

ОТЗЫВ

официального оппонента о диссертации Мироновой Дины Марковны на тему: «Автоматизированная классификация древних рукописей (на материале 525 списков славянского Евангелия от Матфея XI-XVI вв.)», представленной на соискание ученой степени кандидата филологических наук по специальности 10.02.21 – Прикладная и математическая лингвистика

В диссертации Д. М. Мироновой апробируются статистические методы измерения степени близости рукописей с применением математического аппарата кластерного анализа. Разработка оптимальных алгоритмов для группировки рукописей – актуальная задача текстологии, занимающейся исследованиями текстов, которые сохранились в большом количестве списков. При рукописном тиражировании текста в течение долгого времени на обширных территориях возникают контаминированные списки, между которыми трудно, а иногда даже невозможно установить генеалогические отношения. Это объясняется тем, что в допечатную эпоху единственным способом обеспечить правильность текста при переписке было обращение к нескольким источникам: дефекты одного антиграфа устранялись благодаря сверке по другим. Возникла ситуация, когда у каждой новой рукописи мог быть не один антиграф, а несколько, причем относящихся к разным текстуальным группам. В результате отразить отношения между списками при помощи традиционной генеалогической стеммы не удастся. В качестве альтернативы генеалогическому методу был предложен ряд статистических методов. В работе Д. М. Мироновой сопоставляются статистические методы Э. Ваттеля и А. А. Алексева и выявляются их достоинства и недостатки. Такое сопоставление проводится впервые.

В первой главе диссертации, с привлечением большой отечественной и зарубежной литературы, дается обзор методов компьютерной текстологии, используемых в отечественной и зарубежной практике, – в частности, метод американских ученых Э. Колвелла и Э. Тьюна, при котором все рукописи сравниваются между собой попарно и процент сходжений заносится в таблицу по степени убывания, метод голландского ученого А. Дееса, предложившего построение неориентированного дерева, и его продолжателя Э. Ваттеля, разработавшего программу для построения глубинной структуры и процедуру оптимизации получаемой стеммы в три этапа, на последнем из которых стемма корректируется текстологом. Особое внимание в обзоре уделяется кластерному анализу, который в 1980-х гг. стал применяться А. А. Алексевым к группировке библейских славянских списков. Детально рассматриваются ключевые для кластерных методов вопросы о критерии присоединения объектов к кластерам и об определении границ кластеров. В варианте, используемом А. А. Алексевым, новая рукопись присоединяется к имеющемуся кластеру, если имеет наиболее высокий процент сходства с одной из рукописей уже образовавшегося кластера, а коэффициент или пороговое значение, определяющее окончательные границы кластеров, не используется. Полученные данные, как у Э. Колвелла и Э. Тьюна, представляются в виде квадратной матрицы, упорядоченной по убыванию процента сходства между рукописями. Метод прост в применении и наглядно представляет результаты классификации. Впрочем, Д. М. Миронова указывает и на недостаток описанного метода: иногда кластеры объединяются между собой раньше, чем отдельные рукописи, которые имеют не очень высокий процент сходства со своей группой и не успевают к ней присоединиться. Такие

рукописи оказываются внизу таблицы, и их приходится, если это возможно, распределять по кластерам с помощью отдельной процедуры.

Во второй главе описывается процесс классификации рукописей славянского Евангелия на материале контрольного отрывка, в котором были выделены 532 узла разночтений. Все узлы приведены в Приложении к диссертации. Сравнение было проведено по 525 рукописям, материал подвергался анализу параллельно по методу Э. Ваттеля и по методу А. А. Алексева. После первоначальной кластеризации были удалены низкорелевантные узлы, в которых случайным образом объединились 2-3 рукописи, после чего осталось 387 значимых узлов, на базе которых был проведен новый расчет. Он показал наличие пяти кластеров, представляющих хорошо известные разновидности текста славянского Евангелия: Поздний текст (363 списка), Преславский текст (96 списков), Древний текст (41 список), Новый литургический текст (4 списка) и редакция Чудовского Нового Завета (3 списка). В диссертации приведены характерные разночтения всех пяти кластеров. В самом большом кластере было выделено ядро Позднего текста (126 списков), к которому примыкают рукописи, принадлежащие к другим редакциям или принадлежность которых к той или иной редакции неизвестна. В кластере Преславского текста была зафиксирована большая степень единства входящих в него рукописей. Внутри кластеров выделены более мелкие подгруппы. Ядро кластера Древнего текста составили четыре древнейших тетра: Мариинское, Зографское, Типографское и Галицкое евангелия. В целом результаты, полученные по методу Э. Ваттеля и по методу А. А. Алексева, совпали, расхождения касались некоторых групп внутри редакций. Метод Э. Ваттеля оказался более трудоемким, а полученные с его помощью результаты – менее наглядными.

В третьей главе оценивается вес разночтений разного типа с точки зрения возможностей оптимизации автоматической классификации рукописей. Здесь подробно обсуждается, что именно нужно считать узлом разночтений и показывается, как меняется вес разночтения в зависимости от того, по какому принципу выделен узел. Далее проводятся подсчеты с последовательным исключением разных типов узлов (лексические и грамматические разночтения, пропуски и добавления, перестановки, наличие и отсутствие союза *и* и частицы *же*) и доказываемся, что исключение любого из перечисленных типов узлов ведет к искажению картины и только если не учитывать наличие и отсутствие союза *и* и частицы *же*, группировка остается той же, что и при учете всех узлов. Другими словами, наличие и отсутствие союза *и* и частицы *же* при проведении кластерного анализа можно не подсчитывать. Кроме того, при помощи специальной программы были выбраны узлы, характерные для каждого выделенного кластера (текстологические приметы), и проведена повторная классификация на основании только этих 64 узлов (они приведены в Приложении к диссертации). Повторная классификация в целом совпала с первоначальной, но оказалась более рельефной: граница между кластерами стала более отчетливой, а рукописи с периферии подтянулись к ядру кластера, то есть кластер стал более компактным.

Таким образом, в диссертации осуществлено разностороннее тестирование двух статистических методов с применением кластеризации. Доказано, что оба метода обеспечивают адекватную классификацию рукописей, принадлежащих любой текстуальной традиции, и могут применяться как для обследования рукописного материала в первом приближении, так и для иллюстрации взаимоотношений между списками. В диссертации осуществлена группировка рукописей внутри Древнего и

Преславского типа, охватывающая, в том числе, рукописи, текстуальный статус которых до сих пор не был определен. Эти новые данные представляют интерес для текстологии славянского Евангелия.

С другой стороны, диссертация демонстрирует пределы возможностей статистического анализа при изучении текстуальной традиции славянского Евангелия. В силу высокой степени совпадения между списками кластеры оказываются слишком аморфными. Так, внутри Позднего текста не выделились редакции А и В, зато в кластер Позднего текста попали рукописи других редакций. В ядро кластера Древнего текста вошли Мариинское, Зографское, Типографское и Галицкое евангелие, хотя между Мариинским и Зографским лежит сверка по греческому оригиналу, а Галицкое евангелие содержит особую восточноболгарскую редакцию текста. Высокую степень единства показал кластер Преславского текста, хотя в нем при филологическом анализе отчетливо выделяются две текстуальные ветви. Эти результаты тоже важно зафиксировать, чтобы понимать, что статистический анализ сам по себе, во всяком случае применительно к тексту славянского Евангелия, дает адекватную картину распределения списков главным образом по очень крупным группам, представляющим собой типы текста.

Ценным вкладом являются предложения Д. М. Мироновой по оптимизации статистического анализа славянского текста Евангелия. Первое сводится к исключению из подсчетов узлов, регистрирующих наличие и отсутствие союза *и* и частицы *же*, которые составляют изрядную долю от всех разночтений. В то же время показано, что исключение какого-либо другого типа разночтений, помимо наличия или отсутствия союза *и* и частицы *же*, искажает общий результат. Впрочем, искаженный результат, как показано в диссертации, может быть использован в эвристических целях: он сигнализирует, например, о том, что для определенного типа текста более показательны лексические инновации, а для другого – грамматические замены. Второй способ оптимизации – автоматизированное выделение текстологических примет кластеров с повторным пересчетом, для которого Д. М. Миронова разработала специальную программу. Поскольку этот способ прошел испытание очень успешно, его, очевидно, следует рассматривать как новую эффективную модификацию кластерного метода и принять на вооружение.

Диссертация в целом написана очень ясно и доказательно. Можно высказать лишь несколько мелких замечаний.

К сожалению, не описаны принципы определения низкорелевантных узлов и не приведены примеры таких узлов, удаленных на втором этапе кластеризации, хотя к их числу принадлежит пятая часть от общего количества всех выделенных узлов. Может быть, стоило бы пометить их в Приложении специальной маркировкой.

Процедура выделения узлов описана в диссертации достаточно четко, но в единичных случаях можно заметить отступления от главного принципа, который сводится к тому, что в рамках одного узла разночтения должны быть противопоставлены только по одному признаку. Это обеспечивает равную близость / удаленность вариантов внутри одного узла. Между тем в разделе 2.1.2 в узел № 6 к стиху Мф. 14.15 включены три варианта: 1 – градьц□, 2 – грады, 3 – вьси. Очевидно, что первые два варианта ближе друг к другу, чем к третьему варианту, поскольку содержат один корень. Следовательно, правильнее было бы разбить вариант № 6 на два, выделив в первом два варианта (1 – градьц□, грады, 2 – вьси), а во втором – два варианта (1 – градьц□, 2 – грады) с исключением варианта «вьси».

Принятое в диссертации определение понятия длины узла разночтений как минимального элемента текста, изменение которого не повлекло за собою изменение других элементов, по сути, безусловно, правильное, может быть, нуждается в уточнении: при анализе переводных текстов длина узлов иногда определяется оригиналом.

Это частности, разумеется, не влияют на высокую оценку обсуждаемой диссертации. В ней представлен результат, интересный и важный с точки зрения как разработки методов автоматизированного анализа текста, так и решения трудных вопросов текстологии славянского Евангелия.

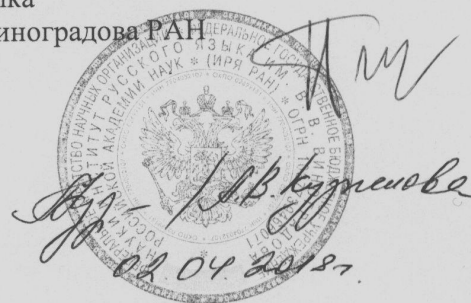
Основное содержание диссертации представлено в 11 публикациях в научных журналах и сборниках, 3 статьи опубликованы в рецензируемых изданиях их перечня ВАК РФ (в том числе одна статья – в издании, включенном в индекс цитирования Web of Science), положения диссертации неоднократно излагались в докладах на научных конференциях в России и за рубежом.

Диссертация написана автором самостоятельно, автореферат диссертации и опубликованные статьи полностью отражают основное содержание работы, научные результаты и положения основательно аргументированы. В диссертационном исследовании Мироновой Дины Марковны «Автоматизированная классификация древних рукописей (на материале 525 списков славянского Евангелия от Матфея XI-XVI вв.)» содержится решение задачи, имеющей существенное значение для прикладной и математической лингвистики, оно соответствует всем требованиям, изложенным в пунктах 9, 10, 11, 13, 14 Положения о порядке присуждения ученых степеней, утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842, а его автор Миронова Дина Марковна заслуживает присуждения искомой ученой степени кандидата филологических наук по специальности 10.02.21 – Прикладная и математическая лингвистика.

2 апреля 2018 года

Пичхадзе Анна Абрамовна,
доктор филологических наук, старший научный сотрудник,
ведущий научный сотрудник отдела лингвистического источниковедения
и истории русского литературного языка
Института русского языка им. В. В. Виноградова РАН

*Подпись руки заверено
Главной специалисткой по кадрам*



Адрес организации –
119019, Москва, Волхонка 18/2
Институт русского языка им. В. В. Виноградова РАН
(495) 695-26-60 ruslang@ruslang.ru