

ХОЗЯИНОВ

Сергей Александрович

**АТРИБУЦИЯ ПУБЛИЦИСТИЧЕСКИХ ПРОИЗВЕДЕНИЙ,
ПРИПИСЫВАЕМЫХ А. С. ПУШКИНУ
(тексты 1830—1836 гг.)**

Специальность 10.02.21 — прикладная и математическая лингвистика

А В Т О Р Е Ф Е Р А Т

диссертации на соискание ученой степени
кандидата филологических наук

Санкт-Петербург

2008

Хозяинов

Работа выполнена на кафедре математической лингвистики Федерального государственного образовательного учреждения высшего профессионального образования «Санкт-Петербургский государственный университет»

Научный руководитель: доктор филологических наук, профессор
Марусенко Михаил Александрович

Официальные оппоненты: доктор филологических наук, профессор
Маркович Владимир Маркович

кандидат филологических наук, доцент
Виландеберк Анна Арнольдовна

Ведущая организация: Балтийский государственный технический
университет «Военмех» им. Д. Ф. Устинова

Защита состоится «17» июня 2008 г. в 16³⁰ ч. на заседании диссертационного совета Д 212.199.17 по защите диссертаций на соискание ученой степени доктора наук в Российском государственном педагогическом университете им. А. И. Герцена по адресу: 191186, г. Санкт-Петербург, наб. р. Мойки, д. 48, корп. 14, ауд. 314.

С диссертацией можно ознакомиться в фундаментальной библиотеке Российского государственного педагогического университета им. А. И. Герцена.

Автореферат разослан «14» мая 2008 г.

НАУЧНАЯ БИБЛИОТЕКА КГУ



0000434452

Ученый секретарь
диссертационного совета
доктор филологических наук,
профессор

Л. А. Пиотровская

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

В исследовании разрабатываются вопросы теоретического и прикладного аспектов проблемы атрибуции анонимных и псевдонимных текстов — в рамках таких областей исследования прикладной лингвистики, как структурное моделирование и формализация единиц языка, вероятностные и статистические модели языка и речи, статистическая стилистика. В работе представлен один из способов решения задач перевода информации, содержащейся в неформализованном виде в тексте, на формализованный искусственный язык — на основе математического аппарата теории распознавания образов.

Актуальность темы исследования. Необходимость точного определения границ публицистического наследия А. С. Пушкина не подлежит сомнению. Так как решающее значение здесь приобретает выбор методов атрибуции, актуализация проблемы именно сегодня объясняется возникновением ряда новых методик атрибуции (М. А. Марусенко; Л. В. Милов, Л. И. Бородин; И. П. Севбо и др.), основанных на применении точных математических методов и позволяющих решать задачи атрибуции на новом уровне.

Объектом исследования является авторство приписываемой А. С. Пушкину публицистики.

Исследование осуществляется на **материале** анонимных, псевдонимных и авторизованных публицистических текстов «Литературной газеты, издаваемой бароном А. А. Дельвигом» (1830—1831 гг., далее — ЛГ), альманаха «Северные цветы» (1832 г.), «Современника, литературного журнала, издаваемого Александром Пушкиным» (за 1836 г.), газеты «Северная пчела» (за 1836 г.), а также неопубликованной статьи «О Татищеве». В общей сложности изучению подвергаются 135 текстов (объемом 3447 предложений), написанных каждым из изучаемых авторов, а также 46 текстов со спорным авторством (объемом 650 предложений). Исследуемый период (1830—1836 гг.) охватывает около 80% всей журналистской деятельности А. С. Пушкина.

Предметом исследования являются индивидуально-стилевые особенности публицистики А. С. Пушкина, П. А. Вяземского, А. А. Дельвига, Н. В. Гоголя и О. М. Сомова.

Целью исследования является атрибуция приписываемых А. С. Пушкину публицистических произведений.

Достижение цели основано на проверке трех взаимоисключающих **гипотез** (первая — основная, другие — альтернативные): (H_0): все 46 статей написаны А. С. Пушкиным; (H_a^1): часть статей принадлежит А. С. Пушкину, а другая часть — одному или нескольким альтернативным авторам, что не исключает возможность совместной работы писателей над отдельными текстами; (H_a^2): А. С. Пушкину не принадлежит ни одна из приписываемых ему статей, и все они являются произведениями других авторов, что не исключает возможность совместной работы этих писателей над отдельными текстами.

С учетом свойств объекта и предмета исследования, согласно указанной цели и сформулированным гипотезам ставятся **задачи**: 1) определить состав и границы корпуса приписываемых А. С. Пушкину текстов со спорным

авторством; 2) для каждого из предполагаемых авторов определить перечень текстов, оптимально характеризующих его публицистику; 3) в целях выполнения предшествующих задач определить критерии спорности и подлинности авторства; 4) на уровне критериев формирования корпусов текстов обеспечить их сопоставимость между собой; 5) обосновать выбор применяемой методики атрибуции; 6) осуществить процедуру атрибуции.

Теоретические основы исследования. Априорно принимаются положения: 1) корректность результатов квантитативно-структурного изучения авторских стилей зависит от степени когерентности исследуемых текстов в хронологическом и жанровом отношениях (Б. Я. Слепак); 2) для адекватного определения особенностей индивидуального стиля писателя необходимо обращение к анализу синтаксического уровня языка текста (И. П. Севбо; Г. Я. Мартыненко; М. А. Марусенко); 3) для повышения качества классификации текстов в задачах атрибуции необходимо применение процедур многомерного анализа (М. А. Марусенко). Методологическая основа эксперимента по атрибуции — методика, разработанная М. А. Марусенко. В рамках этой методики процедура атрибуции интерпретируется как задача классификации текстов и основывается на применении методов распознавания образов. Методика апробирована на материале ряда сложных проблем авторства, которые не могли получить своего решения на пути применения как традиционных филологических, так и современных математических средств атрибуции. Морфолого-синтаксический анализ текста основан преимущественно на подходе, принятом в «Русской грамматике» 1980 г.

Методы исследования: 1) метод сравнительно-исторического анализа использован для определения и описания основных направлений, периодов и тенденций в истории изучения проблемы авторства публицистики, приписываемой А. С. Пушкину; 2) методы классификации применяются для систематизации изученных материалов; 3) метод сопоставительного анализа применен для характеристики современных методов атрибуции; 2) методы распознавания образов (методы математического моделирования, снижения размерности параметрического пространства, построения решающего правила классификации) применяются в эксперименте по атрибуции.

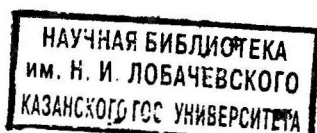
Основные положения, выносимые на защиту:

1. Проблема авторства приписываемой А. С. Пушкину публицистики не получает достаточно точного решения на пути применения традиционных филологических методов атрибуции.

2. Предложенное В. В. Виноградовым деление принципов атрибуции на субъективные и объективные избыточно по отношению к задаче раскрытия методологического аспекта названной проблемы авторства.

3. Методология атрибуции, основанная на применении методов теории распознавания образов, предполагает корректную реализацию всех стадий функционирования знаковых систем в эксперименте по атрибуции.

4. Названная методология атрибуции позволяет существенно повысить уровень решения проблемы авторства приписываемой А. С. Пушкину публицистики.



5. Предлагаемый вариант решения этой проблемы авторства является оптимальным в смысле минимизации влияния субъективного фактора на конечный результат эксперимента по атрибуции.

Научная новизна исследования: 1) предложено систематизированное и подробно иллюстрированное изложение проблемы авторства приписываемой А. С. Пушкину публицистики, дана периодизация изучения проблемы; 2) решение этой проблемы впервые осуществлено математическими методами атрибуции; 3) впервые с опорой на эти методы предложено решение проблемы авторства текстов сверхмалых объемов (от двух предложений); 4) пересмотрена классификация принципов атрибуции, разработанная В. В. Виноградовым; 5) введено понятие «коэффициент значимости решения» для оценки результатов вероятностной процедуры атрибуции.

Теоретическая значимость результатов исследования определяется тем, что в нем: 1) раскрыты содержательный и методологический аспекты проблемы авторства публицистики, приписываемой А. С. Пушкину; 2) на материале исследованных периодических изданий 1830—1836 гг. составлены указатели текстов, приписываемых А. С. Пушкину, П. А. Вяземскому, А. А. Дельвигу и О. М. Сомову; 3) показана избыточность деления принципов атрибуции на субъективные и объективные в рамках традиционной классификации принципов атрибуции (В. В. Виноградова); 4) дан критический обзор ряда практических реализаций некоторых современных методик атрибуции, показано, что они не вполне соответствуют задачам изучения авторских стилей; 5) логические этапы процедуры атрибуции, основанной на применении методов распознавания образов, охарактеризованы с точки зрения функционирования в них знаковых систем.

Практическая значимость результатов исследования определяется возможностью их применения: 1) в подготовке публицистики А. С. Пушкина, А. А. Дельвига, П. А. Вяземского, Н. В. Гоголя и О. М. Сомова к научному изданию; 2) в поиске решений проблемы атрибуции текстов малых объемов; 3) при разработке правил параметризации стилей на материале русскоязычных текстов; 4) для решения частных задач по атрибуции.

Рекомендации об использовании результатов исследования. Полученные результаты могут быть использованы: 1) в качестве исходных данных для решения прикладных задач: а) по атрибуции публицистических текстов, приписываемых П. А. Вяземскому, А. А. Дельвигу и О. М. Сомову (приложение 3); б) по изданию публицистики А. С. Пушкина и этих писателей (глава 3 (комментированные перечни публицистических текстов писателей), приложения 1 и 2); 2) в практике преподавания соответствующих теоретических и прикладных научных дисциплин: а) при чтении курсов по истории русской литературы и журналистики первой трети XIX в.; б) при чтении специальных курсов по методам атрибуции.

Апробация результатов исследования. Результаты исследования освещены в докладах на XXXV и XXXVII Международных филологических конференциях (СПбГУ, Факультет филологии и искусств, 2006 г. и 2008 г. соответственно), изложены в трех публикациях общим объемом 1,4 п. л.

Объем и структура работы. Работа состоит из введения, трех глав, заключения, списка литературы, списка сокращений и четырех приложений. Основное содержание исследования изложено на 188 страницах. Список литературы включает 113 наименований. Объем приложений 1—4 составляет 76 страниц. Приложение 1 информирует о том, кем и в какой период изучения проблемы впервые был приписан Пушкину тот или иной публицистический текст (в соответствии с периодизацией, представленной в главе 1). Приложение 2 в форме комментированного указателя раскрывает разработанную в главе 1 рабочую классификацию этих текстов. Приложение 3 представляет собой комментированный указатель публицистических текстов, приписываемых П. А. Вяземскому, А. А. Дельвигу и О. М. Сомову. В приложении 4 описаны правила лингвистического анализа текстов (в качестве иллюстраций даны примеры из исследуемых публицистических произведений).

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснована актуальность темы исследования, указаны его объект, материал и предмет, определены цели и задачи, сформулированы гипотезы, названы теоретические основы и методы исследования, перечислены выносимые на защиту положения, описана научная новизна, теоретическая и практическая значимость результатов исследования, даны рекомендации по использованию, наконец, приведены сведения об апробации результатов исследования, его объеме и структуре.

В **главе 1** — «Содержание и методы изучения проблемы авторства приписываемой А. С. Пушкину публицистики» — описаны содержательный и методологический аспекты изучаемой проблемы авторства, сформулирована атрибуционная гипотеза. Первый аспект раскрыт посредством решения двух задач: 1) определения и описания направлений и этапов изучения проблемы; 2) выявления тенденций и закономерностей, характерных для истории ее изучения. Методологический аспект представлен в работе двояко: 1) с точки зрения изучения объекта исследования общими филологическими методами; 2) в плане применения к изучению этого объекта специальных критериев, учитывающих его специфику.

Раздел 1 главы посвящен краткой характеристике ЛГ — издания, с которым связана основная и наиболее сложная часть проблемы. Оно выходило в Санкт-Петербурге раз в пять дней — с 1 января 1830 года по 30 июня 1831 года. Каждые 36 номеров газеты, охватывающие полгода издания ЛГ, формировали один том, к которому редакцией составлялось оглавление, в основном уточняющее авторов прозаических и стихотворных художественных произведений. Авторы статей и заметок в большинстве случаев не указывались ни в оглавлении, ни на страницах самой газеты. Это было вызвано необходимостью оградить издание и его сотрудников от нежелательных вмешательств цензурного комитета. За время существования газеты были изданы три тома; последний, 37-й номер ЛГ за 1831 год, был первым и

единственным в составе IV тома. Редакторской работой в основном занимались О. М. Сомов, П. А. Вяземский и А. С. Пушкин.

В разделе 2 главы 1 описаны три периода изучения проблемы: 1) становления (1850—1900-е гг.); 2) развития и изучения (1900—1930-е гг.); 3) систематизации материалов и исследований (конец 1930-х — 2000-е гг.).

Первый период связан с деятельностью П. В. Анненкова, В. П. Гаевского, П. А. Ефремова и П. О. Морозова. Они приписали А. С. Пушкину 21 статью. Почти все эти статьи печатаются в собраниях сочинений А. С. Пушкина и доныне. Авторство большинства из этих статей подтверждается документально. К началу XX в. были определены ключевые вопросы, затрагивающие период сотрудничества А. С. Пушкина в ЛГ: о степени его участия в газете, об истинном смысле «политических и общественных убеждений Пушкина» (выражение П. В. Анненкова), о достоверности документальных свидетельств, т. е. осуществлена постановка всей проблемы.

Второй период делится на два этапа: 1) начальный (1900—1910-е гг.); 2) этап критического изучения проблемы (1920—1930-е гг.). Наибольшее количество спорных атрибуций создано на *первом этапе*. Н. О. Лернер приписал А. С. Пушкину 16 статей (как и П. В. Анненков), но в сочинения поэта в итоге вошли только три. Значительный вклад в разработку проблемы на этом этапе внесли А. Г. Фомин, Н. К. Замков, М. Л. Гофман и С. А. Венгеров. *Второй этап* составляют труды Б. В. Томашевского, Н. А. Сиявского и М. А. Цявловского, В. П. Красногорского, В. В. Виноградова. Эти ученые стремятся к критическому осмыслению прежнего опыта изучения проблемы, к более глубокому и детальному ее анализу. Однако на этом этапе А. С. Пушкину приписывается еще около 20 публицистических текстов.

Для *третьего периода* характерна тенденция к обобщению и систематизации материалов и исследований, к подведению итогов. С начала 1940-х гг. заметно снижаются темпы поиска новых статей. За весь этот период поэту не было приписано и 10 текстов. В основном систематизация материала представлена в указателях: 1) «Пушкин в печати. 1814 — 1837» Н. А. Сиявского и М. А. Цявловского (издания 1914 и 1938 гг.); 2) «Пушкин в печати за 100 лет. 1837 — 1937» (1938), составленный К. П. Богаевской под редакцией М. А. Цявловского, продолжающий библиографию изданий текстов Пушкина, отраженную в предыдущем указателе; 3) указатель содержания ЛГ, составленный Е. М. Блиновой (1966); 4) указатель содержания «Современника» 1836—1837 гг., составленный Е. М. Рыскиным (1967). Примечания к двум последним содержат много подробных сведений об изучаемой проблеме.

Важнейшим обобщающим исследованием по проблеме является книга «Проблема авторства и теория стилей» В. В. Виноградова (1961). В 1966 г. выходит в свет коллективная монография «Пушкин: Итоги и проблемы изучения». Она ставит новые задачи: так, К. И. Тюнькин в разделе «Критико-публицистическая проза» отмечает, что достигнутые в изучении проблемы успехи не исключают необходимости дальнейшей разработки отдельных вопросов проблемы. Вероятно, первое обращение к историко-методологическому аспекту изучения проблемы находим в книге

Н. Е. Мясоедовой «Подвиг честного человека» (2004). Здесь описан путь развития критериев атрибутирования — от использования документальных свидетельств на начальном этапе изучения проблемы к разработке в трудах Н. О. Лернера, Б. В. Томашевского и В. В. Виноградова методических приемов и критериев атрибуции разной степени сложности и объективности.

В завершение раздела 2 главы 1 рассмотрен вопрос о том, насколько подробно отражена изучаемая проблема в собраниях сочинений А. С. Пушкина, Н. В. Гоголя, П. А. Вяземского и А. А. Дельвига. Наиболее информативны в этом отношении два издания сочинений Пушкина. Издание 1907—1915 гг. Брокгауза и Ефрона, под редакцией С. А. Венгерова, содержит подготовленный Н. О. Лернером раздел «Новые приобретения Пушкинского текста». Академическое издание 1899—1929 гг. обнаруживает подробный многоаспектный комментарий к публицистике А. С. Пушкина, подготовленный Н. К. Козминым. Большую роль в изучении проблемы сыграло приложение Б. В. Томашевского к «Полному собранию стихотворений» А. А. Дельвига (1934) — «Библиография изданий сочинений Дельвига», содержащее перечень статей, рукописи которых исследователь обнаружил среди бумаг Дельвига.

Характерной чертой критического подхода к изучению проблемы является процесс разработки специальных критериев атрибуции. Описанию этого процесса посвящен **раздел 3** главы 1.

Первым к разработке критериев атрибуции обратился Н. О. Лернер. В своих изысканиях он опирался на замечания П. В. Анненкова и В. П. Гаевского относительно степени участия А. С. Пушкина в ЛГ. Но документальные данные не позволяли найти все неподписанные статьи поэта в газете. Пытаясь отыскать эти статьи, Н. О. Лернер выходит за рамки, заданные П. В. Анненковым и В. П. Гаевским. На этом пути он вырабатывает специальные критерии атрибуции, которые становятся основой аргументации его новых гипотез: в тексте следует искать следы характерных элементов стиля автора и системы его убеждений; метод исключения позволяет признать автором статьи А. С. Пушкина в том случае, если нет возможности обосновать иное авторство.

Впоследствии исследователи не раз отмечали, что критерии Лернера недостаточно точны. Одним из первых это сделал В. Я. Брюсов. Он говорил, что филологический анализ лишь тогда позволит «почти с полной достоверностью» устанавливать принадлежность той или иной статьи Пушкину, когда будет составлен «полный словарь пушкинского языка», будут сделаны «длинные ряды статистических подсчетов (относительно употребления Пушкиным особых выражений, оборотов речи <...> и т. д.)» и параллельно будет осуществлена такая же работа в отношении «других писателей пушкинской эпохи»¹. Не располагая подобными данными, определить автора текста можно лишь по наличию авторизованной рукописи этого текста, по подписи или всем известному псевдону писателя. В решении

¹ Брюсов В. Я. Новооткрываемый Пушкин // Биржевые ведомости. — 1916. — 21 октября. — № 15875. — С. 2.

проблемы могут помочь также документальные свидетельства современников писателя, стоявших «близко к литературным кругам» и к нему самому.

Б. В. Томашевский заметил, что все подлинные статьи А. С. Пушкина появились в «Литературной газете» во время пребывания поэта в Петербурге: до 4 марта 1830 г., с 19 июля по 10 августа 1830 г. и с мая 1831 г. Из этого он заключил, что А. С. Пушкин не мог написать статьи, дата публикации которых не вписывается в эту хронологию. Однако, как заметила Н. Е. Мясоедова, Пушкин пользоваться правительственными каналами связи, предоставленными в распоряжение П. А. Вяземского Н. А. Мухановым, и таким образом оперативно реагировать на все текущие изменения¹. Характерно, что даты опубликования большинства приписываемых А. С. Пушкину статей относятся к названным Б. В. Томашевским периодам, особенно к первому, часть которого (январь—февраль) поэт занимался в ЛГ редакторской работой. Интересно также, что все статьи, приписываемые А. А. Дельвигу, относятся к периоду март—ноябрь 1830 г., когда он выполнял обязанности издателя ЛГ. С ноября 1830 г. последние легли на плечи О. М. Сомова. С его именем связано лишь около 30 статей, опубликованных в ЛГ до этого — в I и II тт. за 1830 г., и более 100 статей, опубликованных после этого — в III-м т. за 1831 г. Согласно этим цифрам, объем труда О. М. Сомова после его вступления в должность издателя ЛГ должен был увеличиться почти в семь раз.

В. В. Виноградов посвятил критическому обзору предшествующих атрибуций отдельное исследование (см. выше, с. 7). Он скептически оценивает критерии Н. О. Лернера и «хронологический» критерий Б. В. Томашевского и формулирует важное требование: «Отрицание пушкинского авторства <...> получает полную силу только при утверждении подлинного имени автора», но замечает, что «не всегда можно с полной достоверностью определить имя настоящего автора статьи по языку, стилю, по содержанию критических высказываний, по особенностям и оттенкам литературного вкуса»². Эта оговорка обозначает границы возможностей традиционных филологических методов атрибуции. Атрибуции В. В. Виноградова отвечают требованиям комплексного подхода к атрибуции, обозначенного П. Н. Берковым. Согласно этому подходу, произведение следует характеризовать одновременно в трех аспектах: биографическом, идеологическом, стилистическом. Каждая атрибуция Виноградова — развернутая цепь рассуждений, содержащая четыре компонента: 1) анализ стилевых и 2) языковых особенностей текста; 3) анализ образной системы текста и авторских литературных приемов; 4) анализ содержания текста в аспекте истории его создания.

В. Э. Вацууро заметил, что подпись «Изд.» не является достоверным свидетельством авторства: в № 1 ЛГ за 1831 г. издана подписанная так статья А. А. Дельвига «Борис Годунов», хотя издателем газеты был О. М. Сомов.

¹ Мясоедова Н. Е. «Подвиг честного человека» / отв. ред. В. Е. Ветловская. — СПб. : [б. и.], 2004. — С. 96.

² Виноградов В. В. Проблема авторства и теория стилей. — М. : Гослитиздат, 1961. — С. 374.

Вацуро полагает, что эта подпись могла быть средством маскировки имени настоящего автора, которое привлекало ненужное автору внимание цензуры¹.

В разделе 4 главы 1 описаны три главные тенденции изучения проблемы: 1) к поиску неизвестных статей поэта; 2) к полемике, критическому пересмотру проблемы; 3) к обобщению и систематизации материала. Первая тенденция определяет развитие и усложнение проблемы, остальные две — ее изучение. Каждый из описанных периодов изучения проблемы характеризуется доминированием той или иной тенденции, но не исчерпывается ею.

Раздел 5 посвящен формированию атрибуционной гипотезы. Из 64 приписываемых Пушкину текстов 24 сегодня считаются написанными им (22 статьи ЛГ и две статьи «Современника» Пушкина). Предполагаемыми авторами (соавторами) приписываемых поэту текстов являются И. А. Бессонов, П. А. Бессонов, А. И. Болтин, П. А. Вяземский, Н. В. Гоголь, А. А. Дельвиг, В. Ф. Одоевский, М. П. Погодин, О. М. Сомов и В. Н. Щастный. Степень изученности тех или иных приписываемых Пушкину текстов различна: 1) авторство текста вариативно и не считается установленным (31 текст); 2) авторство текста инвариантно, но не считается установленным, так как не принято издателями сочинений Пушкина (шесть текстов); 3) авторство текста инвариантно и доказано на основе анализа документальных источников (из 22 текстов ЛГ, автором которых признан Пушкин, это 14 статей; прочие относятся к двум следующим случаям); 4) авторство текста инвариантно и считается установленным, но не подтверждается документально (шесть статей); 5) авторство текста считается установленным, но вариативно (семь статей).

Как правило, атрибуция признается «правильной» или при наличии неоспоримых документальных свидетельств в пользу гипотезы, или ввиду согласия издателей сочинений с ней. Но нередко наблюдаются разногласия между различными издателями, а документальные данные интерпретируются по-разному. Здесь важен сам факт научной полемики, свидетельствующий о том, что разные аспекты изучения текста обнаруживают в нем черты, позволяющие связать его с разными авторами. Если таких черт нет, отсутствует полемика. Разделение приписываемых А. С. Пушкину текстов на две группы произведено по трем формальным признакам, характеризующим историю изучения каждого текста: 1) наличие альтернативной точки зрения — гипотезы, атетезы, сомнения (независимо от хронологии их возникновения); 2) наличие прямых документальных доказательств авторства — рукописи, свидетельств современников или самого автора; 3) вхождение текста в состав собраний сочинений автора в качестве написанного им произведения. Неоспоримость авторства констатируется: 1) при положительной характеристике по признакам 2 и 3, отрицательной — по первому; 2) при положительной характеристике по признаку 3, отрицательной — по двум первым. В прочих случаях авторство текста рассматривается как спорное. Перечень текстов со спорным авторством приведен в таблице 1 (*N* — объем текста в предложениях).

¹ Вацуро В. Э. К изучению «Литературной газеты» Дельвига — Сомова // Временник Пушкинской комиссии, 1965. — Л., 1968. — С. 23—36.

Таблица 1. Тексты, приписываемые А. С. Пушкину

Код	Текст	N
T01	Сань-Цзы-Цзин (ЛГ, 1830, № 1, с. 5—6)	15
T02	Письмена Вавилонские (ЛГ, 1830, № 1, с. 7)	9
T03	Обман зрения в Персии (ЛГ, 1830, № 1, с. 7—8)	14
T04	Краткая всеобщая география (ЛГ, 1830, № 3, с. 22—23)	8
T05	Всеобщий атлас (ЛГ, 1830, № 3, с. 23)	3
T06	«Когда Макферсон...» (ЛГ, 1830, № 5, с. 40)	8
T07	Литературные новости (ЛГ, 1830, № 6, с. 48)	8
T08	Сен-Марс, или Заговор (ЛГ, 1830, № 7, с. 53)	12
T09	Письмо к издателю Литературной газеты (ЛГ, 1830, № 7, с. 55—56)	17
T10	«Старый генерал Щ...» (ЛГ, 1830, № 8, с. 66)	5
T11	«Шувалов, заспорив однажды...» (ЛГ, 1830, № 8, с. 66)	2
T12	«Острая путка...» (ЛГ, 1830, № 10, с. 82)	2
T13	«Кажется русским любителям чтения...» (ЛГ, 1830, № 11, с. 90)	6
T14	«На нынешней неделе поступил в продажу...» (ЛГ, 1830, № 11, с. 90)	2
T15	Невский альманах на 1830 год (ЛГ, 1830, № 12, с. 96)	9
T16	«Англия есть отечество карикатуры...» (ЛГ, 1830, № 12, с. 98)	8
T17	«В третьем нумере Московского Вестника...» (ЛГ, 1830, № 16, с. 129—130)	18
T18	«Кто-то заметил, что русские шутки...» (ЛГ, 1830, № 18, с. 146)	7
T19	«В 39-м № Северной пчелы...» (ЛГ, 1830, № 20, с. 161)	7
T20	«Требуем ли публика извещения...» (ЛГ, 1830, № 20, с. 162)	9
T21	«Все благоразумные люди предвидели...» (ЛГ, 1830, № 23, с. 186)	5
T22	«„Мне все советуют,“ говорил однажды...» (ЛГ, 1830, № 24, с. 193—194)	8
T23	«С некоторых пор, Журналисты...» (ЛГ, 1830, № 36, с. 293)	13
T24	Василий Шуйский (ЛГ, 1830, № 38, с. 16)	19
T25	«Новые выходки...» (ЛГ, 1830, № 45, с. 72)	9
T26	«В газете: <i>Le Furet</i> ...» (ЛГ, 1830, № 45, с. 72)	2
T27	«В нынешнем году Северная пчела отличилась...» (ЛГ, 1830, № 45, с. 72)	10
T28	«Сочинитель письма из Рима...» (ЛГ, 1830, № 50, с. 114)	13
T29	«Горестно видеть...» (ЛГ, 1830, № 53, с. 139—140)	25
T30	<i>Histoire du roi Bohême et de ses sept châteaux</i> (ЛГ, 1830, № 55, с. 153—155)	10
T31	<i>La Confession</i> (ЛГ, 1830, № 60, с. 193—195)	45
T32	Киргиз-Кайсак (ЛГ, 1831, № 5, с. 42)	13
T33	«Г. Издатель Телеграфа объявил...» (ЛГ, 1831, № 5, с. 42)	4
T34	«В №№ 17 и 18 Московского телеграфа...» (ЛГ, 1831, № 8, с. 66—67)	31
T35	«В Московском вестнике на 1830 год...» (ЛГ, 1831, № 18, с. 148)	7
T36	«Издатель Телеграфа...» (ЛГ, 1831, № 27, с. 221—222)	19
T37	«Повесть сия сочинена Батюшковым...» (Северные цветы, 1832, с. 1—2)	2
T38	Пять стихотворений барона Дельвига (Северные цветы, 1832, с. 3—4)	6
T39	История средних веков (Современник, 1836, т. I, с. 302)	2
T40	«Вот книги...» (Современник, 1836, т. I, с. 318—319)	11
T41	Французская академия (Современник, 1836, т. II, с. 14—52)	15
T42	«Для очистки совести нашей...» (Современник, 1836, т. II, с. 311—312)	10
T43	Письмо к издателю (Современник, 1836, т. III, с. 321—329)	85
T44	От редакции (Современник, 1836, т. IV, с. 299—300)	6
T45	Несколько слов о Современнике (Северная пчела, 1836, № 86, с. 341—344)	34
T46	О Татищеве	77

Глава 2 — «Методы проверки атрибуционной гипотезы» — посвящена теоретико-методологическому аспекту проблемы атрибуции.

В разделе 1 рассмотрена классификация принципов атрибуции. Самую подробную типологию принципов атрибуции разработал В. В. Виноградов. Он выделяет *субъективные* (коммерческие, конъюнктурные, эстетические, психологические и идеологические) и *объективные* (документально-рукописные (археологические), исторические (биографии, свидетельства современников и т. д.), историко-идеологические и сопоставительно-идеологические, историко-стилистические, художественно-стилистические, «лингво-статистические») принципы атрибуции. Изучение проблемы авторства приписываемой А. С. Пушкину публицистики показало, что выделение субъективных принципов атрибуции в особую группу избыточно.

Во-первых, В. В. Виноградов и сам замечает, что *объективные* историко-стилистические принципы «иногда открывают дорогу субъективным оценкам. Степень их убедительности до некоторой степени зависит от развитости словесно-художественного вкуса и чутья исследователя...»¹.

Во-вторых, поскольку любая методология атрибуции опирается на достигнутые наукой знания, степень ее точности (объективности) зависит от этих знаний, т. е. обусловлена исторически. Так, в послереволюционном пушкиноведении был пересмотрен взгляд на «систему убеждений» поэта. Это позволило К. И. Тюнькину заключить, что в дореволюционном изучении публицистики поэта «подинно научного ее анализа сделано не было. <...> над старым пушкиноведением <...> тяготели <...> представления о Пушкине якобы приверженце “чистого искусства” в области эстетики и стороннике самодержавия в области политики»². Все же именно тогда была определена большая часть истинно пушкинской публицистики, напечатанной в ЛГ.

В-третьих, вне опоры на анализ документальных данных, сопоставительный анализ содержания, языка и стиля текста аргументация исследователей не обнаруживает применения каких-либо принципов атрибуции (если понимать последние как набор правил, задающих условия и ограничивающих сферу действия *методических приемов*).

Вслед за В. В. Виноградовым легко проследить процесс развития принципов атрибуции. Это сделано в разделе 2. В основе метода атрибуции текста «по объективно-историческим качествам его стиля <...>, его литературно-художественной техники»³ лежат принципы, выделенные Ф. Е. Коршем в работе о подлинности окончания «Русалки» Пушкина (1898): 1) следует изучать стилистику художественной речи текста в ее сравнении с

¹ Виноградов В. В. Теория литературных стилей и принципы атрибуции анонимных и псевдонимных произведений // О принципах определения авторства в связи с общими проблемами теории и истории литературы: науч. сес. (тез. докл. и сообщ.). — Л., 1960. — С. 4.

² Пушкин: итоги и проблемы изучения / ред. Б. П. Городецкий, Н. В. Измайлов, Б. С. Мейлах. — М.; Л.: Наука, 1966. — С. 492.

³ Виноградов В. В. Лингвистические основы научной критики текста // Вопросы языкознания. — 1958. — № 2. — С. 10.

обликом литературного языка эпохи; 2) нельзя опираться на художественные качества языка и стиля, поскольку они всегда *понимаются субъективно*; 3) автором атрибутируемого текста может быть признан А. С. Пушкин лишь в том случае, если стиль и индивидуальная техника словесного творчества, наблюдаемые в этом тексте, идентичны аналогичным характеристикам произведений А. С. Пушкина; 4) нельзя опираться на представления о «непогрешимости» пушкинского стиля. Таким же образом рассмотрены остальные два метода. Метод отбора типических признаков текста наиболее полно сформулирован П. Н. Берковым и имеет две вариации: 1) Ф. И. Витязев предлагает определять автора по набору ключевых слов; 2) Н. А. Морозов предлагает сопоставлять так называемые лингвистические спектры — графики частот встречаемости слов в тексте. Методы изучения рукописей произведений, подробно описанные С. М. Бонди, позволяют выявить факты и закономерности, описывающие последовательность работы автора над текстом.

В разделе 3 главы 2 перечислены этапы развития точных методов атрибуции в русской науке. Описание этих этапов было дано Л. В. Миловым¹. Метод анализа лингвистических спектров, разработанный в 1915 г. Н. А. Морозовым, — первый в русской науке опыт применения методов статистики к задачам атрибуции. В 50-60-е гг. XX в. начинает развиваться квантитативное направление в изучении языка и стиля текста. Для периода 60—70-х гг. XX в. характерно интенсивное применение методов статистики в лексикографии и грамматике, а затем и в синтаксисе языка, речи и авторского стиля. Наибольшее развитие получили подходы, основанные на анализе синтаксиса: 1) направленный на построение и анализ графов синтаксических связей, характеризующих синтаксическую структуру авторской фразы (например, метод И. П. Севбо); 2) направленный на изучение взаимосвязей между синтаксическими структурами текста (методология М. А. Марусенко).

В разделе 4 обсуждены современные проблемы и методы атрибуции, обоснован выбор используемой методики.

В рамках квантитативно-структурного изучения авторских стилей ставятся две важные проблемы: 1) определения специфики стиля автора; 2) точности атрибуции. Первая проблема осложняется тем, что авторский стиль изменчив и зависит от влияния хронологического, жанрового и тематического факторов. На конкретное решение проблемы влияет и то, какие уровни языка изучаемого текста оказываются предметом исследования. Наиболее распространены морфология, лексика и синтаксис. Способы решения проблемы точности атрибуции определяются спецификой используемого метода. В силу природы квантитативных методов мера этой точности поддается вычислению.

В плане постановки задачи исследователи Петрозаводского университета (В. Н. Захаров, А. А. Рогов, Ю. В. Сидоров) развивают подход, разработанный Г. Хетсо. А именно: изучая статьи, приписываемые Ф. М. Достоевскому, они в основном исходят из предположения, что «стилистический инвариант» (термин

¹ От Нестора до Фонвизина: новые методы определения авторства / Л. В. Милов, Л. И. Бородин, Т. В. Иванова и др. — М.: Прогресс, 1994. — С. 7—8.

Г. Хетсо) писателя представлен распределением частей речи на первых и последних трех позициях предложений. Очевидно, что при таком подходе остальные позиции предложений не исследуются, а это означает значительные потери полезной информации.

Существует ряд методик, основанных на анализе текста как реализации так называемой цепи А. А. Маркова. В качестве цепей Маркова рассматриваются последовательности пар букв или пар грамматических классов в тексте (Л. И. Бородин; Д. В. Хмелев; И. О. Тарнопольская).

Нельзя сказать, насколько адекватна задачам атрибуции методика такого типа, разработанная Д. В. Хмелевым, так как в проведенном им эксперименте по атрибуции не были соблюдены условия количественно-структурного изучения стилей (соответствие жанра и времени написания сравниваемых произведений). В этом смысле не меняет существа дела расширение объекта анализа за счет грамматических классов слов и их семантико-грамматических разрядов. Весьма ограничена сфера применения этого метода: его успешность констатируется при атрибуции текстов объемом не менее 70 тыс. символов.

В основе метода И. О. Тарнопольской лежат наблюдения, показывающие, что для текстов одного автора одинаков, а для текстов разных авторов различен коэффициент диаграммной энтропии, достигший так называемой «точки стабилизации» — значения, остающегося неизменным при дальнейшем увеличении длины выборки. Исследователь ставит задачу определения оптимальной длины выборки, достаточной для достижения «точки стабилизации». Для разных текстов искомая длина колеблется в интервале от 10 до 17 тысяч знаков, что почти на порядок меньше длин текстов, исследуемых Хмелевым. Начальная точка поиска оптимальной длины — 1200 знаков, а длина каждого шага принята исследователем равной 1000 знаков.

Остается неясным, адекватно ли формализованное описание текста как реализации цепи Маркова в виде последовательности пар букв в тексте какому-либо уровню языка. Предположение Д. Хмелева, что такое описание отражает порядок следования морфем, требует специальных доказательств.

Наиболее последовательный метод атрибуции, основанный на изучении частот парной встречаемости грамматических классов слов, был разработан Л. И. Бородиным еще в 70-е гг. XX в. Исследователь констатирует необходимость вычленения из общего фона грамматических связей только самых сильных. «Сила» связи при этом понимается статистически: редко встречающиеся связи случайны и несущественны для стиля автора. Для вычленения сильных грамматических связей вводится некоторое пороговое значение, с которым сравниваются полученные частоты.

Многие исследователи говорят о том, что для успешного решения задач атрибуции необходимо обращение к изучению синтаксиса текстов¹. Как сказано выше, механизмы такого изучения разработаны в рамках подходов, отраженных в методе И. П. Севбо и методологии М. А. Марусенко. Для

¹ *Марусенко М. А.* Атрибуция анонимных и псевдонимных произведений методами распознавания образов. — Л.: Изд-во Ленингр. ун-та, 1990. — С. 15—16, 18—19, 22.

решения изучаемой нами проблемы авторства используется последняя, так как ее точность не зависит от количества сопоставляемых авторов (что определяется применением: а) математических методов выделения стилеразличающих параметров из первичного набора параметров; б) методов многомерной классификации в процедуре атрибуции). Лингвистический анализ текста в рамках этого подхода направлен на описание структурной организации синтаксиса текста. Информация о линейной организации синтаксиса не менее ценна для характеристики стиля текста, так как отражает порядок следования синтаксических структурных элементов. Важной задачей является поиск путей совместного использования синтаксических параметров, отражающих как структурные, так и линейные порядки внутри синтаксиса текста.

Раздел 5 главы 2 посвящен характеристике основных логических этапов методологии атрибуции М. А. Марусенко с точки зрения функционирования в них знаковых систем. Как отмечает А. Соломоник, на результаты конкретного исследования влияет корректность реализации в нем каждой из трех стадий функционирования знаковых систем¹: 1) выбора знаковой системы (следует помнить, что решение практических задач может потребовать выбора нескольких систем); 2) представления изучаемого явления в формализованном виде (в виде последовательного ряда дискретных знаков); 3) обработки и трансформации зашифрованной последовательности знаков по правилам избранной для этого знаковой системы. Оценить методологию помогает анализ ее логических этапов, практических реализаций этой методики. Согласно подходу А. Соломоника, вызывает сомнения корректность реализации первой стадии в рамках методик атрибуции, основанных на анализе встречаемости парных сочетаний букв в тексте: такой выбор не учитывает сложности лингвистического аспекта задач атрибуции. Не всегда однозначны и результаты автоматизации второй стадии, так как это является особой сложной задачей.

В этом отношении последовательно используемая нами методология атрибуции. Она состоит из трех логических этапов: 1) формирования атрибуционной гипотезы; 2) проверки гипотезы; 3) интерпретации результатов проверки гипотезы. Рассмотрение каждого этапа как самостоятельного акта познания позволяет трижды проследить схему функционирования знаковых систем в исследовании по атрибуции. Так, второй этап состоит из двух процедур: 1) преобразования изучаемых текстов в математические модели; 2) процедуры атрибуции. Для осуществления первой из них применяются метаязыки языковой знаковой системы, включающие описание системы, методов и правил ее изучения и функционирования. Согласно избранным уровням лингвистического анализа, этими метаязыками являются морфология и синтаксис. Для фиксации наблюдений используется математическая система записи (наблюдаемые явления кодируются посредством чисел), более высокий уровень иерархии знаковых систем. Основанием для представления изучаемых явлений в виде математических моделей является одно из фундаментальных

¹ Соломоник А. Философия знаковых систем и язык. — Изд. 2-е, доп. и испр. — Мн.: МЕТ, 2002. — С. ii—iv.

свойств знаковых систем — гомоморфизм. Процедура атрибуции реализует третий этап функционирования знаковых систем: последовательность явлений текста, зашифрованная в виде ряда дискретных знаков (математическая модель), обрабатывается по правилам избранной для этого знаковой системы (методы распознавания образов).

Глава 3 — «Атрибуция приписываемой А. С. Пушкину публицистики методами теории распознавания образов» — описывает эксперимент по атрибуции. Согласно используемой методике, процесс атрибуции включает ряд формализованных процедур¹. Каждой из них посвящен отдельный раздел.

Раздел 1 посвящен процедуре определения априорного алфавита классов. *Априорный класс (АК)* — это множество объектов, объединенных общими свойствами. В нашем случае такими объектами являются тексты, авторство которых достоверно (априорно) известно. Необходимость последующей статистической обработки результатов лингвистического анализа этих текстов требует того, чтобы каждый АК был внутренне однороден. Для выполнения этого требования тексты, входящие в класс, проходят тест по четырем критериям: 1) *подлинности авторства* (оно признается достоверным в случаях: а) если текст издан с подписью автора; б) если на авторство текста указывают документальные источники (автограф текста; свидетельства современников, не противоречащие фактам биографии писателя и прямо указывающие на его авторство; свидетельства автора) и при всем этом не существует альтернативных точек зрения на авторство текста); 2) *жанрового соответствия* (большинство текстов АК — обзоры литературы и рецензии; к ним примыкают стилистически однородные с ними статьи и заметки полемического характера и политического содержания, нередко завуалированного под вид литературной или иной критики и полемики (Мясоедова 2004)); 3) *хронологического соответствия* (состав АК ограничен текстами, написанными в период с 1830 по 1837 г.); 4) *законченности произведения* (незавершенные тексты, остававшиеся не изданными в этот период, не включаются в АК). Текст, не отвечающий этим требованиям, не попадает в АК (за редкими исключениями). Структуру АК Ω_1 (Вяземский П. А.) составили 18 текстов объемом 946 предложений, АК Ω_2 (Гоголь Н. В.) — 29 текстов (613 предложений), АК Ω_3 (Дельвиг А. А.) — 36 текстов (591 предложение), АК Ω_4 (Пушкин А. С.) — 43 текста (1001 предложение), АК Ω_5 (Сомов О. М.) — 9 текстов (296 предложений). При определении границ предложений мы пользуемся элементами «формально-пунктуационного метода структуризации текста»².

Раздел 2 главы 3 посвящен процедуре определения *априорного словаря параметров* (АСП) и анализа исследуемых текстов на языке этих параметров. АСП включает в себя параметры, позволяющие описывать морфолого-

¹ В поисках потерянного автора : этюды атрибуции / М. А. Марусенко, Б. Л. Бессонов, Л. М. Богданова и др. — СПб. : Филол. фак. С.-Петерб. гос. ун-та, 2001. — С. 8—17.

² Гринбаум О. Н. Компьютерные аспекты стилистики // Прикладное языкознание : учебник / отв. ред. А. С. Герд. — СПб. : Изд-во С.-Петерб. ун-та, 1996. — С. 454—464.

синтаксическую структуру текстов, созданных на разных языках и относящихся к разным жанрово-стилистическим разновидностям. АСП составили 49 параметров, позволяющих количественно измерить: 1) структурные элементы синтаксиса предложения (число элементарных предложений в цельном предложении; число сочиненных, главных и подчиненных предложений; число групп однородных членов и число членов этих групп; число причастных, деепричастных оборотов, обособленных членов; число групп существительных и членов групп существительных и т. д.); 2) частеречное наполнение текста (число знаменательных и служебных слов, число существительных, прилагательных, местоимений, спрягаемых и именных форм глагола, наречий, предлогов, сочинительных и подчинительных союзов в предложении). Лингвистический анализ текстов был произведен вручную — с опорой на «Русскую грамматику» 1980 г. и «Толковый словарь русского языка» С. И. Ожегова и Н. Ю. Шведовой 2006 г. издания. Случаи, в которых эти издания не позволяли найти решение, оговорены особо.

В разделе 3 описана процедура определения словаря *информативных параметров* (ИП), позволяющих классифицировать *распознаваемые объекты* (РО), т. е. тексты со спорным авторством. Словарь определен на основе данных пяти матриц. Каждая матрица отражает результаты лингвистического анализа текстов одного из АК и обладает размерностью $N \times n$, где $N=49$ (число параметров), а $n=100$ (число отобранных предложений). В таблице 2 даны статистические характеристики всех АК по ИП, процедура определения которых описана ниже.

Таблица 2. Статистические характеристики АК по ИП

Параметр	Ω_1		Ω_2		Ω_3		Ω_4		Ω_5	
	$\bar{x}_i$	σ_i	$\bar{x}_i$	σ_i	$\bar{x}_i$	σ_i	$\bar{x}_i$	σ_i	$\bar{x}_i$	σ_i
X09	0,650	0,857	0,650	0,821	0,510	0,689	-0,330	0,551	0,520	0,759
X18	5,030	3,836	3,890	2,930	4,260	3,296	3,330	3,039	4,670	3,671
X24	1,100	1,235	1,510	1,411	1,010	1,185	0,790	0,957	1,300	1,124
X32	1,660	1,451	1,740	1,368	1,500	1,219	1,320	0,909	1,990	1,878

Процесс формирования словаря ИП основан на применении так называемой схемы Бонгарда и предполагает осуществление двух этапов¹.

На первом этапе из АСП выделяется подмножество параметров, релевантных для различения одной или более пар авторов. Исходя из предположения о том, что значения параметров распределены по нормальному или близкому к нему закону, в качестве критерия классификации параметров мы использовали t-критерий Стьюдента:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)^{0,5}}. \quad (1)$$

¹ В поисках потерянного автора... — С. 9—11.

Критическое значение критерия при уровне значимости $\alpha=0,05$ равно 1,96. Если при сравнении двух АК наблюдаемое значение t -критерия для i -го параметра выше 1,96, то с доверительной вероятностью 0,95 это свидетельствует о наличии статистически существенных различий между этими АК по данному параметру. Классификация позволила выделить подмножество из 29 параметров, наименее мощные из которых релевантны для различения одной пары авторов (АК), наиболее мощные — пяти и семи пар авторов. В таблице 3 приведены результаты сопоставления всех АК по ИП.

Таблица 3. Результаты сопоставления АК по t -критерию Стьюдента

Параметр	$t_{наб.}$									
	Ω_1/Ω_2	Ω_1/Ω_3	Ω_1/Ω_4	Ω_1/Ω_5	Ω_2/Ω_3	Ω_2/Ω_4	Ω_2/Ω_5	Ω_3/Ω_4	Ω_3/Ω_5	Ω_4/Ω_5
X09	0,000	1,273	3,141	1,136	1,306	3,236	1,163	2,040	0,098	2,026
X18	2,362	1,522	3,474	0,678	0,839	1,327	1,661	2,074	0,831	2,812
X24	2,187	0,526	1,984	1,198	2,714	4,223	1,164	1,444	1,776	3,455
X32	0,401	0,844	1,986	1,391	1,310	2,557	1,076	1,184	2,189	3,211

На втором этапе реализуется процедура свертывания параметрического пространства на подмножестве релевантных параметров, результатом которой и является словарь ИП. Центральным методический прием этой процедуры — обработка корреляционной матрицы связей параметров, выполняемая путем вычисления средних внутри- и внегрупповой корреляции, а на их основе — критерия эффективности каждого релевантного параметра:

$$\bar{r}^m = \frac{\sum_{i=1}^m |r_{ij}| - 1}{m-1}, \quad \bar{r}^{n-m} = \frac{\left(\sum_{i=1}^n |r_{ij}| - 1 \right) - \left(\sum_{i=1}^m |r_{ij}| - 1 \right)}{n-m-1}, \quad E_i = \frac{\bar{r}_i^{n-m}}{\bar{r}_i^m}, \quad (2), (3), (4)$$

где m — число релевантных параметров (29), r_{ij} — коэффициент корреляции i и j -го параметров в матрице, n — число параметров из АСП (49). Наиболее эффективны параметры, обладающие наименьшей внутригрупповой и наибольшей внегрупповой корреляциями, т. е. наибольшим значением E_i .

Для определения искомых величин были составлены две матрицы:

1) объектно-признаковая матрица размерностью $N \times n$, где $N=500$ (общее количество объектов — отобранных предложений всех классов), $n=49$ (количество параметров), представляющая собой пространство объектов, в котором каждый вектор x^j (столбец) имеет длину N (равную числу строк), и задан значениями, которые j -й параметр принимает на каждом из объектов пространства объектов;

2) корреляционная матрица связей параметров $R = \{\bar{\rho}_{jk}\}_{j,k=1}^n$, размерностью $n \times n$, где $n=49$, а выборочные коэффициенты корреляции $\bar{\rho}_{jk}$ представляют собой косинус угла α_{jk} в N -мерном пространстве (объектов) между векторами x^j и x^k :

$$\bar{\rho}_{jk} = \cos \alpha_{jk}. \quad (5)$$

Значения $\bar{\rho}_{jk}$ лежат в интервале $-1 \leq \bar{\rho}_{jk} \leq 1$. Матрица симметрична, т. е. $\bar{\rho}_{jk} = \bar{\rho}_{kj}$, $j, k = \overline{1, n}$, а каждая ячейка ее главной диагонали содержит значение корреляции j -го параметра с самим собой, т. е. $\bar{\rho}_{jj} = 1$, $j = \overline{1, n}$.

Для того чтобы словарь ИП был достаточно мощным, мы применили критерий E_i к параметрам, различающим от четырех до семи пар авторов (см. таблицу 4, где N — число пар авторов).

Таблица 4. Показатели корреляции и эффективности параметров

N	Параметр	$\bar{r}_i^m$	$\bar{r}_i^{n-m}$	E_i	N	Параметр	$\bar{r}_i^m$	$\bar{r}_i^{n-m}$	E_i
7	X28	0,4163	0,2199	0,5281	4	X26	0,4738	0,2802	0,5914
5	X24	0,2620	0,1987	0,7585		X32	0,3960	0,2533	0,6395
	X42	0,4332	0,2465	0,5690		X53	0,4921	0,2822	0,5734
4	X09	0,3571	0,2261	0,6330		X54	0,5069	0,2686	0,5300
	X18	0,5159	0,3276	0,6351		X55	0,5033	0,2699	0,5362

Решающее правило отбора ИП: $E_i > 0,6$. Характеристики параметров, удовлетворяющих условию, приведены в таблице 5.

Таблица 5. Словарь ИП

Параметр	$\bar{r}_i^m$	$\bar{r}_i^{n-m}$	E_i
X09 (число подчиненных предложений первой степени)	0,3571	0,2261	0,6330
X18 (число служебных слов)	0,5159	0,3276	0,6351
X24 (число наречий)	0,2620	0,1987	0,7585
X32 (число подлежащих)	0,3960	0,2533	0,6395

Раздел 4 главы 2 описывает процедуру построения математических моделей (определения координат эталонов АК и РО), образуемых статистическими характеристиками индивидуальных стилевых особенностей исследуемых текстов. Следует предположить, что данные выборки, на материале которых был определен набор ИП, отражают свойства совокупностей неодинаково точно. Мету этой точности позволяет оценить так называемая относительная стандартная ошибка оценки¹:

$$V_{\bar{x}} = \frac{V}{\sqrt{n}} \sqrt{1-f}, \quad (6)$$

где $V = \frac{s}{\bar{x}}$ — коэффициент разброса значений признака (отношение выборочного среднего квадратичного отклонения признака (s) к выборочному среднему значению признака ($\bar{x}$) для совокупности); $f = \frac{n}{N}$ — доля отбора (отношение объема выборки (n) к объему совокупности (N)).

Под оценкой понимается среднее значение некоторого параметра, полученное на основе данных выборки. Например, в отношении АК

¹ В поисках потерянного автора... — С. 14—15.

Ω (Пушкин А. С.) $V_{\bar{x}}$ для характеристики текста, описываемой параметром X09, составляет 0,16 (16%); X18 — 0,09 (9%); X24 — 0,11 (11%); X32 — 0,07 (7%). Определив заранее приемлемую величину $V_{\bar{x}}$, можно вычислить необходимый объем выборки — n :

$$n = \frac{N}{1 + \left(\frac{V_{\bar{x}}}{V} \right)^2} \quad (7)$$

Объем выборки n для каждого ИП был определен при $V_{\bar{x}}=0,05$. Для каждого АК n был определен по его значению для X09, поскольку по всем АК это значение n оказалось больше, чем для других ИП (см. таблицу 6, где N — общий объем АК, n — итоговый объем выборки).

Таблица 6. Объем выборки при $V_{\bar{x}}=0,05$ для эталонов АК

Параметр	Класс	N	$\bar{x}_i$	σ_i	$n(V_{\bar{x}}=0,05)$
X09	Ω_1 (Вяземский П. А.)	946	0,650	0,857	400
	Ω_2 (Гоголь Н. В.)	613	0,650	0,821	312
	Ω_3 (Дельвиг А. А.)	591	0,510	0,689	326
	Ω_4 (Пушкин А. С.)	1001	0,330	0,551	527
	Ω_5 (Сомов О. М.)	296	0,520	0,759	219

Выборки были осуществлены методом случайного отбора. АК Ω_5 (Сомов О. М.) и все РО были подвергнуты сплошному обследованию, так как их объемы невелики. Координатами эталонов АК и РО являются статистические характеристики распределений значений ИП ($\bar{x}_i$ и σ_i) для этих АК и РО. Координаты АК Ω_4 (Пушкин А. С.) и пяти атрибутированных ему РО (см. ниже) приведены в таблице 7 (где N — объем объекта в предложениях).

Таблица 7. Координаты эталонов некоторых РО и АК Ω_4 (Пушкин А. С.)

Объект	N	Параметр	$\bar{x}_i$	σ_i	Объект	N	Параметр	$\bar{x}_i$	σ_i
Ω_4	527	X09	0,455	0,701	T23	13	X09	0,615	0,625
		X18	3,833	3,242			X18	2,923	2,165
		X24	1,070	1,326			X24	0,769	0,799
		X32	1,499	1,118			X32	1,385	0,487
T04	8	X09	0,500	0,500	T29a	16	X09	0,250	0,433
		X18	3,375	1,728			X18	5,750	5,379
		X24	0,875	0,599			X24	1,188	0,882
		X32	1,250	0,433			X32	2,000	1,768
T15	9	X09	0,333	0,667	T43	85	X09	0,529	0,776
		X18	3,111	2,961			X18	3,776	2,763
		X24	1,222	1,397			X24	1,000	1,328
		X32	1,222	0,416			X32	1,482	0,876

В разделе 5 главы 2 описана процедура атрибуции.

Она складывается из работы двух алгоритмов распознавания — детерминированного (ДАР) и вероятностного (ВАР).

ДАР классифицирует РО в многомерном параметрическом пространстве (в нашем случае — 4-мерном). В качестве классифицирующей функции используется t -критерий Стьюдента с уровнем значимости $\alpha=0,05$ (формула (1)). Здесь нас интересуют значения критерия ниже 1,96, указывающие на отсутствие статистически значимых различий между РО и АК. Решение о принадлежности РО X_i одному из АК принимается лишь в том случае, если значение t -критерия ниже порогового во всех измерениях параметрического пространства, т. е. решающее правило имеет вид:

$$\begin{cases} X_i \in \Omega_A, t_{набл.,j}^{A,X_i} \leq t_\alpha, j = \overline{1, n}. \\ X_i \in \Omega_B, t_{набл.,j}^{B,X_i} \leq t_\alpha, j = \overline{1, n}. \end{cases} \quad (8)$$

Атрибуция на этапе работы ДАР носит итерационный характер. После определения принадлежности РО X_i некоторому АК объем и статистические характеристики последнего меняются: АК включает в себя распознанный объект. Координаты эталонов изменившихся классов мощностью j объектов вычисляются по формулам¹:

$$\bar{x}_i = \frac{(\bar{x}_1 n_1 + \bar{x}_2 n_2 + \dots + \bar{x}_j n_j)}{\sum n_j}, \sigma_i = \sqrt{\frac{\sum \sigma_j^2 n_j}{\sum n_j}}. \quad (9), (10)$$

Так, пять текстов (т. е. РО), атрибутированных А. С. Пушкину на первой итерации ДАР (Т04, Т15, Т23, Т29а, Т43), были включены в структуру АК (Ω_4), вследствие чего его координаты изменились: $n=658$; $\bar{x}_{X09}=0,462$, $\sigma_{X09}=0,702$; $\bar{x}_{X18}=3,839$, $\sigma_{X18}=3,218$; $\bar{x}_{X24}=1,058$, $\sigma_{X24}=1,303$; $\bar{x}_{X32}=1,500$, $\sigma_{X32}=1,089$.

ВАР приписывает вероятность (степень достоверности) каждому из возможных решений. Искомые вероятности определяются по формуле²:

$$P_{ji} = \frac{1}{d_{ji}} \left(\sum_{k=1}^n \frac{1}{d_{jk}} \right)^{-1}, \quad (11)$$

где d_{ji} — расстояние между j -м объектом и i -м классом, d_{jk} — расстояние между j -м объектом и k -м классом, n — число классов. В качестве функции расстояния между объектами используется взвешенное евклидово расстояние (EP)³:

$$d(a, b) = \left[\sum_{i=1}^n \omega_j (x_{aj} - x_{bj})^2 \right]^{0,5}, \quad (12)$$

где n — размерность евклидова пространства E_n , a и b — две точки в нем с координатами $a(x_{a1}, x_{a2}, \dots, x_{an})$, $b(x_{b1}, x_{b2}, \dots, x_{bn})$, ω_j — весовой коэффициент j -й переменной. Значение ω_j можно принять равным единице при условии, что все признаки выражены в общих единицах измерения. Выполнить это условие

¹ В поисках потерянного автора... — С. 13.

² Марусенко М. А. Указ. соч. — С. 58.

³ Марусенко М. А. Указ. соч. — С. 52.

позволяет стандартизация матрицы данных (Z), образованной средними значениями ИП для классов и объектов. Элементы стандартизованной матрицы X вычисляются по формуле¹:

$$x_{ij} = \frac{z_{ij} - \bar{z}_j}{\sigma_j}, i = \overline{1, N}, j = \overline{1, n}, \quad (13)$$

где z_{ij} — соответствующий элемент матрицы данных Z , $\bar{z}_j$ — среднее значение j -го параметра, σ_j — квадратичное отклонение j -го параметра. Фрагменты матриц взвешенных ЕР $d(X_i, \Omega_j)$ и вероятностей принадлежности объектов классам $P(X_i \in \Omega_j)$ приведены (для текстов, приписываемых только А. С. Пушкину) в таблице 8.

Таблица 8. Матрица взвешенных ЕР и вероятностей (X_i — объект)

X_i	$d(X_i, \Omega_j)$					$P(X_i \in \Omega_j)$				
	Ω_1^2	Ω_2^1	Ω_3	Ω_4^1	Ω_5	Ω_1^2	Ω_2^1	Ω_3	Ω_4^1	Ω_5
T05	1,837	2,099	0,884	0,634	1,206	0,119	0,105	0,248	0,346	0,182
T10	3,294	3,259	2,551	1,989	3,010	0,165	0,167	0,213	0,274	0,181
T11	2,481	2,261	2,388	2,245	2,574	0,192	0,211	0,200	0,212	0,185
T12	4,202	4,298	3,840	3,971	3,798	0,191	0,187	0,209	0,202	0,211
T16	3,074	3,129	2,102	1,459	2,510	0,148	0,145	0,216	0,311	0,181
T35	1,761	2,355	0,780	0,870	1,468	0,138	0,103	0,312	0,280	0,166
T37	2,906	3,813	3,508	4,133	3,669	0,245	0,187	0,203	0,172	0,194
T46	2,572	3,072	1,565	1,283	2,244	0,151	0,126	0,248	0,302	0,173

ВАР не чувствителен к свойствам распределений, определяющим условия применимости критерия Стьюдента (нормальность сопоставляемых распределений и равенство их дисперсий), и может быть использован в качестве теста результатов ДАР. В двух случаях тест показал результаты, отличные от итогов ДАР, и в трех — от итогов ВАР, основанного на итогах ДАР и учитывающего изменения координат эталонов классов (см. табл. 9).

Таблица 9. Тест результатов ДАР и ВАР посредством ВАР

X_i (объект)	$P(X_i \in \Omega_j)$					Тестируемый алгоритм	Прежний результат
	Ω_1^2	Ω_2^1	Ω_3	Ω_4^1	Ω_5		
T02	0,185	0,148	0,239	0,234	0,194	ВАР	Нет решения
T21	0,227	0,233	0,180	0,159	0,201	ВАР	Ω_1 (Вяземский П. А.)
T26	0,206	0,206	0,200	0,193	0,194	ВАР	Ω_3 (Дельвиг А. А.)
T29a	0,202	0,146	0,223	0,175	0,255	ДАР	Ω_4 (Пушкин А. С.)
T45	0,385	0,387	—	0,228	—	ДАР	Ω_1 (Вяземский П. А.)

В аспекте определения межтекстовых расстояний более существенна не сама вероятность, а отношение p_1/p_2 , или «оценка значимости вероятностей»².

¹ Марусенко М. А. Указ. соч. — С. 76.

² Якубайтис Т. А., Складевич А. Н. Вероятностная атрибуция типа текста по нескольким морфологическим признакам. — Рига: ИЭВТ, 1982. — С. 15.

Атрибутируя текст пяти авторам, получим значения вероятностей пяти рангов. Так, для текста Т05 первым рангом обладает вероятность принадлежности текста А. С. Пушкину (0,346), вторым — А. А. Дельвику (0,248) и т. д. Оценка значимости вероятности первого ранга по отношению к вероятности второго ранга — 1,395. Но применение всего набора оценок значимости вероятностей первого ранга затрудняет определение общего веса каждого решения ВАР. Поэтому целесообразно ограничиться использованием среднего значения набора оценок для каждого решения ВАР. Это значение можно назвать коэффициентом значимости принятого решения:

$$\bar{k}_{P(X_i \in \Omega_j)} = \frac{1}{n-1} \sum \frac{P(X_i \in \Omega_j)}{P(X_i \in \Omega_k)}, X_i \in \Omega_j, k = \overline{1, n}, k \neq j, \quad (14)$$

где $P(X_i \in \Omega_j)$ — вероятность первого ранга, n — число классов.

Общие результаты эксперимента по атрибуции сведены в таблице 10. Точность решений ДАР очень высока (значительно выше 95%), коэффициент их значимости не вычисляется, и на его месте в таблице значится «Д».

Таблица 10. Общие результаты атрибуции по итогам работы ДАР и ВАР

Текст	Автор	$\bar{k}_i$	Соавтор	Текст	Автор	$\bar{k}_i$	Соавтор
T01	Пушкин	2,333	Дельвиг	T24	Пушкин	1,694	Дельвиг
T02	—	—	—	T25	Сомов	3,179	—
T03	Гоголь	1,365	—	T26	Дельвиг	1,080	—
T04	Пушкин	Д	—	T27	Пушкин	1,748	Дельвиг
T05	Пушкин	2,375	Дельвиг	T28	Пушкин	6,109	—
T06	Пушкин	1,182	—	T29a/T29b	Пушкин	Д/1,431	—
T07	Пушкин	1,523	—	T30	Дельвиг	2,817	Сомов
T08	Сомов	2,517	Дельвиг	T31	Сомов	4,086	—
T09	Вяземский	1,462	—	T32	Сомов	2,607	Дельвиг
T10	Пушкин	1,525	—	T33	Дельвиг	1,294	—
T11	Пушкин	1,079	—	T34	Дельвиг	2,383	Пушкин
T12	Сомов	1,072	—	T35	Дельвиг	2,071	Пушкин
T13	Сомов	1,725	—	T36	Вяземский	Д	—
T14	Пушкин	1,400	—	T37	Вяземский	1,301	—
T15	Пушкин	Д	—	T38	Гоголь	Д	—
T16	Пушкин	1,851	Дельвиг	T39	Пушкин	1,520	—
T17	Дельвиг	1,511	Вяземский	T40	Пушкин	1,711	—
T18	Гоголь	1,272	—	T41	Вяземский	1,244	—
T19	Вяземский	Д	—	T42	Вяземский	1,698	—
T20	Гоголь	1,412	—	T43	Пушкин	Д	—
T21	Вяземский	1,311	—	T44	Пушкин	1,368	—
T22	Пушкин	1,784	Дельвиг	T45	Вяземский	Д	—
T23	Пушкин	Д	—	T46	Пушкин	1,840	Дельвиг

ЗАКЛЮЧЕНИЕ

В заключении сформулированы основные **выводы**.

1. За всю историю пушкиноведения А. С. Пушкину было приписано 64 публицистических текста, а 24 из них сегодня считаются написанными им. Авторство 46 из этих 64 текстов спорно. Филологические методы атрибуции не позволяют получить однозначное решение проблемы авторства этих текстов.

2. По результатам проведенного нами эксперимента из 46 текстов со спорным авторством А. С. Пушкину атрибутируется 22 текста. Проблема авторства одного текста не получила однозначного решения.

3. Подтверждаются предположения (М. П. Еремин, Н. Е. Мясоедова) о том, что роль А. С. Пушкина в издании ЛГ значительней, чем принято думать: 17 из упомянутых 22 текстов были опубликованы в ЛГ.

4. Опровергается принятое в сочинениях Пушкина представление о том, что он написал статью «Французская академия» (приписанную Н. О. Лернером) и анекдоты «В газете: *Le Furet...*» (приписанный В. В. Виноградовым) и «Острая шутка...» (приписанный С. А. Венгеровым и В. В. Виноградовым).

5. Подтверждается предположение о совместной работе писателей над некоторыми текстами. Возможность соавторства констатируется в отношении 13 текстов, а одним из соавторов всегда оказывается А. А. Дельвиг. Остальные соавторы — А. С. Пушкин, П. А. Вяземский и О. М. Сомов.

6. Вопрос о том, прекращал ли А. С. Пушкин сотрудничать с ЛГ в периоды своих отъездов из Петербурга, получает отрицательный ответ: ему атрибутируется пять текстов, напечатанных в ЛГ в эти периоды (Т22, Т23, Т24, Т28 и Т29).

7. Авторство ряда текстов, связываемых с именами П. А. Вяземского (пять статей), А. А. Дельвига (24 статьи) и О. М. Сомова (102 статьи), спорно.

8. С *определенными оговорками* использованную методику можно применять в задачах атрибуции текстов критически малых объемов.

ПУБЛИКАЦИИ ПО ТЕМЕ ИССЛЕДОВАНИЯ

Основные результаты исследования изложены в публикациях:

1. Хозяинов С. А. Проблема авторства публицистических произведений А. С. Пушкина периода 1830—1836 гг. // Материалы XXXV Междунар. филол. конф. — СПб.: Филол. фак. С.-Петерб. гос. ун-та, 2006. — Вып. 21: Секция математической лингвистики / отв. ред. А. С. Герд, Т. Г. Скребцова: в 2 ч. — Ч. 2. — С. 26—32 (0,4 п. л.).
2. Хозяинов С. А. Проблема авторства приписываемой А. С. Пушкину публицистики: содержание и методы изучения // Вестник Санкт-Петербургского университета. Серия 9. Филология. Востоковедение. Журналистика. — Вып. 3. — Ч. II. — 2007. — С. 274—280 (0,55 п. л.).
3. Хозяинов С. А. Некоторые проблемы и методы квантитативно-структурного изучения авторских стилей // Известия Российского государственного педагогического университета им. А. И. Герцена. — 2008 (30 апреля). — № 28 (63). — С. 378—384 (0,45 п. л.).

Хозяинов

Подписано в печать 13.05.2008
Формат 60х84 1/16. Бумага офсетная. Печать офсетная.
Усл. печ. л. 1,4. Тираж 100 экз.
Заказ № 828.

Отпечатано в ООО «Издательство "ЛЕМА"»
199004, Россия, Санкт-Петербург,
В.О., Средний пр., д.24, тел./факс: 323-67-74
e-mail: izd_lemma@mail.ru

