.

Кроме того, «бременятяжъки» соотносятся по принципу семантического контраста с легким бременем, которое возлагает Иисус Христос на уверовавших в него: *«Възьмете иго Мое на себе и наоучитеся от Мене, яко кротъкъесмь и съмерень с[ъ]р(д)[ъ]ц[ъ]мь, и обрящете покои д[у]шамвашимь. Иго бо мое бл[а]го, и беремя Мое льгко есть»* (Мф. 11:29–30). Предметное значение выражения «бременятяжъкы» контрастирует с метафорой *легкое бремя*. Но именно благодаря этой контрастной соотносительности оно также приобретает дополнительный метафорический смысл «ложные ценности, тяжесть мирского, „бытового“ бытия» [18].

В “Житие Феодосия Печерского” содержатся микроцитаты библейских мотивов (рис. 10). В Житие обозначены темы и Святой земли, и время Ветхого, Нового Заветов, и Константинополь. Они могут быть изучены в корпусе.

Тъмъже и мы, братие, потъщимъся рвѣнителе быти и подражателе житию преподобнааго Феодосия и ученикомъ его, ихъ же тогда предъ собою къ Господу посла, да тако сподобимъ ся слышати гласъ он; еже отъ владыки и вседръжителя речеться: «Придѣте убо, благословении отъца моего, примѣте уготованое вамъ царствие». (27г) ⁹

наипаче же яко и въ странѣ сей такъ сий мужъ явися и угодникъ Божий. О семь бо и самъ Господь пророче: [...] «Мнози будутъ послѣднии пръвии», ибо сии послѣднии вѣщии пръвѣхъ отъць явися [...]; сего же Христось въ послѣднииимъ родѣ семь такого себе съдѣльника показа и пастуха инокыимъ. (26в)

Тъмъже и божъствный Давыдъ о семь прѣдъизвѣща, рекый: «Очи господьни на правдѣныя и уши его въ молитву ихъ». (57в)

О семь бо и самъ господь пророче: «Яко мнози приидуть отъ востокъ и западъ и възлягутъ с Авраамъмъ и с Исакъмъ и Ияковъмъ въ царствии небесньмъ. (26в)

Рис. 10 - Микроцитаты библейских мотивов в “Житие Феодосия Печерского”

Значимая роль в Житие отдана образу Иисуса Христа. В.Н. Топоров [69] объясняет значение этого образа для “Жития Феодосия Печерского”

Особая роль принадлежит имени Иисуса Христа (не менее двух с половиной десятков употреблений). Оно открывает и закрывает текст ЖФ, ср: «Благодарю тя, владыко мой, господи Иисусе Христе, яко сподобиль мя еси» (26а) ¹⁵ и «[...] въписахъ на славу и чсть великому Богу и Спасителю нашему Иисусу Христу...» (67в). Но такое положение этого имени не исчерпывается композиционным заданием (рамка текста, Grenzsignal и т. п.). В данном случае существенно, что появление имени в этом месте образует знак перехода от текста к внетекстовой сфере, к пространству и времени автора текста и индуцируемым ими пространству и времени читателя текста, т. е. к серии повторов, воспроизведений, выступающих как своего рода ритуальный элемент ¹⁶. Но и в самом тексте ЖФ роль обра-

следующим образом (рис. 11).

Рис. 11 - Цитата В.Н. Топорова из книги “Святость и святые в русской духовной культуре. Том I. Первый век христианства на Руси”

Таким образом, актуальным является изучение библейских цитат с помощью НКРЯ. Это перспективное направление для работы с Национальным корпусом русского языка. С помощью средств НКРЯ можно изучить синтаксические, морфологические, лексические, фонетические особенности тех или иных слов и выражений. К примеру, слово “владыко”. С помощью средств НКРЯ можно узнать его лексическое значение, узнав контекст употребления слова и пр. (рисунок 12).

[перейти на страницу поиска](#) [сбросить подкорпус](#) [настройки](#) [формат KWIC](#) [English](#)

Результаты поиска в древнерусском корпусе

Объём всего корпуса: 201 документ, 639 847 слов.

"владыко"

Найдено: 1 документ, 1 вхождение.

Поискать в других корпусах: [основном](#), [старорусском](#), [церковнославянском](#), [корпусе берестяных грамот](#).

Страницы: 1

1. Нестор. Житие Феодосия (1080-е) [\[омонимия снята\]](#) [Все примеры \(1\)](#)

жити идишь и намъ не вѣдоуцелья его. какоже многашады въсхотѣ тако сътворити. нъ оумолигъ вывааше о томъ ꙗко князь и ꙗко вельможь. вратни о томъ паче молишиса. и такоже и тоу тако тѣмъ мнѣшемы ꙗко. Таче по сиухъ блаженнаго зливѣ възгрозивъши. и огню оуже люте распальшю и. и не могли кътомуу ничьтоже. въздеже на одрѣ рекъ. вола бѣжна да воздетъ. и какоже изволиша кмоу о мѣстѣ тако да сътворитъ. нъ оваче молю ти са **владыко** мон. мнѣстивѣ боуди дѣи мои да не сърациеть киа противныиухъ лѣукавьство. нъ да примодуть ю айган твои провадице ю. сквозѣ пренарьство тьмыниухъ тѣхъ мытарьство. приволадице ю къ твоего мнѣсарьдина свѣтоу. и си рекъ оумѣльче кътомуу не могли ничьтоже [Нестор. Житие Феодосия (1080-е)] [\[омонимия снята\]](#)

Рис. 12 - Поиск в НКРЯ в “Житие Феодосия Печерского” слова “владыко”

Также морфологической особенностью многих слов из цитат библейских мотивов “Жития” являются окончания творительного падежа существительного и прилагательного (ью/ию, ой/ою). Данные окончания являются древнерусскими и церковнославянскими флексиями творительного падежа существительных. В творительном падеже у слов 3-го склонения нередко используется окончание -ию на месте русского -ью. . Во многом различия с современным русским языком вызваны приверженностью к сложившемуся церковному традиционному стилю написания, присущему жанру жития. Исторические корпуса НКРЯ дают возможность фронтального исследования грамматических феноменов.

Заключение

Таким образом, в результате проведенного исследования можно сделать следующие выводы. Для организации работы с корпусными данными, организации эффективного поиска материала по заданным критериям, в корпусной лингвистике разрабатываются специфические методы исследования текстов, основанные на соединении количественных методов статистического анализа с организацией поиска материала на основе аннотирования текстов. Так, корпус – это определенная выборка, основанная на заданных критериях, связанных с конкретным аспектом лингвистических исследований; унифицированный, большой, а также структурированный массив языковых данных в электронном виде. Он непосредственно служит

исследователям и пользователям для решения каких-либо лингвистических задач.

Созданию русскоязычных корпусов предшествовали разработки корпусов текстов в ряде европейских стран. Первым национальным корпусом, ставшим своего рода образцом для создания всех последующих корпусов текстов, стал Британский национальный корпус (BNC). Отечественная корпусная лингвистика возникает в 1980-е гг. и связывается с появлением работ А. П. Ершова и В. М. Андриященко, посвященных машинному фонду русского языка.

В НКРЯ непосредственно представлены тексты самых разных сфер, жанров и функциональных стилей.

Структура НКРЯ регулярно корректируется, поскольку развитие корпусной лингвистики вызывает непрерывную исследовательскую рефлексию и продолжение разработки устройства языковых корпусов.

В связи с тем, что корпусная лингвистика является достаточно молодой дисциплиной, корпусные методы до сих пор остаются несистематизированными и не описанными в полной мере в научной литературе. Среди методов корпусной лингвистики выделяются следующие группы: филологические, теоретико-лингвистические, математические (статистические), а также методы информационных технологий.

Древнерусский текст привлекает внимание различных исследователей-лингвистов и литературоведов, которые анализируют разные его аспекты. Однако, несмотря на большой интерес к древнерусским текстам, остаются нерешенными многие проблемы, в частности, проблема изучения форм и способов реализации категорий текста. Древнерусский подкорпус в НКРЯ включает оригинальные древнерусских произведения, выполненные на Руси переводы с греческого и памятники южнославянского происхождения, переписанные на Руси. Из оригинальных произведений в подкорпус входят все летописи, созданные в древнерусский период.

Особый интерес возникает при изучении структуры и семантики древнерусского текста. В основе анализа в рамках данной работы лежит произведение «Житие Феодосия Печерского»—памятник древнерусской литературы о жизни преподобного Феодосия, одного из основателей и игумена Киево-Печерского монастыря, написанное монахом того же монастыря преподобным Нестором Летописцем.

С помощью средств НКРЯ можно проанализировать различные представленные в корпусе древнерусские тексты, включая «Житие Феодосия Печерского».

Для анализа структуры и семантики «Жития Феодосия Печерского» в НКРЯ необходимо обратиться к разделу «Древнерусский корпус» и выбрать подкорпус «Житие Феодосия» (древнейшее восточнославянское житие: описание жизни Киево-Печерского игумена преподобного Феодосия, написанное монахом Нестором. Текст приводится по Успенскому сборнику).

Разногласия исследователей в вопросе о времени написания «Жития Феодосия Печерского» существуют по сей день. В НКРЯ «Житие Феодосия Печерского» появилось в 2022 году - в Древнерусском корпусе и подкорпусе «Житие Феодосия». Характеризуется как «древнейшее восточнославянское житие: описание жизни Киево-Печерского игумена преподобного Феодосия, написанное монахом Нестором. Текст приводится по Успенскому сборнику». Содержит 167 предложений, 19571 словоформ. Разновидность языка: церковнославянский.

Существуют различные датировки написания Жития: 1080-е годы (исходя из описания ряда событий произведение было написано не позднее 1088 года) или начало XII века. В НКРЯ дата создания «Жития Феодосия Печерского»: 1080-е. Дата списка: XII – XIII в. По мнению А.А. Шахматова, «Житие Феодосия Печерского» было написано до 1088 г. и основанием для такого заключения ему послужили факты из «Жития Феодосия Печерского».

Обращение к древнерусскому и старорусскому корпусам НКРЯ (даже без снятой омонимии в последнем), несмотря на сравнительно небольшой

объём, уже является весьма действенным методом получения новых примеров, фронтального просмотра материала, пополнения исторических грамматик и (для лексико-грамматических феноменов) словарей. В процессе исследования решены следующие задачи: исследованы корпусные методы изучения художественной литературы; дана характеристика методов корпусной лингвистики в изучении художественных текстов; рассмотрены корпусные методы в изучении древнерусской литературы; проведен анализ структуры и семантики «Жития Печерского» средствами НКРЯ.

Древнерусский подкорпус моложе многих других подкорпусов НКРЯ, и практика его использования еще невелика. Он, безусловно, нуждается в усовершенствованиях, в частности в унификации параметров разметки для всех текстов, в оптимизации словаря лемм, в расширении метаданных, в развитии поисковых возможностей. Ресурс для этого имеется, поскольку потенциал осуществленной разметки памятников шире того, что представлено на данный момент в древнерусском подкорпусе. Так, может быть добавлен поиск по греческим леммам и их славянским соответствиям; по альтернативным разборам словоформ, которые допускают неоднозначную трактовку; по ошибочным и испорченным словоформам и др. Тем не менее уже сейчас подкорпус позволяет получать разнообразную информацию, а главное, открывает неожиданные аспекты в изучении древнерусской письменности и ставит перед исследователями нетривиальные вопросы.

Цель исследования - изучение «Жития Феодосия Печерского» средствами корпусной лингвистики – достигнута.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Акимова, Э. Н. Древнерусский текст в лингвистических исследованиях XX–XXI вв. [Текст] / Э. Н. Акимова. Саранск: Изд-во Мордов. ун-та, 2012. - С. 1-10.
2. Андриющенко В.М. Концепция и архитектура машинного фонда русского языка. М., 1989.
3. Балута А.А. Эволюция средств выражения ирреальной модальности в сложноподчиненном предложении русского языка (на материале первичных евангельских текстов), дис. ... канд. филол. наук, Москва, 2005.
4. Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие. — М.: Эдиториал УРСС, 2001. — 360 с.
5. Баталина К.Е. Абстрактные имена существительные и категории сакрального текста как средства экспликации концептов христианской картины мира в евангельских чтениях: На материале Апракоса Мстислава Великого 1115–1117 гг., дис. ... канд. филол. наук, Нижний Новгород, 2005. – С. 164.
6. Брукер Н.И. Конструкции цепного нанизывания с синтаксической частицей ЖЕ. [Текст] / Н.И. Брукер. [в:] Семантика и прагматика языковых единиц, Душанбе, 1990. - С. 163–174.
7. Ваулина С.С. Эволюция средств выражения модальности в русском языке (XI–XVII вв.). [Текст] / С.С. Ваулина. Ленинград: Издательство Ленинградского университета, 1988.
8. Вербицкая Л.А., Казанский Н.Н., Касевич В.Б. Некоторые проблемы создания национального корпуса русского языка // Научно-техническая информация. Сер. 2. 2003. № 6. С. 2–8; Шаров С.А. Представительный корпус русского языка в контексте мирового опыта // Там же. С. 9–18.

9. Виноградов В.В. Избранные труды: О языке художественной прозы. [Текст] / В.В. Виноградов. Москва: Наука, 1980. – С. 70.

10. Воейкова М. Д. Проблемы использования подкорпуса устной разговорной речи (на примере анализа русских диминутивов) // Национальный корпус русского языка: 2006—2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009, 353—373.

11. Гришина Е. А. Устная речь в Национальном корпусе русского языка // Национальный корпус русского языка: 2003—2005. М.: Индрик, 2005, 94—110.

12. Гришина Е. А. О маркерах разговорной речи (предварительное исследование подкорпуса кино в Национальном корпусе русского языка) // Компьютерная лингвистика и интеллектуальные технологии. Труды международной конференции «Диалог'2007» (Бекасово, 30 мая — 3 июня 2007 г.), 147—156.

13. Гришина Е. А., Плунгян В. А. Перспективы развития Национального корпуса русского языка // Национальный корпус русского языка: 2003—2005. М.: Индрик, 2005, 330—339.

14. Гришина Е. А., Савчук С. О. Корпус устных текстов в НКРЯ: состав и структура // Национальный корпус русского языка: 2006—2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009, 129—149.

15. Дич Н. Л. О текстах XIX века в Национальном корпусе русского языка // Национальный корпус русского языка: 2003—2005. М.: Индрик, 2005, 89—93.

16. Добрушина Н. Р. Как использовать Национальный корпус русского языка в образовании? // Национальный корпус русского языка: 2003—2005. М.: Индрик, 2005, 308—329.

17. Добровольский Д.О. Корпус параллельных текстов и литературный перевод // Научно-техническая информация. Сер. 2. 2003. № 10. С. 13–18; Копотев М.В., Мустайоки А. Принципы создания Хельсинкского аннотированного корпуса русских текстов // Научно-техническая

информация. Сер. 2. 2003. № 6. С. 33–37; Копотев М.В. Корпусная лингвистика в Финляндии (обзор ресурсов) // Там же. С. 37–43; Михайлов М.Н. Стыковка параллельных текстов в автоматическом режиме: иллюзии и перспективы // Научно-техническая информация. Сер. 2. 2003. № 10. С. 18–27.

18. Житие Феодосия Печерского. Поэтика и проблематика. — Текст: электронный // Myfilology.ru – информационный филологический ресурс: [сайт]. – URL: https://myfilology.ru/russian_literature/russkaya-literatura-xi-xvii-vekov/zhitie-feodosiia-pecherskogo-poetika-i-problematika/ (дата обращения: 27.04.2022).

19. Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005. – 48 с.

20. Захаров В.П. Поисковые системы Интернета как инструмент лингвистических исследований // Русский язык в Интернете. Казань, 2003.

21. Кагарлицкий Ю.В., Пичхадзе А.А., Шаров С.А. Проблемы создания электронного корпуса переводных памятников древнерусской письменности XI–XII вв. // Научно-техническая информация. Сер. 2. 2003. № 10. С. 1–5; Кибрик А.А., Подлеская В.И. К созданию корпусов письменной устной речи: принципы транскрибирования // Там же. С. 5–12; Чардин И.С. Лингвистические корпуса с синтаксической разметкой и их применение // Научно-техническая информация. Сер. 2. 2003. № 6. С. 18–25.

22. Ковалев Н.С., Древнерусский литературный текст: Проблема исследования смысловой структуры и эволюции в аспекте категории оценки, Волгоград: Издательство Волгоградского государственного университета, 1997.

23. Колесов В.В. Древнерусский литературный язык. [Текст] / В.В. Колесов. Ленинград: Издательство Ленинградского университета, 1989.

24. Клименко Л.П. Проблема языка, состава текста и жанра старопечатных памятников письменности XVII–XVIII вв. [Текст] / Л.П.

Клименко. Нижний Новгород: Издательство Нижегородского государственного университета, 1997. – С. 3 - 4.

25. Копотев М. В., Янда Л. А. Национальный корпус русского языка // Вопросы языкознания. 2006. № 5. С. 149—155.

26. Кустова Г.И., Ляшевская О.Н., Падучева Е.В., Рахилина Е.В. Национальный корпус русского языка как инструмент семантикограмматического исследования лексики // Международная конференция «Корпусная лингвистика–2004». Тезисы докладов. СПб., 2004. С. 50–52; Ляшевская О.Н., Плунгян В.А., Поляков А.Е., Савчук С.О., Сичинава Д.В. Обработка текстов для Национального корпуса русского языка: технологическая цепочка // Там же. С. 54–57; Шаров С.А., Савчук С.О. Типология текстов для представительного корпуса // Там же. С. 93–94.

27. Кустова Г. И., Ляшевская О. Н., Падучева Е. В., Рахилина Е. В. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. — М., 2005, 155—174.

28. Кустова Г. И., Ляшевская О. Н., Падучева Е. В., Рахилина Е. В. Опыт семантического расширения морфологической разметки: таксономическая классификация лексики в Национальном корпусе русского языка // Научная и техническая информация, сер. 2. Информационные процессы и системы, 2005, N 6.

29. Курзина Е.С. Древнерусский сборник как гипертекст (к постановке проблемы). [Текст] / Е.С. Курзина [в:]. Книга и мировая цивилизация. Материалы одиннадцатой международной научной конференции по проблемам книговедения, Москва: Наука, с. 9–13.

30. Кустова Г. И., Толдова С. Ю. НКРЯ: семантические фильтры для разрешения многозначности глаголов // Национальный корпус русского языка: 2006—2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009, 258—276.

31. Лихачев Д.С. Поэтика древнерусской литературы. [Текст] / Д.С. Лихачев. [в:] Избранные работы: в 3 т., Ленинград, т. 1, 1987. - С. 261–648.
32. Лихачев Д.С. Развитие русской литературы X–XVII веков. Эпохи и стили. [Текст] / Д.С. Лихачев. [в:] Избранные работы: в 3 т., Ленинград, т. 1, 1987. - С. 24–260.
33. Ляшевская О. Н. О частотном словаре Национального корпуса русского языка // Слово и словарь = Vocabulumetvocabulary: сб. науч. тр. по лексикографии. Гродно: ГрГУ, 2007.
34. Ляшевская О. Н., Плунгян В. А., Сичинава Д. В. О морфологическом стандарте Национального корпуса русского языка // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. — М., 2005. 111—135.
35. Мамонтова В. В. Корпусная лингвистика в современной языковедческой парадигме // Актуальные вопросы современной науки. 2010. №12.
36. Матхаузерова С. Две теории текста в русской литературе XVII в. [Текст] / С. Матхаузерова. Ленинград: Труды Отдела древнерусской литературы, т. 31, 1976. - С. 271–284.
37. Молдован А. М., Плунгян В. А., Сичинава Д. В. Национальный корпус русского языка: новые перспективы русистики // Русский язык: исторические судьбы и современность: III Международный конгресс исследователей русского языка (Москва, МГУ им. М. В. Ломоносова, филологический факультет, 20-23 марта 2007 г.): Труды и материалы / Сост. М. Л. Ремнева, А. А. Поликарпов. — М., МАКС Пресс, 2007, 26—27.
38. О древнерусском корпусе: [Электронный ресурс]. – URL: https://ruscorpora.ru/new/help-old_rus.html (Дата обращения: 25.04.2022).
39. Опарина А.В. Специфика проявления авторской модальности в списках «Повести временных лет»: Лексико-грамматический аспект, дис. ... канд. филол. наук, Тамбов, 2004.

40. Орехов Б. В. Параллельный корпус переводов «Слова о полку Игореве»: итоги и перспективы // Национальный корпус русского языка: 2006—2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009, 462—473.

41. Оскольская С. А. Корпус письменных текстов XIX века: сферы употребления и жанровое разнообразие // Национальный корпус русского языка: 2006—2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009, 46—51.

42. Отин Е. С. Модальные отношения в конструкциях чужой речи и средства их выражения в русском языке XIII—XVII вв., автореф. дис. ... канд. филол. наук, Днепропетровск, 1963.

43. Официальный сайт Национального корпуса русского языка: [Электронный ресурс]. – URL: <https://ruscorpora.ru/new/> (Дата обращения: 10.04.2022).

44. О репрезентативности корпусов см. важную статью с высоким индексом цитируемости: Biber D. Representativeness in Corpus Design // *Literary & Linguistic Computing* 8 (1993). P. 243–257, а также содержательный обзор литературы вопроса: Grönqvist L., Helgadóttir S. *Literature review on representativeness of linguistic resources* (2002) // <http://www.gslt.hum.gu.se/~leifg/gslt/doc/rep021202.doc>. Научный подход к этой проблематике, включая разработку детального инвентаря признаков текстов, характерен уже для разработок англоязычных Брауновского [1960-е годы] и Ланкастерско-Осло-Бергенского корпуса – см. Johansson S., Leech G.N., Goodluck H. *The Lancaster-Oslo-Bergen Corpus Manual* [1978] // <http://khnt.hit.uib.no/icame/manuals/lob/INDEX.HTM>).

45. Перцов Н. В. О роли корпусов в лингвистических исследованиях // Труды международной конференции «Корпусная лингвистика–2006». — СПб.: Изд-во С. Петерб. ун-та, 2006, 318—331.

46. Пиккио Р. Древнерусская литература. [Текст] / Р. Пиккио. Москва: Языки славянской культуры, 2002.

47. Пильщиков И. А., Старостин А. С. Основные проблемы автоматизации базовых процедур ритмико-синтаксического анализа силлаботонических текстов // Национальный корпус русского языка: 2006—2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009, 298—316.

48. Пичхадзе А. А. Корпус древнерусских переводов XI-XII вв. и изучение переводной книжности Древней Руси // Национальный корпус русского языка: 2003—2005. М.: Индрик, 2005, 251—262.

49. Плунгян В. А., Сичинава Д. В. Национальный корпус русского языка: опыт создания корпусов текстов современного русского языка // Л. Н. Беляева и др. (ред.). Труды Межд. конференции «Корпусная лингвистика-2004». СПб: СПбГУ, 2004, 216—238.

50. Плунгян В.А. Корпус как инструмент и идеология: о некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении. – 2008. - № 2 (16). – С. 7-20.

51. Плунгян В.А., Сичинава Д.В. Морфологическая информация в Национальном корпусе русского языка // II Международный конгресс исследователей русского языка «Русский язык: исторические судьбы и современность». Москва, 18–21 марта 2004 г. Труды и материалы. М.: МГУ, 2004. С. 483–484.

52. Плунгян В. А. Зачем мы делаем Национальный корпус русского языка? // Отечественные записки. 2005. № 2, 296—308.

53. Потапова Р. К. Новые информационные технологии и лингвистика. – 2-е изд. – М. : УРСС, 2004. – 320 с.

54. Припадчев А.А. Гносеология, прагматика и семантика в диахронии синтаксиса текста. [Текст] / А.А. Припадчев. Воронеж: Курьер, 1992. – С. 13.

55. Райскина В. А., Дубнякова О. А. Современные методы корпусной лингвистики при анализе текста (на примере корпуса ВФМ) // Актуальные вопросы современной науки: сборник научных трудов. Новосибирск: Издательство ЦРНС, 2015. Вып. 40 / под общ. ред. С. С. Чернова. С. 146-155.

56. Рахилина Е. В., Кобрицов Б. П., Кустова Г. И., Ляшевская О. Н., Шеманаева О. Ю. Многозначность как прикладная проблема: лексико-семантическая разметка в национальном корпусе русского языка // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2006» (Бекасово, 31 мая — 4 июня 2006 г.) / Под ред. Н. И. Лауфер, А. С. Нариньяни, В. П. Селегея. — М.: Изд-во РГГУ, 2006.

57. Ринберг В.Л. Древнерусские сложные конструкции и их современные параллели (Спецкурс по русскому языку). [Текст] / В.Л. Ринберг. Киев: Издательство Киевского государственного университета. — С. 29.

58. Рыков В.В. Корпус текстов как новый тип словесного единства // Труды Междунар. семинара «Диалог-2003». М.: Наука, 2003. С. 15–23.

59. Рычкова, Л. В. Специализированный лингвистический корпус, создаваемый для исследования социокультурной специфики регионального варианта русской речи // Писменото наследство и информационните технологии : материали от V международна науч. конф., Варна, 15–20 септ. 2014 г. / отг. ред. В. А. Баранов, В. Желязкова, А. М. Лаврентьев. — София ; Ижевск : Кирило-Методиевски научен център БАН; Ижевский гос. техн. ун-т, 2014. — С. 167–170.

60. Рычкова, Л. В. Отражение языков для специальных целей в СМИ Гродненщины // Русский язык : система и функционирование (к 75-летию филологического факультета БГУ) : сб. материалов VI Междунар. науч. конф., Минск, 28–29 окт. 2014 г. : в 2 ч. / редкол.: И. С. Ровдо (отв. ред.) [и др.]. — Минск : Изд. центр БГУ, 2014. — Ч. 2. — С. 266–271.

61. Рычкова, Л. В., Станкевич А. Ю. «Диалог» языков на страницах СМИ Гродненщины // Актуальные проблемы теории дискурса: материалы Междунар. электрон. конф., Актюбинск, 30 мая 2014 г. / «Actualproblemsofthetheoryofdiscourse» proceedings of the international scientific

e-conference, May 30, 2014. – Актобе : Актюбинский регион. гос. ун-т им. К. Журбанова. – С. 51–61.

62. Рычкова, Л. В., Станкевич, А. Ю. Многоцелевой лингвистический корпус региональных СМИ / Л. В. Рычкова, А. Ю. Станкевич // Информационно-коммуникативные технологии в лингвистике, лингводидактике и межкультурной коммуникации : сб. науч. ст. Вып. 6. / М-во образования и науки РФ [и др.] ; под ред. А. Л. Назаренко. – М. : Центр дистанц. образования : Факультет иностр. яз. и регионоведения МГУ им. М. В. Ломоносова; Изд-во «Университетская книга», 2014. – С. 488–497.

63. Савчук С. О. Корпус текстов первой половины XX века: текущее состояние и перспективы // Национальный корпус русского языка: 2006—2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009, 27—45.

64. Савчук С. О, Сичинава Д. В. Обучающий корпус русского языка и его использование в преподавательской практике // Национальный корпус русского языка: 2006—2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009, 317—334.

65. Селезнева А.А. Тенденции формирования функционально-смысловых типов речи в летописном тексте (на материале «Повести временных лет»), дис. ... канд. филол. наук, Москва, 2004.

66. Сичинава Д.В. Об одном лингвистическом параметре типологии текстов: коэффициент под/над // Научно-техническая информация. Сер. 2. № 10. 2003. Р. 27–34.

67. Сичинава Д. В. Национальный корпус русского языка: очерк предыстории // Национальный корпус русского языка: 2003—2005. М.: Индрик, 2005, 21—30.

68. Станкевич, А. Ю. Технология сбора и систематизации электронного контента для корпуса русскоязычных СМИ Гродненщины // Карповские научные чтения : сб. науч. ст. Вып. 8 : в 2 ч. / редкол.: А. И. Головня (отв. ред.) [и др.]. – Минск : Беларус. Дом печати, 2014. – Ч. 2. – С. 6–10.

69. Топоров В. Н. Святость и святые в русской духовной культуре: в 2 т. – М.: «Гнозис»; Школа «Языки русской культуры», 1995. – Т. 1., С. 665.

70. Трестерова З. К некоторым особенностям связности древнерусского текста и их роли в перестройке «типа» синтаксиса. [Текст] / З. Трестерова. [в:] Восточные славяне: языки, история, культура, Москва: Наука, 1985. – С. 15.

71. Черепанова О.А. Лексико-грамматические средства выражения модальности в русском языке XI–XVII вв., автореф. дис. ... канд. филол. наук, Ленинград, 1965.

72. Шахматов А. А. Несколько слов о Несторовом житии Феодосия // Сборник отделения русского языка и словесности. — 1896. — Т. 64, № 1.

73. Шмелев А. Д. Языковые факты и корпусные данные // Русский язык в научном освещении, 2010, 19 (1), 236—265.

74. Biber D., Conrad S, Reppen R. Corpus Linguistics: Investigating Language Structure and Use. Cambridge: Cambridge University Press, 1998. P. 248.

75. Biber D. A Typology of English Texts // Linguistics. Vol. 27. 1989. P. 3–43.