

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

имени М.В. ЛОМОНОСОВА

На правах рукописи

ШПИРКО СЕРГЕЙ ВАЛЕРЬЕВИЧ

**МАТЕМАТИЧЕСКИЕ МЕТОДЫ КЛАССИФИКАЦИИ СПИСКОВ
ДРЕВНЕРУССКИХ ТЕКСТОВ**

07.00.09 – Историография, источниковедение и методы исторического
исследования

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата исторических наук

Москва – 2019

Работа выполнена на кафедре исторической информатики исторического факультета ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова»

- Научный руководитель** – Бородкин Леонид Иосифович, член-корреспондент РАН, доктор исторических наук, профессор
- Официальные оппоненты**
- Гимон Тимофей Валентинович, доктор исторических наук, ФГБУН «Институт всеобщей истории РАН», заведующий отделом истории Византии и Восточной Европы, ведущий научный сотрудник
 - Горский Антон Анатольевич, доктор исторических наук, профессор, ФГБУН «Институт российской истории РАН», руководитель Центра источниковедения истории России, главный научный сотрудник
 - Усачев Андрей Сергеевич, доктор исторических наук, доцент, ФГБОУ ВО «Российский Государственный Гуманитарный Университет», профессор кафедры теории и истории исторической науки факультета истории, политологии и права Историко-архивного института

Защита диссертации состоится «3» июня 2019 г. в 14 часов на заседании диссертационного совета МГУ.07.01 Московского государственного университета имени М.В. Ломоносова по адресу: 119192, г. Москва, Ломоносовский проспект, д. 27, корп. 4, исторический факультет МГУ, ауд. А-419.

E-mail: ot-dissovet@hist.msu.ru

С диссертацией можно ознакомиться в отделе диссертаций научной библиотеки МГУ имени М.В. Ломоносова (Ломоносовский просп., д. 27) и на сайте ИАС «ИСТИНА»: <http://istina.msu.ru/dissertations/188297828>

Автореферат разослан «__» _____ 20__ г.

Ученый секретарь
диссертационного совета,
кандидат исторических наук

Н.Г. Абрамова

Общая характеристика работы

Актуальность темы исследования. Прогресс в изучении средневековых текстов во многом определяется совершенствованием методов их исследований. Круг вопросов, поставленных и решаемых в рамках настоящего междисциплинарного исследования, лежит в русле фундаментальной идеи привлечения новых математических идей, методов и алгоритмов к решению исторических проблем. В этом смысле идея применения теории нечетких множеств к историко-текстологическим исследованиям является весьма актуальной. Адаптация данной теории применительно к текстологическим задачам дает в руки исследователя достаточно универсальный и адекватный инструмент, помогающий проверять и выдвигать те или иные текстологические гипотезы.

Объект исследования в настоящей работе выбран с таким расчетом, чтобы представлять как древнеславянские переводные тексты («Закон Судный людем» краткой редакции), так и оригинальные древнерусские («Предсловоие покаянию»). Что касается «Закона Судного людем» (в дальнейшем ЗСЛ), то можно без преувеличения сказать, что традиция его изучения насчитывает уже не одно столетие. Для XIX века в этой связи достаточно назвать такие известные имена как Г.А. Розенкампф¹, А.С. Павлов², Н.С. Суворов³. В XX веке изучение ЗСЛ было продолжено в трудах А.И. Соболевского⁴,

¹ Розенкампф Г.А. Обзорение Кормчей книги в историческом виде. М., 1829. 758 с.

² Павлов А.С. Первоначальный славяно-русский Номоканон. Казань, 1869. 81 с.; Павлов А.С. Номоканон при большом Требнике. М., 1897. 520 с.

³ Суворов Н.С. Следы западно-церковного католического права в памятниках древнего русского церковного права. Ярославль, 1888. 294 с.; Суворов Н.С. К вопросу о западном влиянии на древнерусское право: по поводу книги проф. А.С. Павлова «Мнимые следы католического влияния в древнейших памятниках юго-славянского и русского церковного права». Москва, 1892. Ярославль, 1893. 412 с.

⁴ Соболевский А.И. Церковно-славянские тексты моравского происхождения // Русский филологический вестник. Варшава, 1900. Т. 43. С.150-217; Соболевский А.И. Материалы и исследования// Сборник отделения русского языка и словесности. СПб., 1910. Т. 88. С. 113-116.

Т.Д. Флоринского⁵, Г.Т. Данаилова⁶, Й. Вашицы⁷, С.В. Троицкого⁸ и др. Количество выявленных (Я.Н. Щаповым, М.Н. Тихомировым, Л.В. Миловым)⁹ к 1960-м годам рукописей ЗСЛ превысило 50, что позволило перейти к следующему этапу исследования этого замечательного текста – этапу классификации его списков. В этой связи необходимо отметить прежде всего фундаментальную монографию М.Н. Тихомирова «Закон Судный людем краткой редакции»¹⁰, а также «Закон Судный людем пространной и сводной редакции»¹¹, вышедшие в 1961 году. В первой монографии (речь идет о «Законе Судном людем краткой редакции») на основе тщательного, «пословного» изучения текстов, их лексических, текстологических особенностей, были выявлены четыре крупные группы списков, получившие название редакций и изводов, которые остаются актуальными вплоть до настоящего времени.

В 1970-х годах классификация ЗСЛ была продолжена Л.В. Миловым и Л.И. Бородкиным уже с помощью количественных методов¹². Речь идет об известном методе групп французского текстолога Ж. Фроже¹³. Модификация этого метода Л.И. Бородкиным позволила не только построить генеалогическую стемму из списков, но и восстановить утраченные ее звенья¹⁴. Предложенный в настоящем исследовании подход следует той же

⁵ Флоринский Т.Д. Древнейший памятник болгарского права // Сборник статей по истории права в честь М.Ф. Владимирского-Буданова. Киев, 1904. С. 404-429.

⁶ Данаилов Г.Т. Един паметник на старото българско право. Закон судний людем. Историко-догматична студия. София, 1901. 59 с.

⁷ Вашица Й. Кирилло-мефодиевские юридические памятники // Вопросы славянского языкознания. М., 1963. Вып. 7. С. 12-33; Vašica, J. Zakonъ sudnyi ljudьmъ – soudní zákoník pro lid // Magnae Moraviae Fontes Historici: leges, textus iuridici, supplementa. IV. Brno, 1971. P. 147-198.

⁸ Троицкий С.В. Святой Мефодий, как славянский законодатель // Богословские труды. М., 1961. Сборник 2. С. 83-141; Троицкий С.В. Святой Мефодий или болгарский князь Борис составил закон судный людем // Богословские труды. М., 1968. Сборник 4. С. 117-126.

⁹ Щапов Я.Н. Византийское и южнославянское правовое наследие на Руси в XI-XIII вв., М., 1978. 291 с.; Древнерусские княжеские уставы XI-XV вв. / ред. Я.Н. Щапов. М., 1973. 241 с.; Закон Судный людем краткой редакции / ред. М.Н. Тихомиров. М., 1961. 179 с.

¹⁰ Закон Судный людем краткой редакции...

¹¹ Закон Судный людем пространной и сводной редакции / ред. М.Н. Тихомиров. М., 1961. 290 с.

¹² Бородкин Л.И., Милов Л.В. О некоторых аспектах автоматизации текстологического исследования (Закон Судный людем) // Математические методы в историко-экономических и историко-культурных исследованиях / ред. И.Д. Ковальченко. М., 1977. С. 235-280.

¹³ Froger D.J. La critique des textes et son automatization. Paris, 1968. 280 p.

¹⁴ Бородкин Л.И., Милов Л.В. О некоторых аспектах... С. 235-280.

парадигме изучения ЗСЛ, что и метод групп Фроже-Бородкина – формализованной генеалогической классификации списков. Привлечение вновь открытых списков ЗСЛ, в том числе входящих в т.н. Древнейшую редакцию (в диссертации использовано 57 списков, один из которых – список Макариева Унженского монастыря¹⁵, упомянутый у Я.Н. Щапова как входящий в Софийскую редакцию, – был вновь разыскан и привлечен к рассмотрению автором диссертации), позволяет по-новому взглянуть как на состав вышеупомянутых редакций, так и на взаимоотношения списков между собой.

В отличие от ЗСЛ, второе рассматриваемое нами произведение – «Предсловие покаянию» – не имеет столь представительной историографии. Изучение «Предсловия покаянию» началось, правда, еще со второй половины XIX века в работах И.А. Чистовича¹⁶, И.И. Срезневского¹⁷, но недостаточное количество имеющихся в распоряжении исследователей списков препятствовало их сколько-нибудь уверенной классификации. Так, В.М. Изергину, проводившему их первую классификацию на рубеже XIX–XX веков, было известно всего семь списков. Он же предложил делить все списки «Предсловия» на две редакции: краткую (более древнюю) и пространную (более позднюю)¹⁸.

После Октябрьской революции 1917 г. изучение текста «Предсловия покаянию» было надолго прервано. Вместе с тем, это произведение является одним из интереснейших древнерусских произведений домонгольской эпохи. Оно было одним из первых, в которых «русская церковь начала воспитывать в нравственных, гуманных идеалах древнерусское общество, взывая к совести

¹⁵ Государственный архив Костромской области. Ф. 1486. Оп. 3. № 4.

¹⁶ Чистович И.А. Библиографическая записка об «Обзоре Русской духовной литературы 862-1720» преосв. Филарета, епископа Харьковского, СПб., 1856 // Известия Императорской Академии наук по отделению русского языка и словесности. СПб., 1857. Т. 6. Вып. 3. С. 237-238 (л. 140).

¹⁷ Срезневский И.И. Древние памятники русского языка и письма. Общее повременное обозрение // Известия Императорской Академии наук по отделению русского языка и словесности. СПб., 1861-1863. Т. 10. Стлб. 303, 645-647; Срезневский И.И. Сведения и заметки о малоизвестных и неизвестных памятниках // Сборник отделения русского языка и словесности Императорской Академии наук. СПб., 1867. Т. 12. С. 327-329.

¹⁸ Изергин В.М. «Предсловие покаянию» (Историко-литературный очерк) // Журнал Министерства народного просвещения. СПб., 1891. Ноябрь. С. 142-184.

новопросвещенного народа»¹⁹. Важность «Предслова» характеризует то, что оно включалось не только в канонические сборники, но и в Требники, старообрядческие Зинары; знали о нем и южные славяне (болгары). О древности «Предслова покаянию» свидетельствуют древнерусские лексизмы, рассыпанные по тексту некоторых его списков: изгойство, куна, поклажаи, рѣзоимство, челадинъ, кърчма. Таким образом, можно утверждать, что правильное толкование и классификация его списков позволит обогатить наши знания об общественно-политических и экономических реалиях той далекой эпохи.

Изучение этого замечательного произведения возрождается в наше время усилиями профессионала-лингвиста Г.С. Баранковой²⁰. Число разысканных исследовательницей рукописей к настоящему времени насчитывает не меньше дюжины, что позволило ей провести новую классификацию «Предслова покаянию». В ходе нее была существенно уточнена классификация В.М. Изергина: выделились новые группы и подгруппы, наметились генеалогические «цепочки» списков. В то же время остается актуальным ряд вопросов, прежде всего — взаимоотношение между вышеупомянутыми двумя редакциями. Более конкретно, текст Софийского № 1262 или Софийского № 1285 списков – старших для краткой и пространной редакции соответственно – является более древним²¹? В ходе настоящего исследования был привлечен к изучению 21 список «Предслова покаянию», часть из которых была разыскана при участии автора диссертации. Таким образом, более представительный набор списков делает осмысленным и корректным применение формализованного подхода, основанного на теории

¹⁹ Изергин В.М. «Предслова покаянию»... с.184.

²⁰ Баранкова Г.С. Текстологические и языковые особенности раннего древнерусского канонического сочинения «Предслова покаянию» //Труды Института русского языка им. В.В. Виноградова. № 16. Лингвистическое источниковедение и история русского языка 2016-2017. М., 2018. С. 9-65; Баранкова Г.С. Сборники дидактического характера, обращенные к иереям и мирянам // Древняя Русь. Вопросы медиевистики. М., 2013. № 3 (53). С. 10-11; Баранкова Г.С., Белякова Е.В. Влияние византийских канонических и учительных текстов на славяно-русскую литературную традицию // Традиции и инновации в истории и культуре. Программа фундаментальных исследований Президиума Российской Академии наук. М., 2015. С. 456-466.

²¹ Российская Национальная библиотека, Софийское собрание, № 1262; Российская Национальная библиотека, Софийское собрание, № 1285.

нечетких множеств, что может по-новому поставить задачу классификации этого произведения, получить более реалистичные оценки соотношения списков.

В настоящий момент известно около 90 списков «Закона Судного людем» краткой редакции и 30 списков «Предслова покаянию». Предложенные в настоящем исследовании методология и метод нечеткой генеалогической классификации могут рассматриваться как открытая система. Поэтому при возможном расширении в будущем набора рассматриваемых списков этот метод может использоваться для возможного уточнения классификации списков рассматриваемых произведений.

Предмет исследования заключается в разработке новой методологии и метода количественной формализованной классификации средневековых текстов с большой рукописной традицией.

Степень разработанности темы. При изучении средневековых русских текстов, дошедших до нас в десятках, а то и в сотнях списках, для исследователя возникает ряд текстологических задач: 1) определение структуры стеммы, установление генеалогического родства внутри групп списков (изводы, редакции); 2) определение возможного архетипа. Но даже если изводы уже историографически установлены, то нередко существуют альтернативные гипотезы относительно их состава, взаимосвязей списков между собой.

Исследователи средневековых русских текстов применяют как традиционные текстологические подходы, так и математические, специально разработанные методы. В основе большинства построенных моделей лежит содержательное предположение относительно процесса копирования списков: чем «ближе» генеалогически пара списков друг к другу, тем меньше различий содержат их тексты²². В результате попарного сличения текстов списков формируется матрица близости, которая используется для их дальнейшего

²² Бородкин Л.И. Математические модели классификации древних текстов // Методы количественного анализа текстов нарративных источников. Сборник статей. М., 1983. С. 8-29.

группирования. Пионерами здесь выступили американские текстологи Э. Колвелл и Э.Тьюн, предложившие использовать для целей классификации кластерные методы²³. В нашей стране метод Колвелла-Тьюна был развит в работах А.А. Алексеева²⁴.

Другое направление связано с генеалогической классификацией (кладистические методы) и основано на применении и адаптации для текстологии универсального принципа экономии (парсиномии), согласно которому в процессе эволюции переход из одного состояния в другое должен произойти за минимальное число шагов²⁵. Кладистические методы могут использовать в своих построениях матрицу близости²⁶. Метод Фроже-Бородкина²⁷ налагает более жесткие требования к принципам копирования списков, но это позволяет не только строить стеммы списков, но и восстанавливать утраченные ее звенья²⁸.

Среди перечисленных формализованных методов выделяется одна общая черта, заключающаяся в *однозначности* проводимой ими классификации. Так, кластерные методы предполагают принадлежность каждого списка к одной и только одной группе, а метод Фроже-Бородкина «привязывает» каждый список в выстраиваемой им генеалогической цепочке к одному и только одному непосредственному списку-предку (антиграфу). Вместе с тем трудно не согласиться, что реальная картина классификации

²³ Colwell E.C. and Tune E.W. The quantitative relationship between MS text-types // Biblical and patristic studies in memory of R.P. Casey/ Ed. by J.N. Birdsall and P.W. Thomson. Freiburg, 1963. P. 25-32.

²⁴ Алексеев А.А. Текстология славянской Библии. СПб., 1999. 256 с.; см. также: Горина Н.Л. Опыт оценки текстологической значимости разночтений (на материале списков славянского Евангелия XI-XV вв.) // Труды Отдела древнерусской литературы. СПб., 1996. Т. 49. С. 323-338; Миронова Д.М. Автоматизированная классификация древних рукописей (на материале 525 списков славянского Евангелия от Матфея XI-XVI вв.): дис. ... канд. фил. наук: 10.02.21/ Миронова Дина Марковна. СПб., 2017. 315 с.

²⁵ Прикладная и компьютерная лингвистика / ред. И.С. Николаев, О.В. Митренина, Т.М. Ландо. М., 2016. 320 с.

²⁶ Dees A. Sur une constellation de quatre manuscrits // Melanges de linguistique et de literature offertes a Lein Geschiere. Amsterdam, 1975. P. 1-9; Wattel E. Constructing Initial Binary Trees in Stemmatology // Studies in Stemmatology II / Ed. by Reenen P. van, Mulken M. van. Amsterdam, 2004. P. 145-166.

²⁷ Предложенный Ж. Фроже и модифицированный Л.И. Бородкиным метод групп основан на применении и адаптации идей генетического развития и мутации к процессу исторического бытования текста средневекового произведения.

²⁸ Froger D.J. La critique des textes et son automatization. Paris, 1968. 280 p.; Бородкин Л.И., Милов Л.В. О некоторых аспектах автоматизации текстологического исследования (Закон Судный людям) // Математические методы в историко-экономических и историко-культурных исследованиях / ред. И.Д. Ковальченко. М., 1977. С. 235-280.

может быть гораздо более сложной. Так, списки могут нести в себе черты двух или нескольких изводов; одни и те же чтения могут появляться в списках, находящихся в совершенно разных ветвях стеммы.

На современном этапе для решения подобного рода задач появляется возможность использования более гибких методов, учитывающих неоднородность текста списков. Так, для задач генеалогической классификации текстов списков автором диссертации разработан подход, основанный на применении теории нечетких множеств²⁹.

В основе этой теории лежит осознание нечеткости (неоднозначности, размытости) суждений как универсальной реальности человеческой жизнедеятельности. Так, при качественном анализе сложных систем аппарат теории нечетких множеств позволяет учитывать силу связи между объектами. Отсюда понятен интерес к этой теории со стороны исследователей гуманитарно-исторической направленности. Так, элементы этой теории уже применялись для анализа структуры текстов³⁰. Разработанный автором настоящего исследования алгоритм использует такие понятия и инструментарий теории нечетких множеств как функция принадлежности, класс нечеткой эквивалентности, нечеткое отношение предпочтения. Все они имеют естественную интерпретацию на профессиональном языке текстологов. Так, классу нечеткой эквивалентности соответствуют группы текстологически близких списков (изводы, виды, редакции), нечеткие отношения предпочтения отражают связи между ними.

Разработанная на основе этого алгоритма программа нечеткой генеалогической классификации с заданным уровнем надежности (или, в противоположную сторону, нечеткости) выделяет классы нечеткой эквивалентности и устанавливает между ними генеалого-преемственные

²⁹ Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений. М., 1976. 165 с.; Мелихов А.Н., Бернштейн Л.С., Коровин С.Я. Ситуационные советующие системы с нечеткой логикой. М., 1990. 272 с.

³⁰ Москин Н.Д., Петров А.М. Использование нечетких теоретико-графовых моделей для анализа синтаксической структуры музыкально-стихотворного фольклора (на материале эпических духовных стихов) // Методы и принципы современных гуманитарных исследований: сборник научных работ аспирантов и молодых ученых. Петрозаводск, 2011. С. 13-23.

связи. Другими словами, предложенный метод позволяет строить генеалогическую стемму групп списков с заданным уровнем детализации (вплоть до отдельных списков).

В отличие от метода Фроже-Бородкина, нечеткая классификация не ориентирована на восстановление в стемме пропущенных звеньев. В то же время, в модели нечеткой классификации учтена возможность существования у списка нескольких антиграфов, существования одних и тех же разночтений в разных ветвях стеммы.

Еще один важный сюжет связан с типологизацией разночтений, точнее, с типологизацией ошибочных разночтений (ошибок), поскольку именно они будут учитываться в процессе проведения классификации. Здесь мы опираемся на фундаментальный принцип, сформулированный К. Лахманом еще в XIX веке, о том, что *наличие общих ошибок* в текстах свидетельствует об *общности их происхождения*, их генеалогической близости. Сгруппировав списки по родственным группам, можно построить генеалогическое древо списков, отражающее их взаимоотношения друг с другом³¹. Здесь и далее под “ошибками” будем понимать как изменения первоначального текста стилистического, идеологического, лексического характера, так и явные механические описки и искажения.

Понятно, что разные типы разночтений имеют различную ценность для классификации. Разработанная автором совместно с профессионалом-лингвистом Г.С. Баранковой типологизация разночтений включает 14 типов: лексическая замена; пропуск или вставка главы; пропуск или вставка предложения (или части предложения); пропуск или вставка словосочетания (не более трех слов); пропуск или вставка частицы “не”; смысловая ошибка; пропуск или изменение названия главы; перестановка слова или нескольких слов; исправление; орфография (включающая нормы т.н. 2-го южнославянского влияния); описка или повтор; русизм; архаизм.

³¹ Лихачев Д.С. Текстология. Л., 1983. С. 10-24.

Значимость каждого различия в формализованной классификации учитывается в виде весового коэффициента, соответствующего его типу (всего 14 типов). Для максимального устранения субъективности при назначении весовых коэффициентов автор диссертации применяет т.н. метод косвенной экспертизы³². В соответствии с этим методом исходная сложная задача заменяется на серию простых, заключающихся в попарном сравнении значимости типов друг относительно друга. Причем данное сравнение проводится в терминах качественной шкалы оценок, максимально приближенных к естественному языку (например, “существенно значимее”, “очевидно значимее”, “несравним” и т.д.). Перевод этих качественных оценок в количественные проводится с помощью специальной таблицы соответствий, разработанной и апробированной в рамках научной школы Д.А. Поспелова³³. Полученная таблица из коэффициентов попарных сравнений типов с помощью формализованной математической процедуры преобразуется в искомый ряд весовых коэффициентов.

Территориальные и хронологические рамки исследования. «Закон Судный людем» (ЗСЛ) является одним из древнейших славянских текстов. Его создание датируют IX–X вв. и возводят к переводческой деятельности святого Мефодия и его учеников (“болгарская”, “моравская”, “македонская” теории происхождения ЗСЛ).

Весьма примечательно, что текст ЗСЛ сохранился исключительно на территории Древней Руси и в гораздо более поздних списках (для его краткой редакции – в составе Кормчих и Мерил Праведных конца XIII–XVII вв.).

«Предсловие покаянию» является оригинальным древнерусским произведением, создание которого относят к XII–XIII векам. Это произведение сохранились в более поздних списках в составе различных канонических сборников XIV–XVII вв., в том числе новгородского,

³² Saaty T.L. Measuring the fuzziness of sets // Journal of Cybernetics. Vol. 4. 1974. P. 53-61.

³³ Мелихов А.Н., Берштейн Л.С., Коровин С.Я. Ситуационные советующие системы с нечеткой логикой. М., 1990. 272 с.

псковского, западнорусского происхождения.

Цели и задачи исследования. Целью настоящего исследования является разработка максимально возможной универсальной модели, алгоритмов и методов, а также комплексной программы, ориентированных на решение задачи классификации средневековых текстов с большой рукописной традицией; проведение на основе разработанной методологии классификации двух средневековых произведений, одно из которых переводного, а другое оригинального происхождения: «Закон Судный Людем» (краткой редакции) и «Предсловие покаянию».

В соответствии с указанной целью, автор ставит перед собой следующие **задачи:**

- выявить источниковедческие проблемы текстологической классификации и пути их решения; охарактеризовать основные механизмы и принципы копирования средневековых текстов в ситуации замкнутой контролируемой традиции (когда переписчик стремится максимально точно передать копируемый текст рукописи)³⁴;

- построить соответствующую математическую модель, допускающую неоднородность текста списков, неоднозначность его отнесения к той или иной текстуально близкой группе; адаптировать и развить аппарат теории нечетких множеств в рамках построенной модели текстологической классификации;

- разработать алгоритм и программу, реализующие решение задачи нечеткой генеалогической классификации;

- провести типологизацию разночтений с учетом их разной степени важности для классификации; адаптировать математический метод косвенной экспертизы с целью максимально объективного формирования весового коэффициента для каждого типа разночтений;

- собрать материал списков двух исследуемых произведений, включая

³⁴ Алексеев А.А. Текстология славянской Библии. Спб., 1999. 256 с.

отыскание новых списков в архивах и библиотеках бывшего СССР;

– провести сличение текстов списков и представить результаты этого сличения в формате программы нечеткой генеалогической классификации; провести кодировку полученных разночтений в соответствии с разработанной типологизацией;

– провести автоматическую классификацию списков двух исследуемых произведений с помощью разработанной программы нечеткой генеалогической классификации, выбирая в качестве входного параметра разные уровни надежности;

– сравнить полученные результаты с предыдущими классификациями;

– построить стеммы двух исследуемых произведений на основе результатов нечеткой генеалогической классификации.

Источниковой базой исследования стал материал 78-ми списков «Закона Судного людем» и «Предслова покаянию», как опубликованных (54 списка ЗСЛ – из издания М.Н. Тихомирова и Л.С. Милова³⁵), так и оригиналов текстов списков ЗСЛ, не вошедших в академическое издание указанного произведения и хранящихся в архивах и библиотеках Москвы, Санкт-Петербурга, Киева, Ярославля, Костромы.

Научная новизна. История бытования средневекового памятника – это цепь изменений первоначального текста. Используя традиционный аппарат, текстологи выделяют группы близкородственных текстов и устанавливают между ними взаимосвязи.

Предлагаемый новый подход – это попытка формализовать этот процесс, перевести его на язык "цифр", что дает возможность получать более точные результаты. Исследования, сочетающие в себе традиционный и инновационный подходы, появлялись и ранее. Отличие предлагаемого подхода состоит в его максимальной универсальности. Использование разработанного автором инновационного метода позволяет расширять

³⁵ Закон Судный людем краткой редакции/ ред. М.Н. Тихомиров. М., 1961. 179 с.

границы исследования, в частности, учитывать неоднородность текста в случае его правки по нескольким редакциям. Используя аппарат теории нечетких множеств, этот метод позволяет выделять с заданной степенью надежности группы близких текстов и устанавливать между ними генеалогическо-преемственные связи, что обуславливает научную новизну предполагаемого проекта.

Предлагаемая классификация опирается не только на традиционные для формализованной текстологической классификации количественные данные (число разночтений, коэффициенты близости текстов списков), но и на качественно новые характеристики, такие как весовые коэффициенты разночтений, учитывающие уровень значимости разночтений для классификации (пропуски, вставки, лексические замены и т.п.), что ранее не использовалось.

В качестве объектов настоящего исследования выбраны два средневековых русских произведения с большой рукописной традицией. Одно из них, «Предсловие покаянию», ранее не становилось объектом формализованной классификации.

Методологическая основа работы – принципы объективности, историзма, научности и системности. В исследовании используются следующие методы: историко-генетический, историко-сравнительный, историко-типологический. Из области математической науки используется теория нечетких множеств, теория матриц, теория графов.

Предлагаемая автором настоящей работы модель основывается на следующих предположениях: чем больше доля ошибок (уклонений от нормальных вариантов разночтений), перешедших из одного списка в другой, тем достовернее гипотеза о том, что первый список генетически предшествует второму. И, наоборот, чем меньше эта доля, тем более независимы списки друг

от друга³⁶. Необходимо при этом отметить, что хотя классификация списков строится на основе теории нечетких множеств, но учитываются и сведения о происхождении списков, истории их копирования.

Подсчитанные доли ошибок заносятся в специальную таблицу, носящую название матрицы нечеткого отношения. Элементы этой матрицы отражают характер генеалогической связи одного списка относительно другого. Поскольку число ошибок в каждом списке различно, то эта матрица не является симметричной (значения ее элементов, симметричных относительно главной диагонали, не совпадают).

В зависимости от значений этих пар элементов матрицы возможны три экстремальных случая. Первый – когда оба числа (доли) близки к единице, то есть большая часть ошибок первого и второго списков совпадают. Это значит, что, либо эти списки являются точными копиями друг друга, либо переписаны с одного антиграфа. Второй случай – когда оба числа близки к нулю. Это значит, что списки не имеют общих ошибок и принадлежат различным ветвям стеммы. Наконец, третий случай описывает ситуацию, когда одно из чисел близко к единице, а второе – к нулю. Это значит, что большая часть ошибок первого списка совпадает с ошибками второго, но лишь малая часть ошибок второго списка совпадает с ошибками первого (на математическом языке это значит, что множество ошибок первого списка включается в множество ошибок второго списка). Отсюда можно сделать вывод о первичности первого списка относительно второго.

Таким образом, матрица нечеткого отношения, в отличие от симметричной матрицы близости кластерных методов, содержит в себе возможность как группирования текстов списков (горизонтальный срез), так и их хронологического упорядочивания (вертикальный срез).

Этот содержательный вывод можно облечь с помощью аппарата теории

³⁶ Шпирко С.В. Применение теории нечетких множеств к задаче генеалогической классификации в текстологическом исследовании // Историческая информатика: Информационные технологии и математические методы в исторических исследованиях и образовании. Барнаул, 2013. № 3. С. 39-51.

нечетких множеств в формализованную процедуру, разбивающую исходное нечеткое отношение на две составляющие: симметричную (отношение эквивалентности) и антисимметричную (отношение порядка). Другими словами, нам удастся разбить исходное множество списков на текстологически близкие классы и установить между ними генеалогические связи. При этом степень подобного разбиения на классы (более мелкие или, наоборот, более крупные) регулируется внешним параметром – уровнем надежности (или противоположным ему понятием – уровнем нечеткости).

Резюмируя вышесказанное, отметим, что в основе исследования лежит новый подход, заключающийся в адаптации развитого аппарата теории нечетких множеств к текстологическим задачам генеалогической классификации списков. Так, предложенный алгоритм использует такие стандартные понятия и инструментарий теории нечетких множеств, как функция принадлежности, отношения эквивалентности и порядка, каноническое разбиение и нечеткое покрытие, линейное упорядочивание. Все используемые понятия имеют естественную интерпретацию на профессиональном языке текстологов. Например, каноническое разбиение означает выделение текстуально близких групп списков, линейное их упорядочивание – формирование отдельных ветвей стеммы.

Проводимое исследование базируется на следующих двух фундаментальных принципах. Во-первых, предполагается, что полученная в результате применения нечеткой классификации генеалогическая стемма допустима лишь с определенной степенью достоверности (которая регулируется уровнем надежности). Во-вторых, в рассматриваемую исходную модель закладывается в максимальной степени принцип универсальности, применимости предложенного алгоритма для классификации списков средневекового произведения. Так, пожалуй, единственным предположением модели является то, что список-потомок с большой вероятностью сохранит все ошибки (уклонения от нормального чтения) своего протографа (списка-

предка).

В то же время, данная модель предполагает предварительный этап сличения списков экспертом-текстологом, лежащий за рамками алгоритма (схема этого этапа приводится в приложении диссертации). Заключительный этап классификации, включающий ее содержательный анализ и интерпретацию, также необходимым образом связан с работой профессионального текстолога.

Теоретическая и практическая значимость. Теоретическая значимость исследования заключается в адаптации и развитии метода нечетких множеств применительно к новой области – текстологической классификации. Практическая значимость заключается в том, что предложенная методология дает основу для проведения нечеткой классификации на материале других средневековых текстов, а также для сравнения полученных результатов с результатами как традиционного текстологического, так и формализованного количественного анализа. Разработанная программа нечеткой генеалогической классификации размещена в открытом доступе на сайте исторического факультета и может быть свободно использована исследователями в своих текстологических исследованиях.

Результаты исследования могут использоваться для разработки курсов по количественным формализованным текстологическим методам.

Апробация результатов исследования. Основные положения и выводы диссертации представлены в 10-ти научных публикациях общим объемом 4 п.л., в том числе 4 статьях в изданиях, рекомендованных для защиты в диссертационном совете МГУ имени М.В. Ломоносова. В ходе работы диссертант выступил на следующих научных конференциях: XIV Международная конференция Ассоциации «История и компьютер» (АИК) «Исторические исследования в цифровую эпоху: информационные ресурсы, технологии, методы» (г. Звенигород, 2014 г.); VII научные чтения памяти профессора А.В. Муравьева, (Москва, 2014 г.); Международная научная

конференция «Творческое наследие А.А. Зимина и современная российская историография (VI Зиминские чтения)» (Москва, 2015 г.); XV Международная конференция АИК «Исторические исследования в цифровую эпоху: информационные ресурсы, методы, технологии» (г. Звенигород, 2017 г.); Международная научная конференция «Маргиналии-2017: границы культуры и текста» (г. Торжок, 2017 г.); IX Международная конференция «Комплексный подход в изучении Древней Руси» (Москва, 2017 г.); XLVI Международная филологическая конференция (Санкт-Петербург, 2017 г.); Круглый стол «Древняя Русь» (г. Москва, 2017 г., Институт российской истории).

Диссертация обсуждена и рекомендована к защите на кафедре исторической информатики исторического факультета МГУ имени М.В. Ломоносова.

В соответствии с поставленными целями и задачами **структура диссертации** включает в себя введение, шесть глав, разбитых на параграфы, заключение, список литературы и источников, приложения.

Положения, выносимые на защиту:

1. Применение математических методов, в частности, теории нечетких множеств, в область текстологической классификации средневекового произведения является обоснованным и полезным при проведении традиционного анализа текстологами; предложенная модель и разработанный на ее основе алгоритм нечеткой классификации списков адекватно описывает механизмы и принципы бытования средневекового текста с большой рукописной традицией; проведенная классификация позволяет независимо от традиционного анализа проверить ряд гипотез, в частности уточнить состав той или иной текстологически близкой группы списков, выявить неявные связи между списками; применяемая нечеткая классификация с заданным уровнем надежности позволяет снять (или, по крайней мере, посмотреть под иным углом зрения) ряд сложнейших вопросов, связанных, в частности, с неоднозначностью отнесения списка к той или иной группе.

2. Нечеткая классификация, выделяя в составе ЗСЛ 4 извода (также как

и авторитетные исследователи этого произведения), содержит при этом ряд уточняющих элементов. Так, нечеткая классификация выделяет среди старших списков ЗСЛ Новгородский 1280/82 годов и Устюжский списки как ближайшие к его архетипу³⁷.

3. Нечеткая классификация устанавливает определяющее влияние Варсонофьевского списка на Троицкий и Софийский изводы; нечеткая классификация устанавливает определяющее влияние Новгородского 1280/82 списка на Чудовский извод; в качестве первоисточника для Чудовского извода нечеткая классификация устанавливает группу списков типа Крестининского; нечеткая классификация подтверждает выделение Ферапонтовского вида внутри Чудовского извода.

4. Нечеткая классификация «Предслова покаянию» согласуется с выводами авторитетных исследователей о существовании 2-х больших редакций этого произведения – пространной и краткой; в качестве первичной нечеткая классификация «Предслова покаянию» устанавливает пространную редакцию по отношению к краткой.

5. Нечеткая классификация «Предслова покаянию» выделяет в составе пространной редакции три устойчивые группировки. В состав “древнейшей” в том числе входят старший список группы Софийский № 1285, Епархиальный № 384³⁸, Мазуринский³⁹ список болгарского происхождения. К одной из двух оставшихся групп⁴⁰ особую близость имеет список Епархиальный № 384; проведенная классификация «Предслова покаянию» выделяет в составе краткой редакций 4 устойчивые группы. В состав “древнейшей” входит Софийский № 1262, являющийся старшим в этой группе.

Основное содержание работы

Во введении очерчивается проблематика исследования, ставятся цели и задачи, дается обзор научной литературы по теме диссертации,

³⁷ О списках ЗСЛ см. Приложение 10 текста диссертации.

³⁸ ГИМ. собр. Епархиальное. № 384.

³⁹ РГАДА. собр. Мазурина (Ф.196). Оп. I. № 1700.

⁴⁰ В состав этой группы входят два списка: РГБ. собр. Троице-Сергиевой Лавры. № 793; РНБ. собр. Кирилло.-Белозерское. № 4/ 1081.

обосновывается его актуальность, теоретическая и практическая значимость, научная новизна.

Глава 1 – **«Историографические и источниковедческие аспекты исследования»** – посвящена важнейшим аспектам традиционной текстологии, связанной с классификацией средневековых текстов. В первом параграфе главы освещается историография этой проблематики, уходящей своими корнями к классическим работам К. Лахмана. На примере работ Ж. Бедье, Ж. Кантэна, В. Грега показаны основные этапы развития европейской и мировой текстологии середины XIX – середины XX в. Особо разбираются теоретические и методологические идеи и разработки русских текстологов – В.Н. Перетца, С.А. Бугославского, а также в известной мере А.А. Шахматова. Оглядываясь в прошлое русской текстологической мысли, можно заключить, что она развивалась в сложном взаимодействии, борьбе и синтезе двух подходов – метода общих ошибок К. Лахмана и метода конъюнктуральной критики А.А. Шахматова. Первый параграф завершается рассмотрением прикладных аспектов текстологического анализа – классификации списков Русской Правды пространной редакции В.П. Любимовым и классификации Кормчих современной исследовательницей М.В. Корогодиной. Второй параграф посвящен рассмотрению двух источников – объектов настоящего исследования: «Закона Судного людем» краткой редакции (ЗСЛ) и «Предсловия покаянию». Приводятся взгляды ведущих ученых относительно авторства, датировки, происхождения ЗСЛ (в частности, рассматривается т.н. болгарская, моравская и македонская теории). Прослеживаются обстоятельства появления и бытования ЗСЛ на Руси, выясняется характер связи ЗСЛ с другими памятниками (Мерило Праведное, Русская Правда), что влечет важные методологические выводы в плане его классификации. Далее рассматривается историография изучения «Предсловия покаянию», раскрываются этапы открытия и исследования этого оригинального древнерусского произведения, приводятся взгляды ведущих ученых относительно датировки и обстоятельств

его создания. В частности, отмечается важность изучения терминов и понятий (*изгойство, наклады, поклажаи, рѣзы, кърчьма*), приводимых в тексте, для выводов относительно его датировки. Подробно рассматриваются результаты традиционной классификации списков этого произведения в работах В.М. Изергина и современной исследовательницы Г.С. Баранковой.

Глава 2 – «**Формализованные количественные методы текстологической классификации**» – посвящена изучению различных аспектов, связанных с формализованными количественными методами классификации средневековых текстов. Глава 2 состоит из четырех параграфов. В первом параграфе рассматривается таблица (матрица) близости списков, являющаяся исходной для большинства количественных методов классификации. Исследуются вопросы построения этой матрицы, способы ее улучшения и корректировки. Во втором параграфе излагается суть подхода Колвелла-Тьюна, основанного на применении математических методов кластеризации к матрице близости и развитого в работах современного исследователя А.А. Алексева. Рассматриваются результаты применения кластерного метода Колвелла-Тьюна к греческой новозаветной рукописной традиции, выявляются проблемы, связанные с неоднозначностью ее классификации. Третий параграф посвящен изучению второго наряду с кластерными методами направления формализованной количественной классификации – кладистическим методам. Большинство этих методов основано на применении универсального принципа парсиномии (экономии) и использовании матрицы близости (А. Деес, Э. Ваттель, метод покрывающего дерева максимальной общей длины, агрегированная модель классификации Л.И. Бородкина). В агрегированной модели классификации используется один из иерархических кластерных методов – метод К-средних. Особенное внимание уделяется изложению метода групп Ж. Фроже, развитого в работах Л.В. Милова и Л.И. Бородкина. Этот метод, в отличие от предыдущих, базируется на применении к целям классификации теоретико-множественного и теоретико-графового подходов. Завершают параграф результаты

применения метода Фроже-Бородкина к материалу списков ЗСЛ.

Глава 3 – «Метод нечеткой генеалогической классификации для задач текстологического исследования» – посвящена теоретическим вопросам, связанным с нечеткой генеалогической классификацией текстов списков средневекового произведения. Глава 3 состоит из четырех параграфов. В первом параграфе дается неформальное введение в модель нечеткой классификации. Рассматриваются идеи и принципы, лежащие в основе этой модели классификации. На ряде примеров демонстрируется плодотворность привлечения к этой модели идей, принципов и аппарата математической теории нечетких множеств. Дается неформальное определение важнейших понятий этой теории – функции нечеткой принадлежности, класса нечеткой эквивалентности. Второй параграф посвящен изучению ключевого понятия для дальнейшего анализа – нечеткого отношения. На конкретных примерах рассматриваются типы нечетких отношений (симметричные, антисимметричные, транзитивные, отношения порядка). Приводится содержательное описание алгоритма нечеткой генеалогической классификации (формальное его изложение с необходимыми доказательствами вынесено в приложение). Третий параграф посвящен изучению матрицы нечеткого отношения предпочтения, являющейся исходной для метода нечеткой классификации. Обосновываются причины рассмотрения такого рода матрицы, приводится алгоритм ее построения на основе попарного сравнения текстов списков и подсчета доли (процента) унаследованных ошибочных разночтений (уклонений от нормы) из одного списка в другой. Обосновывается необходимость рассмотрения в качестве таковых разночтений неустранимых ошибок (в терминах Д.С. Лихачева), включая вставки или пропуски, перестановки, изменения в названиях глав, лексические замены и т.д. Приводятся примеры построения матрицы нечеткого отношения. Четвертый параграф посвящен сравнению результатов работы метода нечеткой генеалогической классификации и рассмотренных ранее формализованных количественных методов (Колвелла-Тьюна и Дееса-

Ваттеля). На одних и тех же исходных данных (матрица близости списков греческой новозаветной традиции) все методы дают схожие результаты, что косвенно подтверждает корректность и правомерность предложенного метода. Указывается, что применение нечеткого подхода позволяет в значительной степени снять проблему т.н. ядра и периферии (неоднозначности отнесения списка к той или иной группировке), дополнив ее задачей выбора надлежащего порога надежности (или противоположного понятия нечеткости) в качестве входящего параметра метода.

Глава 4 – **«Методы количественной оценки значимости различий для нечеткой классификации»** – посвящена типологизации различий и количественной оценки их важности при вычислении матрицы нечеткого отношения. Глава 4 состоит из четырех параграфов. В первом параграфе приводится типологизация различий Л.П. Жуковской и Н.Л. Гориной как пример лингвистического подхода, направленного на выделение текстуально близких групп списков. Во втором параграфе приводится типологизация различий В.П. Любимовым как пример комплексного подхода, учитывающего и лингвистический и текстологический анализ различий. В третьем параграфе автором настоящего исследования приводится совместная с профессионалом-лингвистом Г.С. Баранковой типологизация различий, учитывающая опыт предыдущих типологизаций. Четвертый параграф посвящен разработке формализованной процедуры оценки значимости типов различий (построения весовых коэффициентов). Обосновывается необходимость привлечения для этой цели математического метода косвенной экспертизы. Суть этого метода состоит в том, что исходная сложная задача назначения каждому типу весового коэффициента заменяется на ряд простых подзадач, заключающихся в попарном сравнении типов. В силу сложности получения количественных оценок сравнения экспертом используется формализация качественных характеристик, разработанная в рамках подхода

Д.А. Поспелова⁴¹. Сформированная на основе качественной шкалы значимости таблица попарных сравнений типов с помощью формализованной математической процедуры⁴² преобразуется в ряд искомым весовых коэффициентов.

Глава 5 – «**Нечеткая генеалогическая классификация списков «Закона Судного людем» краткой редакции»**» – состоит из трех параграфов. Первый параграф посвящен отдельным методологическим аспектам, связанным с нечеткой классификацией переводного текста с замкнутой контролируемой традицией. Вторым параграф посвящен собственно нечеткой генеалогической классификации – приводятся стеммы ЗСЛ с разным уровнем надежности, позволяющие выделить текстуально близкие группы списков и установить между ними генеалого-преемственные связи. Идея проведения нечеткой классификации ЗСЛ состоит в том числе в необходимости учета новых списков, которых не было в распоряжении предыдущих исследователей. В третьем параграфе приводится сравнение результатов нечеткой генеалогической классификации с классификацией ЗСЛ методом групп Л.И. Бородкина и Л.В. Милова, классификациями списков Русской Правды пространной редакции В.П. Любимова, М.Н. Тихомирова, и классификацией списков Кормчих М.В. Корогодиной. В связи с этим отметим, что достоверность классификации возрастает, если схожие результаты получаются с применением разных методов и подходов. Так, нечеткая классификация согласуется в целом с предыдущими результатами, однако применение теории нечетких множеств позволяет учесть влияние не только одного списка-предка. При этом варьирование внешнего параметра нечеткой классификации – порога надежности – позволяет уточнить состав отдельных групп списков, выявить неявные взаимосвязи как между группами, так и отдельными списками.

⁴¹ Мелихов А.Н., Берштейн Л.С., Коровин С.Я. Ситуационные советующие системы с нечеткой логикой. М., 1990. 272 с.

⁴² Saaty T.L. Measuring the fuzziness of sets // Journal of Cybernetics. Vol. 4. 1974. P. 53-6.

Глава 6 – «**Нечеткая генеалогическая классификация списков «Предслова покаянию»**» – состоит из двух параграфов. Первый параграф посвящен отдельным методологическим аспектам, связанным с нечеткой классификацией оригинального древнерусского, характеризующегося открытой рукописной традицией. Второй посвящен собственно нечеткой генеалогической классификации – приводятся стеммы «Предслова покаянию» с разным уровнем надежности, позволяющие выделить текстуально близкие группы списков и установить между ними генеалогические связи. Нечеткая генеалогическая классификация «Предслова покаянию» дает схожие результаты, что и традиционный текстологический анализ. В то же время, варьирование внешнего параметра нечеткой классификации – порога надежности – позволяет выделить новые группы, уточнить их состав, выявить неявные взаимосвязи как между списками. В завершении параграфа приводятся основные выводы, позволяющие утверждать правомерность и корректность предложенного формализованного подхода к построению текстологической классификации.

В **заключении** автором приводятся ответы на ранее поставленные задачи, подводятся итоги исследования.

Публикации в рецензируемых научных изданиях, индексируемых в базах данных Web of Science, Scopus, RSCI, и в изданиях, рекомендованных для защиты в диссертационном совете МГУ имени М.В. Ломоносова по специальностям 07.00.02 – Отечественная история и 07.00.09 – Историография, источниковедение и методы исторического исследования:

1) *Шпирко С.В., Баранкова Г.С.* О применении одного математического метода для задачи формализованной текстологической классификации древнерусского произведения «Предслова покаянию» // Клио. 2018. № 9. С. 13-21. (0,48 п.л.). Импакт-фактор РИНЦ 0,141.

2) *Шпирко С.В., Баранкова Г.С.* О некоторых аспектах построения формализованной генеалогической классификации текстов списков

средневекового произведения с применением теории нечетких множеств (на материале «Закона Судного людем») // Исторический журнал: научные исследования. 2017. № 1. С. 56-64. (0,48 п.л.). Импакт-фактор РИНЦ 0,185.

3) *Шпирко С.В., Баранкова Г.С.* Применение одного математического подхода к задаче генеалогической классификации текстов списков средневекового произведения (на примере «Закона Судного людем») // Исторический журнал: научные исследования. 2017. № 5. С. 37-69. (0,8 п.л.). Импакт-фактор РИНЦ 0,185.

4) *Шпирко С.В.* О новых подходах по применению формальных математических методов в изучении средневековых текстов // Электронный научно-образовательный журнал История. 2015. № 8 (41). DOI: 10.18254/S0001256-0-1. (0,9 п.л.). Импакт-фактор РИНЦ 0,062.

Другие публикации по теме диссертации:

1) *Шпирко С.В., Баранкова Г.С.* Применение одного математического подхода к задаче генеалогической классификации текстов списков средневекового произведения (на примере «Закона Судного людем»). Тезисы доклада // Древняя Русь. Вопросы медиевистики. 2017. № 3. С. 166-187. (0,5 п.л.). Импакт-фактор РИНЦ 0,062.

2) *Шпирко С.В., Нестеров А.Ю.* Программное приложение Programma's Edit редактирования текстов, интегрированное в программный комплекс PFuzClass нечеткой генеалогической классификации для текстологических задач // Историческая информатика. 2017. № 4. С. 67-77. (0,4 п.л.). Импакт-фактор РИНЦ 0,515.

3) *Шпирко С.В., Баранкова Г.С.* Изгойство и изгой в древнерусском каноническом сочинении «Предсловие покаянию» (к проблеме датировки и текстологической группировки списков памятника) // Международная конференция «Маргиналии-2017: границы культуры и текста». Тезисы доклада. Москва, 2017. С. 6.

4) *Шпирко С.В.* Методы текстологической генеалогической классификации: матрица близости vs матрица нечеткого отношения // Историческая информатика. Барнаул, 2016. № 3. С. 13-25. (0,4 п.л.). Импакт-фактор РИНЦ 0,515.

5) *Шпирко С.В.* Об одном подходе к генеалогической классификации в текстологии с применением теории нечетких множеств (на примере средневекового славянского произведения «Закон Судный людем») // Информационный бюллетень Ассоциации История и Компьютер. Тезисы доклада. Барнаул, 2016. № 45. С. 150-151.

6) *Шпирко С.В.* Применение теории нечетких множеств к задаче генеалогической классификации в текстологическом исследовании // Историческая информатика. Барнаул, 2013. № 3. С. 39-53. (0,4 п.л.). Импакт-фактор РИНЦ 0,515.