

Федеральное государственное бюджетное образовательное учреждение высшего
образования

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

На правах рукописи

МИРОНОВА ДИНА МАРКОВНА

**АВТОМАТИЗИРОВАННАЯ КЛАССИФИКАЦИЯ ДРЕВНИХ
РУКОПИСЕЙ**

(На материале 525 списков славянского Евангелия от Матфея XI-XVI вв.)

Специальность 10.02.21 – Прикладная и математическая лингвистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата филологических наук

Санкт-Петербург

2017

Работа выполнена в федеральном государственном бюджетном образовательном учреждении
 высшего образования
 «Санкт-Петербургский государственный университет»

Научный руководитель:	Алексеева Елена Леонидовна, Кандидат филологических наук, доцент, ФГБОУ ВО «Санкт-Петербургский государственный университет», доцент кафедры математической лингвистики.
Официальные оппоненты:	Пичхадзе Анна Абрамовна, доктор филологических наук, ведущий научный сотрудник Отдела лингвистического источниковедения и истории русского литературного языка ИРЯ РАН им. Виноградова.
	Бонч-Осмоловская Анастасия Александровна, кандидат филологических наук, доцент, Национальный исследовательский университет «Высшая школа экономики», доцент Школы лингвистики.
Ведущая организация:	ФГБОУ ВО «Ижевский государственный технический университет имени М. Т. Калашникова»

Защита состоится «_____» _____ 2018 г. в 16-00 часов на заседании совета Д 212.232.23 по защите диссертаций на соискание степени доктора филологических наук, на соискание степени кандидата филологических наук при ФГБОУ ВО «Санкт-Петербургский государственный университет» по адресу: 199034, Санкт-Петербург, Университетская наб. д. 11, ауд. 191.

С диссертацией можно ознакомиться в Научной библиотеке им. М. Горького Санкт-Петербургского государственного университета (г. Санкт-Петербург, Университетская набережная, д. 7/9) и на сайте: <https://disser.spbu.ru/disser/soiskatelyu-uchjonoj-stepeni/dislist/details/14/1511.html>.

Автореферат разослан «_____» _____ 2018 г.

Ученый секретарь
 диссертационного совета
 кандидат филологических наук

К. В. Манерова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Сегодня в славянской медиевистике все большее распространение получает использование компьютерных методов, направленных на решение различных задач, таких как создание корпусов текстов, сравнение текстов рукописных памятников одного произведения и выявление разночтений (колляция) с последующей подготовкой критического издания или классификацией рукописей, преобразование текста критического издания в структурированный корпус¹.

Одной из задач компьютерной текстологии в рамках автоматизации лингвистических исследований является классификация памятников регулярного копирования, использовавшихся в средневековье при совершении богослужения и сохранившихся в большом числе списков. Решение такой задачи направлено на репрезентативный отбор источников для критического аппарата издания.

Актуальность темы данного исследования обусловлена неоспоримой важностью изучения библейской традиции как на языках оригиналов, так и переводных версий. Эта традиция играет исключительную роль в истории европейской культуры и представлена большим числом памятников, равно как большим числом посвященных ей исследований. Однако методика автоматизированной классификации древних рукописей остается мало разработанной, нет ни одного исследования, которое можно было бы считать окончательным, и выработка оптимального компьютерного метода призвана способствовать решению основных задач библейской текстологии. В данном диссертационном исследовании проблема автоматизированной классификации рассматривается с позиций кластерного анализа и компьютерной стемматологии.

Объектом диссертационного исследования являются церковнославянские рукописи Евангелия XI–XVI вв. **Предметом** исследования является разработка оптимального алгоритма классификации рукописных источников путем их разбиения на группы по признаку текстовой близости при помощи кластерного анализа и методов стемматологии с последующим определением характерных особенностей выявленных групп.

¹ Обширная электронная коллекция славянских средневековых рукописей создана в Ижевском государственном техническом университете под руководством В. А. Баранова, в Институте русского языка им. В. В. Виноградова РАН разработаны электронные базы данных древнерусских рукописных памятников. Ряд текстологов используют программу автоматизированной колляции Collate, разработанную П. Робинсоном в Кембриджском университете. Методы компьютерной классификации рукописей применяются на этапе подготовки критического издания, в частности, издания Евангелия от Иоанна и Евангелия от Матфея в славянской традиции, позволяя более точно и полно атрибутировать тексты списков традиции.

Целью данного исследования является формализация критериев отбора текстовых фрагментов для текстологического анализа и последовательное практическое исследование влияния различных типов таких фрагментов на результат объединения рукописей в осмысленные группы посредством автоматического кластерного анализа.

Для достижения поставленной цели в диссертации решается ряд **задач**:

- 1) описание методов компьютерной текстологии, используемых для классификации рукописных традиций в отечественной и зарубежной текстологии;
- 2) выявление методов компьютерной текстологии, применимых к церковнославянской евангельской рукописной традиции;
- 3) на основании методов, применимых к церковнославянской евангельской рукописной традиции – кластерного анализа и трехуровневого метода компьютерной стемматологии, – разработка оптимальной методики автоматизации отбора текстовых фрагментов, позволяющей с наименьшими ограничениями и наименьшими затратами усилий решать поставленную задачу;
- 4) подробное описание методики вычленения узлов разночтений, примененной в данной работе;
- 5) подготовка списка узлов разночтений для пассажа Евангелия от Матфея 14.14. – 14.34 по 525 рукописям;
- 6) осуществление классификации 525 списков Евангелия от Матфея по выявленным узлам разночтений с помощью описанных методов и проведение сравнительного анализа результатов;
- 7) исследование зависимости автоматизированной классификации рукописей от типа учитываемых узлов разночтений и определение текстологической значимости каждого типа узлов;
- 8) разработка программно реализованной процедуры выделения текстологических примет для групп рукописей.

Материалом исследования послужили коллажи пассажа Евангелия от Матфея 14.14 – 14.34 объемом в 300 словоупотреблений по 525 церковнославянским рукописям. Исследование выполнялось на материале церковнославянской евангельской традиции на более чем 1500 рукописях, полностью или частично содержащих текст славянского перевода Евангелий². Очевидно, что пока нет

² До начала подготовки коллажей самый полный список источников церковнославянского Евангелия, составленный Л. П. Жуковской, включал около 500 источников XI-XV вв. При подготовке издания Евангелия от Иоанна был составлен список из 1500 единиц. Оба эти списка были составлены *de visu*. Уже после окончания коллажей М. Гарзанити опубликовал список из 2500 единиц, составленный по опубликованным описаниям. В действительности рукописей Евангелия больше, чем 2500; множество рукописей XVI-XVII вв. до сих пор не включено в описания.

возможности сравнить такое большое число рукописей по всему тексту, поэтому было решено ограничиться одним текстовым пассажем. При его выборе учитывалось, что по способу представления евангельского текста существует два основных типа рукописей: четьи (четвероевангелия) с последовательным расположением текста четырех евангелий, и служебные (апракосы), в которых текст приводится не подряд, а в том порядке, в каком он читается в церкви в течение церковного года. Соответственно, нужно было подобрать отрывок, который был бы непрерывным и в апракосах. В ходе коллаций было замечено, что рукописи XV–XVI вв. в значительном большинстве входят в одну группу, на базе которой в XVI в. появились печатные издания. Это наблюдение позволило считать, что 525 рукописей, подвергшихся коллациям, обеспечивает достаточно полное представление материала, при том что в это число вошло подавляющее большинство рукописей XI–XIV вв. и значительное число рукописей XV в.

Методы исследования рукописной традиции, использованные в работе, включают кластерный анализ, основанный на объединении объектов в группы по сходству, и трехуровневый метод, подразделяющий процесс построения стеммы на три этапа, первый из которых опирается на кластерный анализ. Для описания узлов разночтений применяются методы морфологического, синтаксического и лексического анализа.

Теоретическую и методологическую основу диссертации составляют труды отечественных и зарубежных ученых, посвященные компьютерной текстологии (И. В. Азарова, А. А. Алексеев, Е. Л. Алексеева, Е. В. Афанасьева, М. Баккер, Дж. Брефельд, В. А. Баранов, М. Вайцман, Н. Л. Горина, Дж. Гриффит, П. Гэлловой, В. Диаринг, А. Кириченко, А. А. Пичхадзе, Б. Салеманс, М. Спенсер, П. Робинсон, Д. Фроже, К. Хау, А. Хруби), кластерному анализу (М. Андерберг, А. Гордон, Э. Дидей, Р. Кормак, В. Кржановский, П. Снит, Б. Эверитт), текстологическим исследованиям древних славянских памятников (И. В. Азарова, А. А. Алексеев, Е. Л. Алексеева, М. Баккер, Г. А. Воскресенский, М. Гарзанити, Н. Л. Горина, Л. П. Жуковская, Д. С. Лихачев, А. А. Пичхадзе, Й. ван дер Так, С. Ю. Темчин, У. Федер), текстологии Нового Завета (Л. Вагане, К. Вахтель, Э. Колвелл, К. Лахман, П. Маас, Б. Мецгер, Г. Минк, А. ден Холландер, Э. Эпп), компьютерной стемматологии (Э. Ваттель, А. Деес, Дж. Зарри, Э. Лангбрек, Э. Маер, Х. Мейеринг, У. Мерисало, М. ван Мулкен, Э. Пул, Б. Салеманс, У. Смелик, К. Уteman, К. Флайт, М. Хогенхаут-Мулдер, А. ден Холландер, Л. Шослер).

Научная новизна результатов исследования состоит в следующем:

1) впервые разные методы компьютерной классификации применены к обширному рукописному материалу;

2) на основании результатов автоматизированной классификации с применением методов компьютерной текстологии Алексева и Ваттеля произведена оценка достоинств и недостатков данных методов, позволившая выделить из них наиболее оптимальный;

3) автором разработана и программно реализована процедура выделения текстологических примет групп рукописей, что позволяет опираться на наиболее релевантные текстологические чтения при построении классификации;

4) опираясь на опыт авторитетных текстологов, автор подробно описывает алгоритм выделения узлов разночтения.

Теоретическая значимость результатов исследования заключается

1) в демонстрации валидности применения методов компьютерной обработки рукописной традиции путем сравнения двух независимо работающих методов;

2) в создании комплексной системы описания узла разночтений, базирующейся на текстологических принципах изучения рукописной традиции и формальных методах классификации рукописей, правил выделения узла разночтений и определения его границ;

3) в разработке методики выделения текстологических примет для классификации рукописных источников текста.

Степень разработанности проблемы

В XX веке разрабатываются формальные методы исследования рукописных традиций. Производится поиск наиболее оптимальных методов определения отношений между существующими рукописными источниками с целью выявления картины исторического развития содержащегося в них текста. В частности, в начале 80-х гг. широкое применение получают вариации кластерного анализа, когда объекты группируются по некоторым признакам сходства или различия. Данные применения метода кластерного анализа в медиевистике приводятся в работах Э. Колвелла, А. Хруби, Э. Пула, Дж. Гриффита, П. Гэлловой, Дж. Брефельд, А. А. Алексева, А. А. Пичхадзе, Н. Л. Гориной, Е. В. Афанасьевой. Методами компьютерной стемматологии для классификации древних текстов пользуются А. ден Холландер, У. Мерисало, М. Ван Мулкен, Л. Шослер, М. Хогенхаут–Мулдер, Х. Мейеринг, К. Утеман, Дж. Зарри, Э. Лангбрек, У. Смелик, Э. Маер.

С развитием технической базы появляется возможность автоматизировать классификацию любого количества источников. Перед современной компьютерной текстологией сегодня стоит задача оптимизации существующих методов с целью решения проблемы установления генеалогических или типологических связей в рамках обширных рукописных традиций.

Как в отечественной, так и в зарубежной текстологии большое внимание уделяется разработке проблемы определения узла разночтений как основания классификации рукописных источников. Однако несмотря на значительную степень разработанности проблемы определения узла разночтений в славянской традиции (в частности, в работах Л. П. Жуковской, А. А. Пичхадзе, Н. Л. Гориной), системного описания правил выделения узлов разночтений и их длины для церковнославянских рукописных памятников явным образом не было сформулировано.

Практическая значимость результатов исследования состоит в возможности применения результатов классификации для подготовки критического издания церковнославянского Евангелия. Данные о полученных группировках рукописей были использованы при подготовке критического аппарата текста издания Евангелия от Матфея, опубликованного в 2005 г. В настоящее время в СПбГУ завершается работа по подготовке к публикации славянских текстов Евангелий от Марка и Луки, в которой использованы результаты данного диссертационного исследования.

Рекомендации по использованию результатов исследования. Полученные результаты могут найти применение в исследовании других рукописных традиций с применением компьютерных методов, а также в курсах по компьютерным технологиям, корпусной лингвистике, текстологии и истории старославянского языка.

В результате исследования сформулированы и **выносятся на защиту следующие положения.**

1. Среди существующих методов компьютерной классификации рукописей наиболее оптимальным для церковнославянской новозаветной традиции является метод кластерного анализа Алексева. Группировка рукописей с применением данного метода позволяет обеспечить наиболее удобное визуальное представление кластеров, а также оставляет возможность корректировки результатов после завершения компьютерной обработки. Данный метод позволяет избежать внесения субъективной оценки на этапе создания классификации и является наиболее надежным инструментом при решении задач автоматизированного текстологического исследования церковнославянского евангельского текста.

2. Изучение разночтений, представленных в исследуемом материале, и обобщение данных, полученных текстологами в работе со славянскими памятниками постоянного копирования, позволяет сформулировать определение понятия длины узла разночтений как минимального элемента текста, изменение которого не повлекло за собой изменение других элементов. Предложенное определение может применяться к любой традиции и позволяет оценивать значимость различных лексико-грамматических и текстовых категорий вне связи с другими категориями.

3. Выведенное правило формального определения примет для автоматизированной классификации рукописной традиции позволяет оптимизировать метод Алексеева с тем, чтобы при классификации можно было учитывать только текстологически релевантные чтения – приметы. Также определены виды узлов разночтений с низкой релевантностью, которые можно не учитывать на этапе подготовки коллажей.

Достоверность результатов подтверждается тем, что сравнительный анализ результатов применения методов Алексеева и Ваттеля показал совпадение группировок рукописей. Кроме того, полученные группы совпадают с результатами, представленными в работах Г. А. Воскресенского, Л. П. Жуковской, С. Ю. Темчина, а также с экспертной оценкой текстологов, отраженной в критическом издании Евангелия от Матфея 2005 г. Программы выполнены на сертифицированном оборудовании в СПбГУ Е. Л. Алексеевой (Кузнецовой) и Д. М. Мироновой и в Свободном университете Амстердама Э. Ваттелем.

Апробация работы. Отдельные аспекты и основные положения диссертационного исследования обсуждались на многих отечественных и международных конференциях и семинарах, в частности: на международных филологических конференциях преподавателей и аспирантов в Санкт-Петербургском государственном университете (2000, 2010, 2011, 2014, 2016, 2017 гг.), на международной конференции “Studies in Stemmatology” (Васенаар, апрель 2000; Амстердам, Свободный университет, октябрь 2000), на шестой международной конференции «From Alpha to Byte» международной ассоциации «Библия и информатика» (Стелленбос, Стелленбосский университет, июль 2000). По теме диссертации опубликовано 11 работ, в том числе статья, опубликованная в издании, включенном в индексы цитирования Web of Science («Bible and Computer : the Stellenbosch AIBI-6 Conference : Proceedings of the Association Internationale Bible et Informatique “From Alpha to Byte”, University of Stellenbosch, 17–21 July, 2000, ed. by J. Cook, 2002») и две статьи в российских изданиях, рекомендованных ВАК РФ («Структурная и прикладная лингвистика» №9, 2012 и «Филологические науки. Вопросы теории и практики» № 3, Ч. 1, 2017).

Объем и структура работы. Диссертация состоит из введения, трех глав, заключения, списка принятых сокращений, списка литературы, списка иллюстративного материала и четырех приложений. Общий объем работы составляет 315 страниц, основное содержание изложено на 182 страницах и содержит 27 рисунков и 25 таблиц, приложения занимают 133 страниц. Список литературы состоит из 135 наименований, включающих 97 наименований на иностранных языках.

СОДЕРЖАНИЕ РАБОТЫ

Во **Введении** обосновываются актуальность темы диссертационного исследования, ее научная новизна и практическая значимость. Формулируются цель и задачи работы, определяются объект, предмет и методы исследования, описывается структура диссертации.

В **Первой главе «Методы компьютерной обработки рукописных традиций»** описаны алгоритмы классификации рукописей методами компьютерной стемматологии и кластерного анализа, позволяющими работать с неограниченным количеством источников и контаминированной традицией.

Проблема восстановления авторского текста по имеющимся рукописям была осознана еще в античную эпоху, когда библиотекари в Александрии попытались восстановить гомеровские тексты. До середины XVIII века лучшие чтения выбирались по смыслу, без учета отношений между текстами в разных рукописях: если некое чтение встречалось только в одной рукописи, но казалось по смыслу наиболее подходящим, оно принималось за чтение архетипа и было предпочтительно при издании текста, даже если другое чтение повторялось в большинстве рукописей. Важные изменения произошли во второй половине XVIII века, когда было сформулировано правило о том, что лучшие чтения необходимо выбирать, опираясь на данные всех источников. Это время и можно считать началом современной текстологии, задачей которой является подготовка научного издания текста, где приводятся данные из разных источников и объясняются особенности различных вариантов чтений.

Наиболее тиражируемые рукописные тексты, являющиеся основным объектом для применения компьютерной текстологии, можно разделить на два типа: литературные памятники и библейские тексты.

Способы копирования библейских и авторских произведений несколько отличались. При работе с библейским текстом от писца требовалась максимальная корректность передачи текста антиграфом. Это требование диктовалось догматическими и литургическими соображениями, использовались различные средства надзора за стабильностью текста. Кроме этого, один литературный памятник обычно сохраняется в небольшом количестве экземпляров, а число библейских текстов исчисляется тысячами, в силу чего возможности исследования данных традиций значительно различаются.

Подход к выбору метода для текстологического исследования рукописной традиции определяется типом традиции. Для памятников, количество списков которых не превышает нескольких десятков экземпляров, широко используются

методы кладистики и стемматологии, целью которых является представление истории текста в виде дерева отношений между рукописями. Однако для библейской традиции, насчитывающей несколько тысяч свидетелей и характеризующейся контаминацией, невозможно построить дерево отношений между отдельными рукописями и необходимо искать метод, позволяющий выявить группы наиболее близких между собою рукописей, и анализировать отношения уже между этими группами.

Революционным стал метод, впервые предложенный двумя американскими текстологами – Э. Колвеллом и Э. Тьюном в 1963 г. для греческой новозаветной традиции. Э. Колвелл и Э. Тьюн предложили отказаться от эталона, сравнивать все рукописи попарно между собой и оценивать их близость по общему тексту, а не по различиям. К концу XX в. возможности вычислительной техники позволили широко и эффективно применять данный метод в текстологии, опираясь на математический аппарат кластерного анализа. Вариации кластерного анализа активно используются в современных текстологических исследованиях как для поиска групп рукописей с близким текстом, так и при построении стеммы или дерева отношений между рукописями. Данный метод применялся в работах А. Хруби, Э. Пула, Дж. Гриффита, П. Гэлловей, Дж. Брефельд, в отечественной текстологии – в работах А. А. Алексеева, А. А. Пичхадзе, Н. Л. Гориной, Е. В. Афанасьевой. При этом различные исследователи пользуются различными коэффициентами близости или расстояния для классификации рукописей в зависимости от особенностей исследуемой традиции.

В 1985 г. проф. СПбГУ А. А. Алексеев предложил план исследования Славянского Евангелия, который включал в себя использование кластерного анализа для возможности обработки неограниченного количества рукописей контролируемой традиции. Проект был успешно реализован, и автоматизация классификации рукописей при помощи кластерного анализа значительно оптимизировала обработку источников.

В 70-е гг. XX в., немного позже выхода в свет работы Э. Колвелла и Э. Тьюна, голландский текстолог А. Деес разработал метод для построения дерева отношений между рукописями. В отличие от предшественников – приверженцев представления рукописной традиции в виде ориентированного дерева – он не ставил задачи обязательного построения ориентированного дерева и обнаружения архетипа. Он предложил начать с построения неориентированного дерева, в котором все конечные узлы – реальные рукописи, а промежуточные узлы – гипотетические рукописи. В 90-е гг. голландский ученый, математик Э. Ваттель в сотрудничестве с текстологами развил метод А. Дееса, формально его описал, сделал ряд доработок и усовершенствований, также положив в основу кластерный анализ. Метод Ваттеля,

получивший название трехуровневого метода, позволяет строить стеммы для любой рукописной традиции, как библейской, так и авторской, с любым количеством рукописей. Формально программа строит дерево не более чем для 250 рукописей, но практически идентичные рукописи можно представить как одну, что дает возможность построить дерево для гораздо большего числа источников.

Э. Ваттель предлагает следующую процедуру построения глубинной структуры. Пусть в некотором узле разночтений рукописи имеют следующие чтения: $a - 1$; $b - 2$; $c - 2$; $d - 1$; $e - 3$, $f - 3$. Тогда в данном узле получаем группы рукописей: ad , bc , ef . Э. Ваттель записывает данные о группировке рукописей для данного узла в виде строки $ad/bc/ef$ и называет такую запись формулой разночтений. Из данных шести рукописей можно составить 15 пар (ab , ac , ad , ae , af , bc , bd , be , bf , cd , ce , cf , de , df , ef), из них в 12-и случаях у рукописей пары будут разные чтения, а в 3-х – одинаковые.

Коэффициент близости между рукописями Э. Ваттель определяет следующим образом.

Пусть x и y – две рукописи, а Φ – совокупность всех формул разночтений. Для каждой формулы (узла разночтений) $\varphi \in \Phi$ вычисляем число E_φ – количество пар с одинаковым чтением в этой формуле, и число D_φ – количество пар с разными чтениями в этой формуле. Пусть W_φ – вес φ , заданный текстологом (при отсутствии ранжирования узлов разночтений по текстологической значимости вес можно считать равным 1). Тогда коэффициент близости $\sigma_\varphi(x, y)$ для рукописей x и y в данной формуле будет равен:

$$\sigma_\varphi(x, y) = W_\varphi \cdot \frac{D_\varphi + E_\varphi}{E_\varphi}$$

для x и y , имеющих одинаковое чтение в φ ,

$$\sigma_\varphi(x, y) = -W_\varphi \cdot \frac{D_\varphi + E_\varphi}{D_\varphi}$$

для x и y , имеющих разные чтения в φ ,

и 0, если для x или y данный узел нерелевантен (например, лакуна).

Коэффициент близости для пары (x, y) по всем узлам разночтений определяется как:

$$\sigma(x, y) = \sum_{\varphi \in \Phi} \sigma_\varphi(x, y)$$

Метод Ваттеля не ограничивает количество рукописей обрабатываемой традиции. Несмотря на то, что одновременно можно обрабатывать максимум 250 источников, метод позволяет работать с материалом по частям и группировать практически идентичные рукописи, так что из них может быть взят только один представитель.

Основными требованиями, предъявляемыми А. А. Алексеевым к методу кластерного анализа, стали экономичность алгоритма и минимальный субъективизм. Кластерный анализ решено было проводить по степени близости между рукописями, в связи с тем, что степень близости подразумевает более естественные группы. В 1995 году в рамках проекта исследования церковнославянского Евангелия от Иоанна научным сотрудником кафедры математической лингвистики Е. Л. Алексеевой была разработана программа кластерного анализа для церковнославянской евангельской традиции.

Для классификации был выбран наиболее простой критерий близости – отношение количества общих чтений к общему числу узлов разночтений для данной пары (из общего числа узлов исключаются те узлы, в которых хотя бы одна из рукописей имеет лауну или индивидуальное чтение): $K = F / V * 100\%$, где K – коэффициент близости, F – количество общих чтений для данной пары, V – количество узлов, по которым сравнивается эта пара.

После ряда экспериментов с критерием объединения рукописей в кластеры остановились на том, что очередной объект должен иметь наиболее высокий процент сходства с одной из рукописей уже образовавшегося кластера. От подбора коэффициента или порогового значения, определяющего окончательные границы кластеров, отказались, представляя полученные данные в виде квадратной матрицы рукописей, упорядоченной по убыванию процента сходства (процентное выражение коэффициента близости).

Недостатком метода является то, что иногда кластеры объединяются между собою раньше, чем отдельные рукописи, которые, имея не очень высокий процент сходства со своей группой, не успевают к ней присоединиться. Такие рукописи оказываются внизу таблицы. Их приходится, если это возможно, распределять по кластерам с помощью отдельной процедуры.

Для корректного сравнения методов Ваттеля и Алексева и определения их достоинств и недостатков работа обоих методов была проверена на одном и том же материале. Во **Второй главе «Классификация рукописной традиции церковнославянского текста Евангелия от Матфея»** рассмотрены результаты применения трехуровневого метода Ваттеля и метода кластерного анализа Алексева к 525 рукописям церковнославянского текста Евангелия от Матфея.

Первым шагом на пути к созданию классификации текстов рукописных источников является колляция текстов. В качестве базового текста для колляций при работе с церковнославянским Новым Заветом был взят текст Остромирова Евангелия. Каждую рукопись по отдельности сравнивали с базовым текстом, выписывая различия (эта работа была проделана рядом исследователей в рукописных хранилищах России и Европы по одной инструкции). Далее нами был составлен упорядоченный по тексту список мест в Мф. 14.14–14.34, в которых хотя бы в одной рукописи текст отличался от текста Остромирова Евангелия, таких мест оказалось 792. После исключения индивидуальных узлов был составлен список из 532 узлов разночтений, и все варианты каждого узла были пронумерованы по порядку.

После того, как по колляциям для каждого узла разночтений был получен список вариантов, данные для работы программы кластерного анализа были представлены в виде таблицы, в которой строки соответствуют сиглам рукописей, а столбцы – номерам узлов разночтений.

Рассмотрим алгоритм составления таблицы узлов разночтений для автоматической обработки рукописей на примере из Мф. 14.15. Допустим, у нас есть 8 рукописей (a–h), в которых фрагмент стиха Мф. 14.15 читается следующим образом:

- a: отъпүсти народы да шьдъше въ окръстьнаѣ градѣца³
- b: отъпүсти народъ да шьдъ въ окръстьнаѣ градѣца
- c: отъпүсти народъ да шьдъ въ окръстьнаѣ грады
- d: отъпүсти народы да въшьдъше въ окръстьнаѣ грады
- e: да въшьдъше въ ближьнаѣ вьси
- f: да отъшьдъше въ ближьнаѣ вьси
- g: отъпүсти народы да отъшьдъше въ окръжьныѣ вьси
- h: отъпүсти народы да шьдъше въ окръжьныѣ градѣца.

На этом отрезке текста мы получим следующие узлы разночтений:

- 1 узел: вариант 1: **отъпүсти народы (народъ)** – a, b, c, d, g, h
вариант 2: *омиссия* – e, f
- 2 узел: вариант 1: **народы** – a, d, g, h
вариант 2: **народъ** – b, c
вариант 0 – e, f (в этих рукописях данный фрагмент отсутствует)
- 3 узел: вариант 1: **шьдъше, шьдъ** – a, b, c, h

³ В основном орфография дается в канонизированной форме по словарю старославянского языка, изданного Чешской Академией наук в 1960–1967 гг. В данной работе в канонических формах диграф **оу** передается графемой **ү**.

- вариант 2: **вѣшьдѣше, отѣшьдѣше** – d, e, f, g
- 4 узел: вариант 1: **вѣшьдѣше** – d, e
 вариант 2: **отѣшьдѣше** – f, g
 вариант 0 – a, b, c, h (в этих рукописях приставки нет)
- 5 узел: вариант 1: **окръстьнаѣ** – a, b, c, d
 вариант 2: **ближьнаѣ** – e, f
 вариант 3: **окръжьныѣ** – g, h
- 6 узел: вариант 1: **градѣца** – a, b, h
 вариант 2: **градѣ** – c, d
 вариант 3: **вѣси** – e, f, g

Представим эти данные в виде таблицы (Таблица 1), где каждая рукопись представлена строкой, в которой для каждого узла различий проставлен номер варианта, зафиксированный в этой рукописи.

Нулем обозначаются два вида явлений – индивидуальные чтения и лакуны, которые не учитываются при сравнении двух рукописей: при подсчете числа узлов, по которым сопоставляются рукописи, принимаются во внимание только те узлы, в которых обе рукописи имеют ненулевое чтение.

Таблица 1 – Пример исходной таблицы для кластерного анализа

	Узел 1	Узел 2	Узел 3	Узел 4	Узел 5	Узел 6
Рукопись а	1	1	1	0	1	1
Рукопись b	1	2	1	0	1	1
Рукопись с	1	2	1	0	1	2
Рукопись d	1	1	2	1	1	2
Рукопись e	2	0	2	1	2	3
Рукопись f	2	0	2	2	2	3
Рукопись g	1	1	2	2	3	3
Рукопись h	1	1	1	0	3	1

Материал, подготовленный по колляциям рукописей церковнославянского Нового Завета, Э. Ваттель преобразовал для программы трехуровневого анализа в виде формул различий.

Классификация 525 рукописей Евангелия от Матфея XI–XVI вв. по отрывку 14.14 – 14.34 проводилась при помощи описанных методов Алексева и Ваттеля. Коэффициент близости между всеми рукописями оказался достаточно высок, не ниже 80%.

Классификация по всем узлам позволила выделить 13 кластеров, из которых в одном объединилось 308 рукописей, в одном – 48 рукописей, в двух – по 4 рукописи, в одном – 3 рукописи, а во всех остальных – по 2 рукописи. После этого список узлов

был пересмотрен и были удалены те низкорелевантные узлы, в которых случайным образом объединялись 2–3 рукописи, а также узлы, противоречащие полученным кластерам. Важно отметить, что каждое характерное чтение разделяется не абсолютно всеми рукописями редакции, а только большинством, причем в каждом узле кластера в равной степени могут быть представлены два или более чтений. Таким образом, осталось 387 значимых узлов разночтений, и на их базе расчеты были проведены во второй раз, что позволило выявить более четкие границы между кластерами, тогда как состав и количество кластеров существенно не изменились.

Пять полученных в итоге классификации четко выделенных кластеров соответствуют известным редакциям церковнославянского Евангелия, описанных в работах Л. П. Жуковской, Г. А. Воскресенского, А. А. Алексева, критическом издании церковнославянского Евангелия от Иоанна. После завершения процедуры классификации ряд рукописей не объединились в кластеры, и для каждой из этих рукописей мы рассмотрели процент сходства с основными кластерами, соответствующими редакциям церковнославянского Евангелия. В ходе этой процедуры численность кластеров стала следующей: Поздний текст – 363 рукописи, Преславский текст – 96 рукописей, Древний текст – 41 рукопись, Новый литургический тетр – 4 рукописи, Чудовский Новый Завет – 3 рукописи.

Самый многочисленный кластер, включающий после удаления низкорелевантных узлов 363 рукописи, объединяет источники с Поздним текстом. В основном они соответствуют Афонским редакциям А и В, выявленным в ходе исследования Евангелия от Иоанна. К ним, однако, примыкают рукописи, для которых в издании Евангелия от Иоанна редакция не указана или же указана другая редакция. 126 рукописей образуют кластер с самым высоким коэффициентом близости в 95.5% и формируют ядро Позднего текста, к которому тесно примыкает еще 181 рукопись также с высоким коэффициентом близости (93–98%), и на периферии которого находятся 52 рукописи, имеющие коэффициент близости с Поздним текстом в 90–95%. Коэффициент близости с другими редакциями у каждой рукописи периферии на два и более процентов ниже.

В целом результаты в таблице кластеризации по методу Алексева и на стемме Ваттеля получились одинаковые. Отличия наблюдались в составе некоторых подгрупп редакции.

Второй крупный кластер соответствует ядру Преславского текста. Коэффициент близости между рукописями ядра – 88.5%, при этом у некоторых рукописей сходство достигает 98%. Этот текст имеет также большую периферийную часть. Всего выделено 96 рукописей данной редакции. Коэффициент близости внутри всей группы, включая периферийную часть, составляет 85%. У данной редакции в наибольшей степени выражен разрыв значений коэффициента близости

внутри редакции и у рукописей редакции с рукописями других редакций: он составляет 8–10%. Рукописи, вошедшие в данную группу, совпадают с составом текстологической семьи Мстиславова евангелия, подробно описанной в работе С. Ю. Темчина. Стемма, как и таблица кластеризации, иллюстрирует значительное единение внутри кластера Преславского текста.

Кластер Нового литургического тетра образован четырьмя рукописями РНБ Q.п.I.44, тетр, XIV в., РНБ Гильф. 18, тетр, XIII–XIV вв., РГБ ф. 178, 891, тетр, XIV в., Пантелеимонов монастырь, Слав. 7, тетр, XIII в. Коэффициент близости внутри кластера – 97.5%, с другими кластерами он составляет 80–95%. Этот кластер всегда находится рядом с кластером Древнего текста, и коэффициент близости с Древним текстом у него наибольший.

Ядро Древнего текста представлено кластером, включающим следующие четыре рукописи: РНБ Глаг.1, XI в. (Зографское евангелие), РГБ ф. 37 Григ. 6, тетр, XI в. (Мариинское евангелие), РГАДА ф. 381, 1, XII в., (Типографское евангелие), ГИМ Син. 404, , 1144 г. (Галицкое евангелие).

Коэффициент близости между перечисленными рукописями достигает 95–97%, с другими кластерами – 88–94%. У примыкающих к ядру кластера рукописей ГИМ Увар. 480 и ГИМ Увар. 93, Остромирова евангелия, РГАДА ф. 188, 816, РНБ Q.I.924, Библиотека Ватикана Cod. Slav. 3 (Ассеманиево ев.), ГИМ Увар. 293, НМ 1536 (Мирославово ев.), Дублин Chester Beatty 23, РНБ Пог. 11 коэффициент близости с ядром составляет 94–96%, что на 2–3% выше, чем коэффициент близости с другими редакциями. Кроме рукописи Cod. Slav. 3 (Ассеманиево ев.) все перечисленные рукописи образуют две ветви на стемме.

Кластер Чудовского Нового Завета представлен следующими тремя рукописями: РНБ Пог. 21, XIV в., тетр, РГБ ф. 304.III, 6, ок. 1399 г., тетр, (Никоновское евангелие), Чудовский Новый Завет, тетр, XIV. Коэффициент близости между ними – 94.5%, коэффициент близости другими редакциями составляет 81–91%. Эти три рукописи объединяются в кластер как в таблице, так и на стемме.

В целом оба метода позволили получить группы рукописей, соответствующие известным в науке редакциям текста церковнославянского Нового Завета. Это единственные два метода, с помощью которых можно обработать любое количество источников контаминированной традиции. При этом по трудоемкости метод Ваттеля требует больших затрат памяти и времени для вычислений, так как в ходе оптимизации изначальной стеммы необходим перебор большого количества промежуточных вариантов деревьев.

В составе ядра каждой редакции при классификации по обоим методам не наблюдается расхождений, однако на периферии рукописи могут располагаться по-разному, сближаясь с рукописями различных редакций. Также, если метод Алексева

позволяет визуально оценить близость каждой рукописи периферии к той или иной редакции путем сравнения процента сходства между данной рукописью и рукописями ядра различных редакций, то метод Ваттеля не предоставляет такой возможности, так как количественная оценка близости рукописей не визуализирована.

На основании проведенного анализа работы обоих методов на материале церковнославянского Евангелия можно сделать выводы о том, что метод Алексеева является наиболее оптимальным для классификации церковнославянской новозаветной традиции.

В Третьей главе **«Возможности оптимизации автоматической классификации рукописей на основании исследования различных типов узлов разночтений»** сформулированы принципы формирования узлов разночтения с целью обеспечения достоверности полученных результатов. На материале церковнославянского перевода Евангелия от Матфея рассмотрены такие понятия, как длина, вес, тип узла разночтений, после чего проводится проверка влияния разных типов узлов разночтений на формирование групп рукописей при применении методов Алексеева и Ваттеля.

Узел разночтения – основание для любой классификации древних рукописей, для любого метода исследования. А. А. Алексеев определяет узел разночтения как такое место в тексте, где «показания текстов рукописей расходятся между собой». От того, как выделяются узлы разночтения и ранжируются ли они по значимости, зависят результаты компьютерной классификации рукописей.

В ходе анализа разночтений, представленных в исследуемом текстовом фрагменте, сформулировано следующее определение длины узла разночтений: *длина узла разночтений соответствует минимальному элементу текста, изменение которого не повлекло за собой изменение других элементов*. В данном определении под минимальным элементом текста, понимается один лексический, грамматический или текстовый признак.

Понятие «веса» узла разночтения или степени его текстологической значимости затрагивается в ряде работ по текстологии, в особенности в работах, связанных с компьютерной классификацией рукописей. Очевидно, что не все разночтения имеют одинаковый текстологический вес, но, как отмечает А. А. Алексеев, без специального исследования текстологической значимости различных содержательных типов узлов разночтений какой-либо вес узлам приписывать нельзя. Поэтому при классификации рукописей Евангелия от Матфея было решено отказаться от приписывания веса различным лингвистическим категориям, различным типам узлов.

В работе рассмотрены следующие традиционные виды узлов разночтений: лексическое варьирование: (корни (14.15 *окръстьнѣнъ* – *ближьнѣнъ*, 14.24 *погрѣжлѣнъ*

– **вѣлаѡ – мѡта сѡ**), варьирование словообразовательных элементов (приставки: 14.14 **вѣшьдѣ – ишьдѣ – пришьдѣ – съшьдѣ**); грамматические замены: варьирование частей речи (существительное – местоимение: 14.29 **онѣ – исѣсъ**, существительное – прилагательное: 14.20 **избѣтъкы – избѣтъчныхѣ**), личные / неличные формы глагола (причастие – аорист: 14.14 **видѣвъ – видѣ**, аорист – инфинитив: 14.29 **приде – приити**), варьирование времени глагола (аорист – перфект: 14.15 **минѡ – минѡлѣ кстѣ**, аорист – настоящее время: 14.17 **глаголаша – глаголюжтѣ**), варьирование неличных форм глагола (супин – инфинитив: 14.23 **помолитѣ сѡ – помолити сѡ**), варьирование времени причастия (14.21 **гдѣцихѣ – гдѣшихѣ**), видовые пары глаголов (14.34 **прѣгдѡхѡвѣше – гдѡхѡвѣше**), старые / новые формы (аориста: 14.20 **гашѡ – гдѡшѡ**, причастия: 14.19 **прѣломѡлѣ – прѣломивѣ**, винительного падежа 14.26 **видѣвъѣше и – видѣвъѣше его**), варьирование числа (число существительного: 14.30 **вѣтрѣ крѣпкѣ – вѣтры крѣпкы**, число местоимения: 14.19 **благослови и – благослови ѡ**), варьирование падежей (причастия: 14.21 **гдѣцихѣ – гдѣцимѣ**, существительного: 14.24 **вѣлнами – отѣ вѣлнѣ**), полные / краткие прилагательные и причастия (14.32 **вѣлѣзѣшема – вѣлѣзѣше**); варьирование строевых элементов (омиссии и чередования союзов и частиц – в основном **и и же**, изредка – **нѣ, ѹже, да, ѹво, гакѡ**); текстовые изменения: оmissии / добавления (14.19 **вѣзьрѣвъ на небо благослови** – оmissия, 14.33 **корабли – корабли пришьдѣше**), перестановки (14.28 **прити къ тебѣ по водамѣ – къ тебѣ прити по водамѣ – прити по водамѣ къ тебѣ**, 14.33 **сынѣ божи – божи сынѣ**).

Среди выявленных 532 неиндивидуальных узлов разночтений насчитывается 241 узел оmissий и добавлений, 148 оmissий и добавлений без учета строевых элементов, 106 вариаций строевых элементов, 158 грамматических узлов, 66 лексических замен и 54 перестановки. С целью определения текстологической релевантности традиционных типов узлов разночтений с помощью методов Алексеева и Ваттеля была проведена классификация всех рукописей с учетом каждого из типов узлов по отдельности.

Лексические замены. В таблице внутри кластеров многие рукописи имеют между собой сходство 100%. Это касается как ядра Древнего текста, так и Позднего и Преславского текста. Перепад процента сходства между кластером с ядром Древнего текста и Поздним текстом составляет 2%. Это свидетельствует о том, что лексически текст редакций очень близок. Все же кластеры Древнего и Позднего текста не перемешиваются, а размещаются друг за другом. На стемме отдельно располагается только ядро Позднего текстов. Другие представители Позднего и Древнего текста перемешиваются.

Грамматические замены. В таблице все рукописи Позднего текста располагаются вместе, процент сходства между рукописями в пределах этих редакций

очень высок: 95-100%. Они образуют в основном однородный кластер, подгруппы практически не выделяются. На стемме также в первую очередь объединяются рукописи с Поздним текстом. Рукописи с Древним текстом образуют обширный кластер в таблице, в который входят и рукописи с Новым литургическим тетром. Близость рукописей кластера между собою составляет 90-99%. Из представителей Древнего текста только три рукописи находятся на периферии таблицы. На стемме также выделяется группа Древнего текста. На периферии размещены четыре рукописи, две из которых совпадают с рукописями на периферии таблицы. Все представители Преславского текста объединяются в одну группу как на стемме, так и в таблице. Из рукописей с Чудовским текстом только две из трех образуют кластер в таблице. На стемме все три рукописи образуют группу, но объединяются достаточно поздно, что свидетельствует о невысокой грамматической однородности внутри группы.

Строчные элементы. В таблице процент сходства между рукописями *всех* редакций достаточно высок (86-100%). Это затрудняет выделение кластеров. На стемме Древний текст и Новый литургический тетр объединяются с Поздним текстом.

Омиссии (включая омиссии строчных элементов). В таблице четко выделяется кластер с Поздним текстом, сходство между рукописями 94-100%. На стемме все представители Позднего текста объединяются в две ветви. Древний текст в таблице образует один небольшой кластер из восьми рукописей со степенью сходства 95-99%. Остальные представители Древнего текста смешиваются с представителями Преславского текста, который образует один большой кластер. Три рукописи с Новым литургическим тетром выделяются в отдельную группу как в таблице, так и на стемме.

Омиссии (без учета омиссий строчных элементов). Поскольку пропуск строчного элемента и пропуск значимого слова могут иметь различную текстологическую значимость, решено было провести классификацию по омиссиям значимых слов отдельно, без учета омиссий строчных элементов.

В таблице с высоким процентом сходства (96-100%) объединяются в один кластер рукописи с Поздним текстом. Вслед за представителями Позднего текста располагаются представители ядра групп Преславского и Древнего текстов. Далее все остальные рукописи этих редакций, Нового литургического тетра и Чудовского текста размещены в таблице вперемешку. На стемме не выделяется в группу ни одна редакция. По данным таблицы получается, что омиссии и добавления релевантны только для Позднего текста и для нескольких рукописей ядра Древней и Преславской редакций. Из стеммы видно, что этот тип узлов разночтений не является релевантным ни для одной редакции.

Сравнивая результаты обработки материала на выборке оmissий / добавлений любых и только значимых частей речи, можно сказать, что оmissии / добавления строевых элементов важны для атрибуции рукописей к различным редакциям, и что оmissии / добавления только значимых слов обладают меньшей текстологической релевантностью, чем оmissии / добавления любых слов.

Перестановки. Для узлов этого типа по несколько представителей ядра Преславского и Древнего текстов показывают в основном сходство 100%. Большая группа рукописей, относящихся к Преславскому и Позднему тексту, практически идентичны (сходство 98-100%). Единственная редакция, все представители которой образуют один кластер с видимым отличием от соседних рукописей (12%), – это Чудовский текст (сходство рукописей между собою – 94 – 96%).

На стемме большинство рукописей с Древним текстом образуют две ветви, рукописи с Чудовским текстом объединяются, представители Преславского и Позднего текстов на стемме смешиваются. Таким образом, видно, что перестановки текстологически релевантны только для Чудовской редакции и нескольких представителей ядра Древнего текста. Следует отметить, что такой результат связан не только с меньшей текстологической значимостью перестановок, но и с их небольшим количеством.

После того как было получено разбиение на группы с учетом 387 релевантных узлов разночтений, из полного списка узлов разночтений написанная нами программа выбирала такие узлы, в которых больше 75% представителей хотя бы одной текстовой группы имеют одно и то же чтение, отличное от чтения большинства представителей другой текстовой группы. Таким образом, были получены текстологические приметы всех текстовых групп (далее – приметы). Для определения примет из каждого из пяти кластеров были взяты рукописи ядра: 316 рукописей Позднего текста, 11 – Преславского текста, 4 – Древнего текста, 4 рукописи с текстом Нового литургического тетра, 3 рукописи с текстом Чудовского Нового Завета.

В рассматриваемом отрывке Мф. 14.14 – 14.34 было обнаружено 64 приметы. После отбора примет была проведена повторная классификация только с учетом данных примет. Результаты разбиения на группы не противоречат изначальному разбиению. В кластер Древнего текста вошли некоторые рукописи, которые до этого при кластерном анализе располагались на периферии.

Доля грамматических и лексических замен среди примет увеличилась, также незначительно возрастает значимость перестановок, а доля строевых элементов и оmissий уменьшилась. Так, грамматические замены составляют 29.7% от общего числа узлов и 46.8% от числа примет. Лексические замены составляют 12.4% от общего числа узлов и 20% от общего числа примет, а перестановки - 10.1% от общего

числа узлов и 15.6% от числа примет. Оmissии составляют 27.8% всех узлов, а среди примет их доля 7.8%; количество строевых элементов среди всех узлов 19.9%, а среди примет 9.3%.

При переписывании текстов писцы могли сверяться с греческим оригиналом, следовательно, разночтения в греческом тексте могут объяснить некоторые разночтения в славянском тексте.

Из 143 узлов, встречающихся в критическом аппарате изданий греческого Нового Завета, в церковнославянском Новом Завете встречаются 85 узлов разночтений, из них перестановок – 14, лексических замен – 7, строевых элементов – 19, оmissий – 29, грамматических замен - 16. Среди разночтений, играющих роль текстологических примет славянского текста, в греческом Новом Завете имеют параллели 5 оmissий, 5 строевых элементов, 3 лексических варианта, 4 грамматических, 7 перестановок.

Пять из шести строевых элементов, все оmissии и семь из десяти примет перестановок в славянском тексте имеют параллели в греческом тексте. Почти все греческие грамматические чередования в Мф. 14.14 – 14.34 находят отражение в узлах разночтений славянского Нового Завета, однако в славянском Новом Завете грамматических чередований гораздо больше, только четыре из тридцати грамматических примет связаны с греческим текстом. Лексические замены, кроме трех случаев, не имеют связи с греческими чередованиями.

В **Заключении** диссертационного исследования сформулированы его основные результаты.

В работе проведен анализ основных современных методов компьютерной классификации рукописей, применимых к большому количеству источников. На основании применения методов Ваттеля и Алексева к 525 церковнославянским рукописям исследуемого пассажа Евангелия от Матфея предложена оптимальная процедура исследования, позволяющая с наименьшими ограничениями и наименьшими затратами усилий решать задачу определения группировки рукописных источников.

Результаты исследования в дальнейшем могут применяться при работе с библейской и другими объемными рукописными традициями на разных языках.

Основные положения диссертации отражены в следующих публикациях.

1. Николаенко⁴, Д. М. Автоматическая классификация текстов древних рукописей / Д. М. Николаенко // Материалы XXIX межвуз. науч.-метод. конф. преподавателей и аспирантов. Вып. 7. : секция прикладной и математической лингвистики, 13–18

⁴ Фамилия Николаенко изменена на фамилию Миронова в связи со вступлением в брак 24.07.1999 г.

- марта 2000 г. / [отв. ред. А. С. Герд]. – СПб. : Изд.-во С.–Петерб. ун-та, 2000. – С. 39–42 (0,15 п. л.).
2. **Nikolaenko, D. Old Church Slavonic Versions of the Gospels. Computer-Aided Classification and the Choice of Variants / Dina Nikolaenko // Bible and Computer : the Stellenbosch AIBI-6 Conference : Proceedings of the Association Internationale Bible et Informatique “From Alpha to Byte”, University of Stellenbosch, 17–21 July, 2000 / ed. by J. Cook. – Leiden ; Boston : Brill, 2002. – P. 475–493 (0,75 п. л.). – ISBN 90 04 12495 0. (WoS CC)**
 3. Mironova, D. Cluster Analysis and the Three Level Method in the Study of the Gospels in Slavonic / Dina Mironova // Studies in Stemmatology II / ed. by P. van Reenen, A. den Hollander, M. van Mulken. – Amsterdam ; Philadelphia : John Benjamins Publishing Company, 2004. – P. 241–267 (1,3 п. л.). – ISBN 90 272 3222 9.
 4. Миронова, Д. М. Классификация рукописей Евангелия от Матфея / Д. М. Миронова // Евангелие от Матфея в славянской традиции / [изд. подгот.: А. А. Алексеев, И. В. Азарова, Е. Л. Алексеева и др.] ; С.–Петерб. гос. ун-т, филол. фак. ; Синод. библиот. Моск. Патриарх. ; Рос. библиот. общ., С.–Петерб. отд. – СПб., 2005. – С. 163–168 (0,25 п. л.). – ISBN 5-8015-0196-7.
 5. Миронова, Д. М. Текстологическая значимость узлов разночтений (на материале 525 списков славянского Евангелия от Матфея XI–XVI вв.) / Д. М. Миронова // Материалы XXXIX междунар. филол. конф. : секция прикладной и математической лингвистики, 15–20 марта 2010 г. / [отв. ред. Т. Г. Скребцова]. – СПб. : Филол. фак. СПбГУ, 2010. – С. 38–44 (0,25 п. л.). – ISBN 978-5-9651-0489-5.
 6. Миронова, Д. М. Текстологически нерелевантные узлы разночтений / Д. М. Миронова // Материалы XL междунар. филол. конф. : секция прикладной и математической лингвистики, 14–19 марта 2011 г. / [отв. ред. Т. Г. Скребцова]. – СПб. : Филол. фак. СПбГУ, 2011. – С. 45–48 (0,2 п. л.). – ISBN 978-5-9651-0567-0.
 7. **Миронова, Д. М. Оценка текстологической значимости узлов разночтений (на материале 525 списков славянского Евангелия от Матфея XI–XVI вв.) / Д. М. Миронова // Структурная и прикладная лингвистика. – Вып. 9. – 2012. – С. 261–275 (0,7 п. л.). – ISSN 0202-2400.**
 8. Миронова, Д. М. Кластеризация рукописей на базе совпадения разночтений как основа публикации славянской традиции / И. В. Азарова, Е. Л. Алексеева, Д. М. Миронова // Материалы XLIII междунар. филол. конф. : секция прикладной и математической лингвистики, 11–15 марта 2014 г. / [отв. ред. М. В. Хохлова]. –

- СПб. : Филол. фак. СПбГУ, 2014. – С. 10–22 (0,7 п. л.; доля авт. 0,25 п. л.). – ISBN 978-5-8465-1453-9⁵.
9. Миронова, Д. М. Применение кластерного анализа в текстологии / Д. М. Миронова // Структурная и прикладная лингвистика. – Вып. 11. – 2015. – С. 155–160 / (0,25 п. л.). – ISSN 0202-2400.
10. Миронова, Д. М. Компьютерная текстология / Е. Л. Алексеева, Д. М. Миронова // Прикладная и компьютерная лингвистика : коллект. моногр. / [под ред. И. С. Николаева, О. В. Митрениной, Т. М. Ландо]. – М. : URSS, 2016. – С. 259–272 / (0,75 п. л.; доля авт. 0,35 п. л.). – ISBN 978-5-9710-3472-8⁶.
11. **Миронова, Д. М. Понятие длины узла разночтений при классификации рукописей / Д. М. Миронова // Филологические науки. Вопросы теории и практики. – 2017. – № 3 (69). Ч. 1. – С. 117–120 (0,25 п. л.). – ISSN 1997–2911.**

⁵ Работа выполнена в соавторстве. Авторство разделено, согласие соавторов на использование результатов и материалов публикации в тексте диссертации получено.

⁶ Работа выполнена в соавторстве. Авторство разделено, согласие соавтора на использование результатов и материалов публикации в тексте диссертации получено.