*На правах рукописи*

**Пименов Иван Сергеевич**

# ФОРМАЛЬНОЕ ВЫДЕЛЕНИЕ ПРИЁМОВ И СТРАТЕГИЙ АРГУМЕНТАЦИИ В ТЕКСТАХ НАУЧНОЙ КОММУНИКАЦИИ

Автореферат
диссертации на соискание ученой степени
кандидата филологических наук

Специальность 5.9.8 – Теоретическая, прикладная
и сравнительно-сопоставительная лингвистика
(филологические науки)

Новосибирск – 2025

Работа выполнена на кафедре фундаментальной и прикладной лингвистики федерального государственного автономного образовательного учреждения высшего образования **«Новосибирский национальный исследовательский государственный университет»**.

Научный руководитель – главный научный сотрудник сектора русского языка в Сибири федерального государственного бюджетного учреждения науки Института филологии Сибирского отделения Российской академии наук, доктор филологических наук, доцент **Тимофеева Мария Кирилловна**.

Научный консультант – старший научный сотрудник лаборатории машинного обучения и анализа данных федерального государственного бюджетного учреждения науки Института математики им. С. Л. Соболева Сибирского отделения Российской академии наук, кандидат физико-математических наук **Саломатина Наталья Васильевна**.

Официальные оппоненты:

заведующая лабораторией языковой конвергенции департамента филологии Санкт-Петербургской школы гуманитарных наук и искусств федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет «Высшая школа экономики», доктор филологических наук, профессор **Колмогорова Анастасия Владимировна**;

профессор кафедры образовательных технологий в филологии федерального государственного бюджетного образовательного учреждения высшего образования «Российский государственный педагогический университет им. А. И. Герцена», доктор филологических наук, профессор **Беляева Лариса Николаевна**.

Ведущая организация – федеральное государственное бюджетное образовательное учреждение высшего образования **«Санкт-Петербургский государственный университет»**.

Защита состоится **4 декабря** в **11:00** на заседании Диссертационного совета 24.1.142.01 по защите диссертаций на соискание ученой степени кандидата наук, на соискание ученой степени доктора наук по специальности 5.9.8 – Теоретическая, прикладная и сравнительно-сопоставительная лингвистика (филологические науки) на базе федерального государственного бюджетного учреждения науки Института филологии Сибирского отделения Российской академии наук.

Адрес: 630090, г. Новосибирск, ул. Николаева, д. 8.

Тел.: +7 (383) 330 15 18.

Электронный адрес: dissovet24.1.142.01@mail.ru.

С диссертацией можно ознакомиться в библиотеке федерального государственного бюджетного учреждения науки Института истории Сибирского отделения Российской академии наук (г. Новосибирск, ул. Николаева, д. 8).

Полный текст диссертации доступен по адресу:
https://www.philology.nsc.ru/dissovet24/disser/Pimenov/Dissertaciya_Pimenov.pdf

Автореферат разослан: **8 октября 2025 г.**

Ученый секретарь диссертационного совета,
доктор филологических наук, профессор Н. Б. Кошкарева

## Общая характеристика работы

**Актуальность проведённого исследования.** Диссертационное исследование посвящено проблеме формального анализа структур аргументации в текстах научной коммуникации (преимущественно коротких научных статьях) на русском языке методами компьютерной обработки. Такие структуры рассматриваются на уровне приёмов и стратегий аргументации. Приёмы соответствуют организованным, повторяющимся в разных текстах структурам из одной или более типовых моделей рассуждения (абстрактных логических схем, регулирующих построение аргументационных связей между высказываемыми утверждениями в процессе доказательства некоторой идеи). Совокупность приёмов, применяемых на уровне целостного полного текста для доказательства его ключевой идеи, образует стратегию аргументации, реализуемую автором этого текста. Полезно сразу отметить, что термин «доказательство» понимается в работе в широком смысле как поддержка одного тезиса другим в аргументационной структуре текста. Расширенное толкование связано с моделированием аргументации в терминах аргументационных схем Д. Уолтона, где разные схемы (модели рассуждения) соответствуют разным видам обоснования в классическом понимании, от доказательств в строгом определении до логических ошибок (таких как ad hominem).

Формальный анализ прагматических языковых явлений, к числу которых относится аргументация, является фундаментальной лингвистической проблемой. Сложность аргументации как явления и теоретическая значимость её изучения иллюстрируются многоаспектностью аргументационных исследований в рамках различных дисциплин: как собственно лингвистики, теоретической и прикладной, так и риторики [Perelman, Olbrechts-Tyteca, 1969], философии [Van Eemeren, Grootendorst, 2004], формальной логики [Hintikka, 1989]. Задача формальной обработки аргументационных структур предполагает использование методов компьютерной лингвистики, в частности, направления Argument Mining, извлечения аргументов из текстов на естественном языке [Lawrence, Reed, 2019]. Однако ввиду сложности аргументации как языкового явления исследования по её компьютерной обработке активизировались сравнительно недавно. При этом такие работы обращены преимущественно к текстам на английском и других зарубежных языках [Lawrence, Reed, 2014].

Таким образом, актуальность диссертационного исследования проявляется в двух аспектах: теоретическом и методологическом. **Теоретическая актуальность** обуславливается важностью изучения формальной организации аргументации как явления прагматического языкового уровня в текстах на русском языке, в частности, в научных статьях (направленных на доказательство исследовательских выводов). **Методологическая актуальность** работы основывается на необходимости интеграции традиционных лингвистических подходов к изучению аргументации и методов компьютерной обработки текстов для формального анализа структур доказательства и рассуждения с учётом

языковой и жанровой специфики текстов научной коммуникации на русском языке. При этом лингвистическая сложность аргументации как прагматического явления обуславливает потребность в развитии и адаптации методов компьютерной обработки текстов, преимущественно применяемых для анализа более простых единиц иных языковых уровней (морфологического, лексического, синтаксического). Соответственно, моделирование и формальный анализ аргументационных структур обеспечивают возможность для расширения методологического аппарата лингвистики в целом.

**Степень разработанности проблемы исследования.** Ввиду сложности аргументации как языкового явления исследования по её компьютерной обработке (Argument Mining) образовали отдельное целостное направление с устоявшимся методологическим аппаратом сравнительно недавно (первая специализированная конференция по Argument Mining состоялась в 2014 г. [Lawrence, Reed, 2014]). Исследования в этом направлении обращены в основном к текстам на английском языке и таким жанрам, как юридические диспуты [Walker, Vazirova, Sanford, 2014], новостные статьи [Wachsmuth et al., 2018], политические дебаты [Lindahl, Borin, Rouces, 2019], онлайн-дискуссии [Hidey, McKeown, 2018]. Тем не менее отдельные работы посвящены анализу аргументации в научных статьях, примером чего является статья [Green, 2015].

Исследования по компьютерной обработке аргументов в текстах на русском языке представлены в ограниченном количестве из-за недостаточного объёма доступных корпусов, как указано в [Котельников, 2018]. Впрочем, значимость темы обуславливает активную разработку аргументационных корпусов русского языка в последнее время: посредством перевода иноязычных корпусов [Fishcheva, Kotelnikov, 2019], через разметку научно-популярных статей [Сидорова и др., 2020] либо сообщений в социальных сетях [Kotelnikov et al., 2022]. Хотя научные статьи на русском языке ранее не становились объектом компьютерного анализа аргументации, известны исследования научно-популярных текстов, например, [Ким, Ильина, 2020; Саломатина и др., 2020; Zagorulko et al., 2020; Ильина, 2023].

**Объект диссертационного исследования** – аргументационные структуры в текстах научной коммуникации на русском языке (преимущественно в научных статьях). **Предметом исследования** являются средства формальной организации элементов аргументации (утверждений, объединяющих их связей, типовых моделей рассуждения в основе этих связей) при доказательстве основной идеи целостного текста, обеспечивающей его смысловую связность.

Основная **гипотеза исследования**: формальный анализ аргументационных структур посредством интеграции сведений об их лингвистическом выражении (языковом оформлении утверждений) и их абстрактной логической организации (на уровне типовых моделей рассуждения в основе связей между утверждениями) позволяет выявлять составные приёмы и стратегии аргументации, характеризующие специфику доказательств в отдельных целостных текстах. На Рис. 1 показан пример искомого составного приёма:

прямоугольные блоки содержат утверждения из конкретных текстов, овальные указывают на типовые абстрактные модели (схемы аргументации, элементарные приёмы), реализуемые при обосновании одних утверждений другими. Содержание утверждений различается между разными текстами (слева и справа), тогда как модели в связях между утверждениями совпадают. Одинаковы и их конфигурации в образуемой структуре (составном приёме) из трёх элементарных приёмов, и эта структура соответствует составному приёму аргументации.

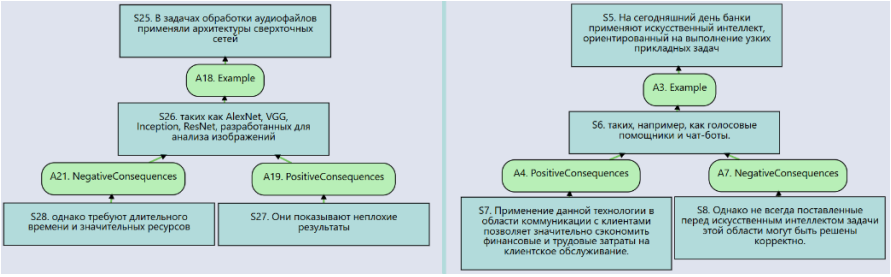


Рисунок 1 – Примеры реализации составного приёма аргументации

В соответствии с основной гипотезой **цель работы** может быть сформулирована следующим образом: разработать и экспериментально проверить систему методов формального анализа аргументации в текстах научной коммуникации для распознавания и характеризации специфики доказательств в отдельных текстах на уровне как их языкового оформления, так и структурной организации. Под экспериментом в работе понимается практическое использование разработанных методов для компьютерной обработки текстовых данных из используемого корпуса с последующей оценкой эффективности методов на уровне как количественных показателей, так и содержательной интерпретации результатов их применения.

Достижение поставленной цели предполагает решение следующих **задач**:

1) определить теоретические основы для формального представления аргументации с учётом как традиционных лингвистических подходов к её анализу, так и специфики методов компьютерной обработки текстов;

2) подготовить корпус текстов (научных статей на русском языке) с многоаспектной аргументационной разметкой (согласно обозначенным теоретическим принципам, причём под многоаспектностью понимается характеризация аргументации на уровнях утверждений, связей между ними, типовых семантических моделей в основе этих связей) для подготовки и проверки автоматических методов обработки аргументации, выраженной на естественном языке. Создание корпуса предполагает разработку методологии многоаспектной разметки аргументации в научных статьях, которая повысит согласованность разметки текстов разными аннотаторами;

3) реализовать набор методов компьютерного анализа аргументационных структур, где отдельные методы предназначены для решения следующих подзадач:

– распознавания предложений, содержащих аргументацию, связей между этими предложениями и отдельных частотных элементарных приёмов (экспертная разметка используется для машинного обучения классификаторов, применяемых для распознавания, и автоматической проверки качества их работы);

– выявления элементарных и составных приёмов аргументации из размеченных текстов с учётом их жанровой специфики в многожанровой коллекции;

– обобщения цепочек рассуждения согласно функциональной близости образующих их моделей рассуждения;

– оценки аргументов по выраженности методов убеждения;

– распознавания элементов методов убеждения (пафос, этос) в аргументах из неразмеченных текстов;

– выявления текстов со сходной организацией аргументации и ранжирования их по сложности рассуждений путём интеграции сведений об их аргументационной специфике во всех отмеченных выше аспектах;

– жанровой классификации текстов на основе сведений о реализованных в них приёмах аргументации;

4) экспериментально проверить реализованные методы на подготовленном корпусе текстов, интерпретировать результаты и обозначить особенности построения аргументации в текстах научной коммуникации на русском языке.

**Материалом исследования** выступает корпус из 11495 тезисов и 10123 аргументов, образующих структуры аргументации для 100 научных статей с ручной разметкой аргументации тремя экспертами (для каждого текста подготовлено по две версии разметки независимо двумя экспертами). Размеченные статьи принадлежат направлениям лингвистики и информационных технологий (по 50 статей обеих тематик), характеризуются объёмом от 800 до 2000 слов (без учёта аннотаций и библиографических списков). Публикация анализируемых статей в научных изданиях указывает на достижение ими академического уровня, достаточного для прохождения рецензирования. Разметка корпуса проведена по ходу исследования, в связи с чем на отдельных этапах задействованы ранние версии корпуса меньшего объёма. На отдельных этапах исследования (при анализе аргументации в текстах разных жанров) привлекаются данные из параллельно размечавшихся корпусов научных новостей и научно-популярных текстов, количественные характеристики которых даны в таблице 1.

Таблица 1 – Количественные характеристики корпусов разных жанров

| Количество единиц | Научные статьи (из библиотеки «КиберЛенинка») | Научные новости (портал «Поиск») | Короткие научно-популярные тексты (из различных журналов) | Научно-популярные тексты (Habr) |
|---|---|---|---|---|
| Число текстов | 100 (с двойной разметкой) | 28 | 95 | 20 |
| Число утверждений | 11495 | 859 | 2005 | 1681 |
| Число аргументов | 10123 | 758 | 1386 | 1415 |

Разметка аргументации проведена в соответствии со стандартом AIF [Rahwan, Reed, 2009] и классификацией типовых моделей рассуждения (схем аргументации) Д. Уолтона [Walton, Reed, Macagno, 2008]. Разметка проведена на трёх уровнях аргументации:

– выявление утверждений (высказываний с аргументацией, принадлежащих целостной структуре рассуждений в тексте);

– распознавание связей между утверждениями (что включает определение их ролей как посылок либо заключений в составе аргументов);

– указание типовых моделей рассуждения из заданного набора для связей (с помощью моделей, также называемых схемами, уточняется семантика перехода от посылок к заключению).

Корпусы с разметкой доступны на платформе ArgNetBank Studio (https://uniserv.iis.nsk.su/arg/corpora), аннотирование текстов осуществлено с помощью инструментов этой платформы.

Для оценки надёжности разметки проведён подсчёт количественных коэффициентов согласия аннотаторов на разных этапах создания корпуса. Использован разработанный в ходе исследования специализированный коэффициент аргументационного согласия (КАС), позволяющий анализировать расхождения в разметке аргументации на всех её уровнях, а также два универсальных коэффициента с учётом случайного согласия между разметчиками, каппа Коэна и альфа Криппендорфа. Значения коэффициентов подсчитаны для каждого уровня разметки (выявление утверждений, связей, схем, в том числе на уровне групп функционально близких схем). В таблице 2 приведены значения двух коэффициентов согласия для рабочей версии корпуса научных статей (каппа Коэна не указана, так как подсчитывается по парам аннотаторов).

Приведённые показатели соответствуют нижним границам оценки согласия разметчиков: случаи формальных расхождений аннотаторов при их фактическом согласии учитываются как несоответствия в разметке (например, случаи разной конфигурации связей, когда оба разметчика указывают связи по одинаковым моделям рассуждения между одними и теми же тезисами, однако один аннотатор строит последовательные связи, а другой – параллельные,

оцениваются компьютером как расхождения в разметке, как и сложные случаи в сегментации утверждений).

Таблица 2 – Коэффициенты согласия аннотаторов в рабочей версии корпуса

| Коэффициент | Утверждения | Связи | Схемы | Функциональные группы схем |
|---|---|---|---|---|
| КАС | 0.87 | 0.54 | 0.52 | 0.71 |
| Альфа Криппендорфа | 0.51 | 0.66 | 0.43 | 0.55 |

Для решения поставленных задач на базе указанных материалов исследования применены **методы** разных групп:

1) анализа аргументации на основе дискурсивного подхода (выявление аргументов с учётом, во-первых, лексико-грамматических маркеров и особых синтаксических конструкций для представления рассуждений, во-вторых, абстрактных логических моделей их организации, в-третьих, контекста реализации аргументов в целостной структуре рассуждений полного текста);

2) корпусной лингвистики (создание аргументационных корпусов, методики разметки, разметка текстов, оценка согласия аннотаторов);

3) обработки естественного языка (морфологический анализ, построение поисковых шаблонов для маркеров аргументации);

4) машинного обучения (классификация и кластеризация текстов и их элементов, извлечение аргументации).

**Теоретическая (источниковая) база** исследования включает работы двух групп. Одну представляют труды по исследованиям и моделированию аргументации с позиций различных дисциплин: в первую очередь, на основе лингвистических подходов, но также и в рамках иных направлений (философии, риторики, формальной логики). Другую группу составляют работы по компьютерной обработке текстов на естественном языке (как в рамках специализированных исследований аргументации, так и для общих смежных задач, таких как автоматическая классификация текстов) и фундаментальные труды по анализу данных разных типов (например, работы с описанием алгоритмов кластеризации, применимых как к текстовым, так и иным данным).

Соответственно, **новизна** диссертационного исследования заключается в следующем:

1) Впервые применены методы формального аргументационного анализа к текстам научной коммуникации на русском языке. Даны строгие определения приема и стратегии аргументации, позволившие построить формальное представление аргументационной структуры текста, коллекции текстов, набора коллекций.

2) Впервые аргументация в текстах на русском языке обрабатывается компьютерными методами на уровне глубинной аргументационной семантики: через многоаспектный анализ элементарных и составных приёмов (в том числе с учётом их функционального сходства), методов убеждения (средств

дополнения фактического содержания рассуждений апелляциями к авторитету и эмоциональным воздействием), совокупных особенностей последовательного обоснования тезисов в отдельных целостных текстах. Предложена функциональная классификация аргументов по структурно-семантическим особенностям их реализации в научных статьях, указаны четыре обобщённых типа таких аргументов («через детализацию», «через анализ причинно-следственный связей», «через рассуждения практического плана», «от авторитета»). Указана специфика реализации методов убеждения в научных текстах, подготовлен словарь их маркеров (в виде многокомпонентных лексико-грамматических шаблонов на основе регулярных выражений).

3) В рамках исследования подготовлен и размечен аргументационный корпус научных статей на русском языке, содержащий 11495 аргументов и 10123 утверждений (тезисов); разметка текстов заключается в трёхуровневом моделировании их аргументационной структуры согласно стандарту Argument Interchange Format и классификации аргументационных схем Д. Уолтона. В процессе создания корпуса разработана методология многоаспектной разметки аргументации в текстах научных статей, которая регулирует действия аннотаторов при разметке. Надёжность аргументационной разметки подтверждается коэффициентами согласия аннотаторов, подсчитанными на всех уровнях аннотирования аргументации. Для анализа расхождений в аргументационном моделировании на всех его уровнях разработан специализированный коэффициент аргументационного соответствия.

4) Разработан программный комплекс для многоаспектной компьютерной обработки аргументации в текстах на русском языке; комплекс образован программными модулями для различных отдельных задач аргументационного анализа: распознавание тезисов в текстах, связей между смежными тезисами и отдельных частотных типов аргументов (для автоматизации экспертной разметки); характеризация аргументов в тексте по методам убеждения (этос, пафос); выявление составных приёмов аргументации, используемых авторами текстов научной коммуникации, из графов, построенных аннотаторами; кластеризация текстов научной коммуникации по сходству аргументационных структур и ранжирование групп текстов по формальной сложности аргументации (которая подсчитывается через интеграцию разноаспектных показателей аргументационной специфики текстов: структурной организации рассуждений, реализуемых приёмов аргументации, интенсивности выражения отдельных методов убеждения); жанровая классификация текстов на основе сведений о реализованных в них приёмах аргументации.

Таким образом, **теоретическая значимость** исследования обусловлена выявлением формальных особенностей построения аргументации в текстах научной коммуникации на русском языке. Определены тематические различия в использовании приёмов аргументации в работах по лингвистике и по информационным технологиям. Предложена функциональная классификация аргументов по структурно-семантическим особенностям их реализации в научных

статьях, указаны четыре обобщённых типа таких аргументов («через детализацию», «через анализ причинно-следственный связей», «через рассуждения практического плана», «от авторитета»). Указана специфика реализации методов убеждения в научных текстах, подготовлен словарь их маркеров (в виде многокомпонентных лексико-грамматических шаблонов на основе регулярных выражений). Результаты диссертационного исследования дополняют теоретическую базу для иных исследований в областях теории аргументации (в рамках лингвистического и формального подходов) и лингвистической прагматики, представляют новые сведения об организации текстов (в первую очередь научных статей), их целостности и связности на уровне аргументационной структуры.

**Практическая значимость** диссертационного исследования заключается в разработке и реализации методов компьютерной обработки аргументации в текстах на русском языке. На основе размеченного корпуса экспериментально оценена эффективность различных алгоритмов для отдельных задач аргументационного анализа, изучены и типизированы ошибки на этапах компьютерного распознавания элементов аргументации (тезисов, аргументов по разным методам убеждения). Корпус с многоуровневой аргументационной разметкой научных статей может использоваться в иных исследованиях по формальному анализу аргументации (в частности, для подготовки методов машинного обучения). Иным аспектом практической значимости является учебно-методический: материалы исследования могут применяться в процессе педагогической деятельности в областях компьютерной лингвистики, лингвистики текста, риторики, академического письма (для развития у обучающихся компетенций в области анализа научной аргументации, критического восприятия информации, для практики навыков написания научных текстов с ясным представлением и обоснованием выдвигаемых тезисов).

**Список используемой литературы** содержит 153 источника, 98 из которых представлены на английском языке.

**Положения, выносимые на защиту**, сформулированы ниже.

1. Формальное представление аргументации позволяет охарактеризовать аргументационную специфику целостного текста в разных аспектах: взаимосвязи языковых средств выражения аргументов и абстрактных схем в основе их связей, использовании приёмов аргументации и методов убеждения. Особенности аргументационной организации отдельных текстов проявляются при сопоставительном анализе их формальных моделей.

2. Авторы разных текстов научной коммуникации используют повторяющиеся между текстами приёмы аргументации, которые образованы небольшим числом типовых моделей рассуждения.

3. Представление и логический анализ фактов в научных статьях при обосновании исследовательских результатов периодически дополняется непрямым эмоциональным воздействием через использование оценочной лексики и

модальных конструкций для привлечения внимания читателей и акцентирования отдельных аспектов работы.

4. Элементы аргументационных структур могут распознаваться методами компьютерной обработки текстов на естественном языке с достижением приемлемых показателей качества, что позволяет автоматически анализировать и неразмеченные тексты.

5. Использование приёмов аргументации и методов убеждения (общих способов аргументационного воздействия на читателя) в текстах научной коммуникации зависит от их жанровой специфики и предметной области. Представление текстов на уровне реализуемых в них приёмов обеспечивает возможность их надёжного разграничения как по жанру и тематике, так и по более специфичным аспектам содержания.

6. Компьютерный анализ характеристик аргументационной структуры в разных аспектах позволяет выявлять тексты научной коммуникации с похожей организацией рассуждений и указывает на взаимосвязь показателей разных уровней аргументации.

**Достоверность результатов исследования** обуславливается подготовкой теоретических и методологических оснований для его проведения на базе фундаментальных работ в областях традиционных исследований аргументации, компьютерной обработки текстов на естественном языке и формального аргументационного анализа. Все приводимые выводы подтверждаются экспериментами, которые проведены в соответствии с общепринятыми стандартами. Для корпуса с аргументационной разметкой, используемого в экспериментах, подсчитаны формальные оценки согласия аннотаторов на разных уровнях моделирования аргументации. Другим свидетельством надёжности результатов является их публикация в ряде рецензируемых научных изданий. Результаты исследования также использованы в работе по грантам, финансируемым Российским научным фондом в рамках проектов № 23-21-00325 и № 23-11-00261.

**Апробация результатов исследования** включала их представление в рамках докладов на следующих конференциях и научных семинарах: Международная научная конференция «Корпусная лингвистика 2021 (2023)» (Corpora-21, Corpora-23) (Санкт-Петербург, 2021 г. и 2023 г.); Международная конференция «Марчуковские научные чтения 2021 (2022, 2024)» (Новосибирск, 2021 г., 2022 г., 2024 г.); 2022 (2024) IEEE 23rd (24th, 25th, 26th) International Conference of Young Professionals in Electron Devices and Materials (EDM) (республика Алтай, 2022 г., 2023 г., 2024 г. и 2025 г.); 2022 International Symposium on Knowledge, Ontology, and Theory (KNOTH) (Новосибирск, 2022 г.); XXV Internation Conference Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2023) (Москва, 2023 г.); Международная научно-практическая конференция «Обработка естественного языка (NLP) для лингвистики, IT и бизнеса» (Томск, 2025 г.); регулярный объединённый семинар Института систем информатики Сибирского отделения Российской академии наук и кафедры программирования Новосибирского государственного университета

«Интеллектуальные системы и системное программирование» (Новосибирск, 2021 г.); международный научно-практический семинар «Экспериментальные исследования языка и речи: Ресурсные базы и технологии» (E-SoLaS-2022) (Томск, 2022 г.); регулярный объединённый семинар Федерального исследовательского центра информационных и вычислительных технологий и Новосибирского государственного университета «Информационные технологии в задачах филологии и компьютерной лингвистики» (Новосибирск, 2022 г.).

**Основное содержание работы** отражено в 20 публикациях, из которых 4 статьи изданы в журналах из перечня российских рецензируемых научных журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук (список ВАК), а ещё 9 работ опубликованы в изданиях, индексируемых в базе данных Scopus.

**Личный вклад соискателя** может быть охарактеризован в следующих аспектах:

1) определения темы исследования, формулировки его целей и задач (как во всём масштабе работы, так и на отдельных этапах), изучении и выборе методологического аппарата, постановки и проверки гипотез;

2) подготовки аргументационного корпуса научных статей (отбор и разметка текстов), определения теоретических оснований и принципов для формального представления структур аргументации, составления методики разметки;

3) разработки и программной реализации методов автоматического анализа аргументации, их применения для формальной обработки структур рассуждения в текстах корпуса, интерпретации полученных данных;

4) подготовки публикаций с представлением результатов исследований, их презентации на международных конференциях и научных семинарах;

5) написании текста диссертационного исследования, формулировании выводов и положений, выносимых на защиту.

**Структура диссертации** образована введением, четырьмя главами (двумя теоретическими и двумя экспериментальными), заключением, библиографией (которая включает 153 наименования) и 6 приложениями. Работа иллюстрирована 5 рисунками, содержит 29 таблиц.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

**Введение** диссертации содержит постановку проблемы формального анализа приёмов и стратегий аргументации в научных статьях. Обосновывается актуальность исследования в двух аспектах (теоретическом и методологическом), обозначаются его объект и предмет, цель и решаемые задачи. Указываются материал и методы исследования, характеризуется его новизна относительно кратко описываемого состояния предметной области. Излагается теоретическая и практическая значимость результатов исследования.

## Глава 1. Теоретические основы представления аргументации

Первая глава диссертации обращена к задаче определения теоретических оснований для моделирования и компьютерного анализа аргументации, выражаемой средствами естественного языка. Отдельно акцентируется значимость анализа аргументации в фундаментальном лингвистическом контексте. Рассматриваются две группы подходов к исследованиям аргументации: традиционные лингвистические и формальные.

В рамках первой группы подходов проводится сопоставление различных лингвистических взглядов на изучение аргументации. Обзор подходов осуществляется на основе классификации представлений о соотношении системы языка и аргументационных структур, предложенной Ж.-К. Анскомбром и О. Дюкро [Anscombre, Ducrot, 1989]. Анализируется разработанная этими авторами теория «радикального аргументативизма», согласно которой аргументационный компонент проявляется не столько в конкретных коммуникативных ситуациях, сколько в системе языка на уровне лексических значений слов. Оценивается применимость данной теории для моделирования аргументации на уровне целостных текстов, рассматривается её расширение за авторством М. Карел в виде теории семантических блоков [Carel, 1995]. Обзор аргументационных исследований в терминах лингвистической семантики дополняется рассмотрением лингвокультурологического подхода Э. Риготти и А. Роччи [Rigotti, Rocci, 2005], согласно которому аргументационный потенциал лексической единицы зависит от культурной важности обозначаемого ею понятия.

В противопоставлении с лексико-семантической традицией (в которой аргументационные значения рассматриваются как явления внутри системы языка, опосредованно от их выражения в процессе коммуникации) обозначается дискурсивный подход к моделированию аргументации. Указывается происхождение этого подхода из концепции естественной (натуральной) логики Ж.-Б. Гриза [Plantin, 2003], разработанной в оппозиции к формальной логике. Абстрактность формальной логики («отсутствие субстанции») уточняется через обращение к лингвистическим формам (предложениям и текстам), посредством которых логические схемы выражаются при повседневном применении языка. Согласно Ж.-Б. Гризу, аргументационный компонент выражается во всех когнитивных и лингвистических операциях при организации связных целостных текстов, а соответственно, и аргументационные структуры проявляются именно при употреблении языка в реальных коммуникативных ситуациях. Развитием естественной логики выступает дискурсивный подход к моделированию аргументации, основные положения которого обозначены в работах Р. Амосси [R. Amossy, 2009]. Согласно дискурсивному подходу, гипотетический аргументационный потенциал отдельных языковых выражений (способность отсылать к другим на основе смысловой сочетаемости в процессе рассуждения или доказательства) ограничивается условиями общения

(жанром дискурса, обсуждаемой темой, статусом говорящего и слушающего, их заявляемыми и скрываемыми намерениями, интертекстуальным влиянием иных дискурсов и т. п.). Так как свободно варьирующиеся условия общения не могут быть зафиксированы в системе языка (но при этом влияют на представление говорящими своих точек зрения наравне с системными значениями выбираемых ими лексических единиц), то структуры аргументации могут исследоваться только на уровне целостных текстов (с учётом специфики их создания как продуктов коммуникации).

Кроме того, как указывает Р. Амосси, анализ аргументационных структур на уровне их текстовой реализации позволяет реконструировать формальную организацию применяемых доводов. Такое восстановление абстрактных схем, задающих переходы от одних утверждений к другим при раскрытии позиции говорящего, оказывается особенно важным при анализе связности текста как лингвистического целого. Однако и формальное строение аргументационных структур не может рассматриваться изолированно от языковых средств, их выражающих: согласно Р. Амосси, именно сочетание обоих типов конструкций (логических и лингвистических) обеспечивает представление в тексте мнений и отношений, характеризующих описание ситуации.

В контексте традиционных лингвистических подходов рассматривается и развитие аргументационной теории в отечественной лингвистике. Анализируются труды по революционному языку в агитации и политической полемике, опубликованные в 20-е гг. XX в. (А. М. Селищев, Л. П. Якубинский), систематизирующая работа А. Н. Баранова в рамках когнитивной парадигмы, современные исследования Е. Р. Иоанесян в развитии теории Ж.-К. Анскомбра и О. Дюкро. Указывается характеризующее отечественные труды гармоничное сочетание самобытной оригинальности и одновременного внимания к международному опыту, а также их специализация в рамках фундаментальной лингвистики, что затрудняет опору на эти работы в компьютерной обработке аргументации.

Таким образом, возможно разграничение двух указанных подходов к аргументационным исследованиям (с лингвистических позиций) в противопоставлении двух общетеоретических парадигм современной лингвистики: структуралистской (системно-семантический подход) и коммуникативной (дискурсивный). Поскольку диссертационное исследование ориентировано на анализ аргументации в научных статьях, теоретической основой выбран дискурсивный подход.

Дискурсивный подход предполагает представление аргументационных структур на уровне целостных текстов. Однако такое представление может выстраиваться посредством различных формальных моделей. Во второй части главы анализируется применимость к задаче трёх наиболее известных аргументационных моделей:

1) шестикомпонентной модели аргумента С. Тулмина [Toulmin, 2003];

2) адаптированной к аргументационным отношениям теории риторических структур (ТРС) У. Манна и С. Томпсон [Mann, Thompson, 1988];

3) модели Argument Interchange Format [Rahwan, Reed, 2009] в сочетании с компендиумом аргументационных схем Д. Уолтона [Walton, Reed, Macagno, 2008].

При рассмотрении двух первых моделей выявляется их ограниченная применимость для моделирования аргументационных структур в целостных текстах. В частности, модель С. Тулмина ориентирована на представление отдельных аргументов как самостоятельных конструкций (с подробным разграничением ролей их отдельных компонентов), но плохо подходит для указания связей между аргументами в аргументационной структуре полного текста [Peldzus, Stede, 2003]. Применение же ТРС для представления аргументационных связей затрудняется рядом различий между риторической и аргументационной структурами текста [Kononenko, Sidorova, Akhmadeeva, 2020]: так, риторическая структура разработана для моделирования линейной связности текста, тогда как аргументационная не обязательно соответствует последовательности выражения тезисов: одно утверждение может использоваться для поддержки двух заключений, не вложенных одно в другое (такая структура не может быть представлена посредством базовой версии ТРС, что выступает её серьёзным ограничением).

В свою очередь, Argument Interchange Format основан на принципах, близких к ключевым положениям дискурсивного подхода. AIF разработан для моделирования аргументации в полных текстах, что заключается в представлении текста в виде **аргументационного графа** – формальной структуры с двумя типами элементов (вершинами, передающими некое содержание, и рёбрами, указывающими на связи между вершинами). Аргументационные графы образуются вершинами двух типов: информационными, которые соответствуют утверждениям в тексте, и вершинами-схемами, которые чередуются с информационными и указывают на абстрактные типовые модели рассуждения (аргументационные схемы), реализуемые при переходе от одного утверждения к другому в процессе аргументации. **Аргумент** в терминах AIF определяется как набор утверждений (тезисов), одно из которых (заключение) обосновывается или оспаривается другими (посылками в случае обоснования) посредством реализации некоторой абстрактной модели рассуждения (например, через построение доказательства «от авторитета», «от примера», «от положительных / негативных последствий», «от причины к следствию», «от корреляции к причине» и т. д.). Пересечения аргументов на уровне отдельных утверждений (один тезис может быть посылкой в одном аргументе и заключением в другом либо посылкой или заключением сразу в нескольких аргументах) обеспечивают их объединение в аргументационной структуре полного текста. Для указания моделей рассуждения предлагается использовать классификацию, в качестве которой авторы AIF рекомендуют компендиум Д. Уолтона.

Данный компендиум не является специализированным и содержит в целях прикладной применимости схемы убеждения разных типов: свойственные как повседневному общению, так и, например, юридическому дискурсу, а также научным текстам различных предметных областей.

Таким образом, использование стандарта AIF для моделирования аргументации позволяет анализировать и лингвистическое оформление утверждений языковыми средствами, и их формальную структурную организацию на уровне абстрактных (логических) аргументационных схем. Два данных аспекта аргументации рассматриваются в их тесной взаимосвязи, что сближает стандарт AIF с теоретическими принципами дискурсивного подхода (о необходимости учитывать и языковое выражение аргументов, и их формальную организацию). Соответственно, итогом первой главы становится обозначение теоретических оснований для экспериментального анализа аргументации: моделирование аргументации производится согласно дискурсивному подходу с использованием стандарта AIF для формального представления аргументационных структур.

## Глава 2. Методы анализа аргументационных стратегий

Во второй главе диссертационного исследования представляются методы формального анализа аргументации в рамках последующих экспериментов. Описание методов включает как обзор существующих подходов к решению отдельных задач в области компьютерной обработки аргументации, так и уточнение алгоритмов, реализованных в диссертационном исследовании.

В частности, во второй главе обозначаются принципы разметки аргументации, дополняющие базовые правила AIF с учётом жанровой специфики доказательств в научных статьях. Эти принципы положены в основу разработанной методики разметки и применены в подготовке аргументационного корпуса, обрабатываемого в экспериментальной части диссертационного исследования. В контексте разметки аргументации описываются коэффициенты согласия аннотаторов, подсчитываемые для оценки надёжности разметки на трёх уровнях аннотирования (утверждения, связи, схемы): универсальные коэффициенты с учётом случайного совпадения между разметчиками каппа Коэна и альфа Криппендорфа (отдельно указывается необходимая для их применения предобработка размеченных данных), а также разработанный в ходе исследования специализированный коэффициент аргументационного соответствия (учитывающий специфику многоуровневой разметки аргументации). Рассматриваются методы частичной автоматизации аргументационной разметки, где основное внимание уделяется компьютерному выявлению тезисов через обучение классификаторов тремя алгоритмами МО (полиномиальный байесовский классификатор, MNB; метод опорных векторов, SVM; базовую нейронную сеть, многослойный перцептрон, MLP) и векторное представление тезисов через статистическую фильтрацию незначимых для аргументации лемм

(посредством компьютерного морфологического анализа, подсчёта мер TF-IDF, Mutual Information и критериев $\chi^2$ и Variance на основе встречаемости различных лемм в аргументативных и неаргументативных утверждениях). По аналогии с выявлением тезисов, к задаче классификации сводится и задача построения связей между тезисами и распознавания элементарных приёмов аргументации.

Во второй главе представлены также методы выявления составных приёмов аргументации, предложено их определение как применяемые авторами разных текстов типовые структуры из нескольких элементарных приёмов (соответствующих отдельным аргументационным схемам). Подчёркивается значимость знаний о таких структурах для вычислительного синтеза аргументов, их идентификации, классификации текстов по специфике аргументации. Распознавание составных приёмов аргументации достигается методами автоматического извлечения изоморфных подграфов из аргументационных графов (изоморфные подграфы соответствуют идентичным подструктурам в полных графах). Приводится формальное описание процедуры построения частотного спектра приёмов аргументации на материале многожанровых коллекций текстов.

Две другие группы методов ориентированы на переход от анализа приёмов к стратегиям аргументации. Предложено определение стратегии аргументации как конструкции высшего уровня, образуемой сочетанием отдельных приёмов аргументации и общих структурно-содержательных особенностей организации рассуждений на протяжении целостных текстов. Описываются методы компьютерной оценки выраженности методов убеждения (трёх общих аспектов аргументационного воздействия: через рациональные доказательства, влияние на эмоции, апелляции к авторитетам) посредством анализа аргументационных схем и лексико-грамматических маркеров на уровне отдельных аргументов и полных текстов (что позволяет сопоставлять различные тексты по соотношению в них методов убеждения). Формулируются критерии принадлежности лексических единиц к маркерам этоса и пафоса, описывается разработанная структура лексико-грамматических шаблонов, выстраиваемых на основе выявленных маркеров, описывается составление словарей этоса и пафоса. Представлены методы выявления текстов со сходными аргументационными структурами и их ранжирования по формальной сложности аргументации (через интеграцию показателей сложности разных типов). Эта задача решается применением алгоритмов кластеризации трёх типов (центроидного *K-means*, иерархического *Ward* и графового *Spectral*) и формальных оценок качества кластеризации (*V-measure*, *Fowlkes-Mallow score*, оценки на основе коэффициента Жаккара).

Представленные во второй главе методы используются в экспериментах, описываемых в третьей и четвёртой главах диссертации.

## Глава 3. Выявление элементов аргументационных стратегий в графах аргументации

В третьей главе диссертационного исследования описаны эксперименты по решению отдельных задач формального выявления элементов аргументационных стратегий. Представлен корпус, использованный на текущем этапе исследования, указаны подсчитанные значения коэффициентов согласия аннотаторов на каждом уровне разметки, рассматриваются ключевые причины расхождений.

Эксперимент по частотному анализу элементарных приёмов аргументации показывает, что 15 типовых моделей рассуждения, применяемых в более чем 20% текстов проанализированной коллекции, покрывают 96% её аргументов. Это свидетельствует о сравнительной однородности доказательств в научных статьях корпуса: авторы исследованных текстов предпочитают строить рассуждения на основе узкого круга типовых моделей и избегают активного применения схем, нетипичных для жанра. Выявлена и тематическая зависимость в использовании элементарных приёмов для двух сравниваемых предметных областей: лингвистики и информационных технологий. Так, лингвистическим статьям свойственно активное приведение примеров для иллюстрации языковых фактов (67% от всего числа реализаций схемы *Example*), изучение глубинных языковых структур через их наглядные внешние проявления (по схемам *Correlation to Cause* и *Sign*, соответственно 64% и 69% аргументов с которыми применяются именно в публикациях по языкознанию) и обращение ко взглядам иных исследователей (по модели *Expert Opinion*, 61%). Прикладная же направленность работ по информационным технологиям влечёт предпочтение практических доказательств через анализ решаемой задачи (*Practical Reasoning*, 60%) либо за счёт эксплицитной оценки возможных результатов (*Positive Consequences* и *Negative Consequences*, причём рассуждения по этим моделям не появляются ни в одной лингвистической статье корпуса). Общая жанровая специфика статей обеих тематических групп проявляется в одинаково интенсивном анализе прямых причинно-следственных связей и основательной систематизации описываемых явлений (по моделям *Cause to Effect* и *Verbal Classification*, которые используются в публикациях по лингвистике и информационным технологиям со сходной частотой, а при этом покрывают соответственно 16% и 12% аргументов всей коллекции при учёте любых аргументационных схем). Отдельно акцентированы результаты экспериментов по компьютерному распознаванию аргументов отдельных типов (*Example*, *Expert Opinion*, *Practical Reasoning*, *Correlation to Cause*, *Verbal Classification*, *Part to Whole*) методами машинного обучения, приводятся оценки качества, указываются типовые причины ошибок (пересечение аргументов на уровне утверждений, многозначные маркеры, реализация аргументов без маркеров).

Эксперимент по компьютерному распознаванию тезисов методами МО показывает возможность извлечения аргументов с помощью таких алгоритмов

для частичной автоматизации ручной разметки и применимость статистической фильтрации значимых для аргументации лемм (на основе их частот в аргументативных и неаргументативных сегментах текста, без экспертного составления маркеров). Качество компьютерного распознавания тезисов оценивается в терминах точности, полноты и усредняющей их F-меры. Лучших показателей по точности распознавания достигает многослойный перцептрон (72.50%, а его полнота и F-мера равны 63.74% и 67.84%), наибольшая же полнота и F-мера отмечается для полиномиального байесовского классификатора (86.81% и 68.70% при точности в 56.83%). Анализ ошибок показывает, что основную часть неправильно распознанных предложений составляют предложения, не содержащие маркеров аргументации или содержащие неоднозначные маркеры. Тезисы в аргументах с отсутствующими в обучающей коллекции (для обучения алгоритмов) моделями рассуждения, тем не менее, распознаются в тестовой при наличии в них риторических и аргументативных маркеров (которые опознаются компьютером за счёт методов формальной фильтрации).

### Глава 4. Стратегии аргументации как совокупность общих приёмов

В четвёртой главе диссертации описаны эксперименты по компьютерному анализу аргументационных структур в более сложных аспектах: идентификации в графах составных приёмов аргументации (как в текстах одного жанра, научных статьях, так и в текстах разных жанров научной коммуникации), выявлению методов убеждения на уровне отдельных аргументов, оценке полных текстов по выраженности эмоционального воздействия и апелляций к авторитету. Представлен эксперимент по кластеризации текстов по формальной сложности аргументации через интеграцию элементов аргументационных структур разных типов (на основе предшествующих экспериментов), результатом чего становится разграничение обобщённых подходов к аргументации (стратегий) в научных статьях из проанализированной коллекции. Заключительный эксперимент состоит в автоматической жанровой идентификации текстов научной коммуникации посредством классификации по приёмам аргументации разной сложности.

Анализ составных приёмов в научных статьях показывает, что построение параллельных доводов чаще происходит по одной и той же модели вне зависимости от темы текста (сочетание нескольких доводов через разные модели является менее типичным). Другой особенностью в употреблении приёмов аргументации является заметное влияние тематической области на сочетаемость между схемами (среди 15 наиболее частотных приёмов треть реализуется только в лингвистических статьях, хотя образующие их отдельные схемы активно используются и в текстах по компьютерным технологиям). При расширении анализа на многожанровую коллекцию (большее число научных статей, научно-популярные статьи Habr и научные новости) отмечена жанровая специфика выявленных приёмов: устойчивая вариативность доказательств

в научных статьях при низком числе уникальных схем, высокая частота полемических приёмов в научно-популярных интернет-текстах, ограниченность новостей преобладающими в них схемами «Expert Opinion» и «Evidence to Hypothesis», непродуктивными в образовании составных приёмов. Посредством эксперимента по кластеризации демонстрируется высокая различительная способность приёмов аргументации в представлении текстов, их эффективность в разграничении текстов как по жанру, так и по тематике.

При выявлении методов убеждения по словарям маркеров отмечена точность работы маркеров на экспериментальном корпусе, равная 78% (причина ошибок – полисемия констант в шаблонах). Показано, что 67% тезисов не содержат маркеров этоса и пафоса, что объясняется прагматической сложностью аргументации (обоснование тезисов производится не применением особых лексических и грамматических констант, а через организацию утверждений с разнообразным содержанием в целостную аргументационную структуру). При анализе текстов по соотношению пафоса и этоса выявлено, что прикладная направленность статей по информационным технологиям влечёт активное использование в них оценочной лексики и конструкций с модальностью долженствования, тогда как теоретический характер лингвистических работ проявляется в нейтральности стиля, но также частом обращении к авторитетам иных исследователей. В эксперименте по компьютерному распознаванию аргументов с этосом и пафосом достигнуты значения F-меры в 65% для этоса и 64% для пафоса. При анализе ошибок показано, что выражение пафоса облегчает выявление этоса, но при этом выражение этоса размывает для алгоритмов выявление пафоса (при эмоциональных отсылках к авторитету оказывается, что классификатор выносит решение преимущественно на основе ссылок на авторитет, тогда как эмоциональные выражения оказывают лишь второстепенное влияние на работу алгоритма).

В рамках перехода на уровень стратегий аргументации предложена обобщающая классификация схем по сходству их значения и специфики реализации в текстах: описаны четыре функциональные группы схем («от авторитета», «от практической значимости», «анализ причинно-следственных связей», «через детализацию»). Эти группы (функциональные аргументационные группы – ФАГ) содержат 14 из 15 схем, покрывающих 96% аргументов в научных статьях корпуса, за исключением схемы «Logical Conflict». Представлена процедура преобразования последовательностей схем в цепочки ФАГ. Анализ преобразованных цепочек показывает, что авторам научных текстов свойственно построение многоаспектных доказательств с активным чередованием аргументов разных ФАГ в последовательном развитии доводов.

Выявление текстов со сходной аргументационной структурой осуществляется путём агрегирования всех ранее рассмотренных показателей аргументации. В эксперименте применены пять наборов аргументационных характеристик: элементарные приёмы, составные приёмы, структурные признаки полных графов аргументации, выраженность методов убеждения, обобщённые

составные приёмы на уровне функциональных групп. При кластеризации текстов по разным сочетаниям признаков разными алгоритмами установлено, что наибольшая надёжность кластеров (определяемая через подсчёт показателей качества кластеризации для каждого случая) достигается на алгоритме Spectral на сочетании трёх типов признаков: обобщённых приёмах по функциональным группам, структурным характеристикам и методам убеждения. В результате анализа построенных кластеров разграничены четыре подхода к аргументации (стратегии) в научных статьях коллекции. Выявлена взаимосвязь между значениями разных показателей (например, при разнообразии приёмов аргументации в тексте отмечается приведение малого числа тезисов без обоснования, тогда как обилие таких тезисов характеризует статьи с частыми апелляциями к авторитету).

Эффективность формальных аргументационных структур в передаче содержательной специфики текстов подтверждена и посредством заключительного эксперимента по жанровой атрибуции. Компьютерное распознавание жанра решается как задача классификации с проверкой эффективности использования приёмов аргументации разной сложности (отдельных схем, цепочек из двух схем в связанных аргументах (биграмм), комплексных приёмов из трёх и более схем, а также полного спектра приёмов аргументации). Представление текстов в эксперименте образовано только сведениями о реализации в них приёмов аргументации (без использования признаков иных уровней, таких как лексических). Для формирования тестовых коллекций использован метод K-fold (пять тестовых коллекций с равной пропорцией текстов трёх разных жанров зафиксированы для всех четырёх наборов признаков). Установлено, что качество жанровой атрибуции возрастает с увеличением сложности используемых приёмов: от 82% F-меры по микро-усреднению для отдельных схем до 89% для биграмм, 91% для комплексных приёмов и 95% для полного спектра приёмов аргументации. Анализ ошибок показывает, что некорректно классифицируемые тексты различаются между приёмами разных типов (отчего использование полного спектра позволяет нивелировать ошибки на приёмах отдельных типов), а причины ошибок могут быть сведены к двум основным: 1) реализация в тексте приёмов, характерных текстам иных жанров; 2) отсутствие в тексте приёмов, свойственных другим текстам этого же жанра. По мере усложнения приёмов аргументации увеличивается число ошибок второго вида, тогда как ошибки первого вида становятся заметно менее частотными. Наибольшее число ошибок в эксперименте характеризует научно-популярные статьи, наименьшее же – научные новости, что в обоих случаях объясняется их жанровой спецификой: научные новости представляют собой наиболее своеобразный жанр из трёх ввиду ориентированности на широкую публику, направленности научной коммуникации вовне, а не к академической либо узкопрофессиональной аудитории, тогда как научно-популярные тексты с онлайн-платформы лишены строгих стилистических ограничений,

свойственных академическим статьям, но при этом посвящены не менее скрупулёзному анализу излагаемого материала, из-за чего близки обоим другим жанрам.

## Заключение

Ключевые итоги диссертационного исследования, полученные в результате экспериментов по формальному анализу приёмов и стратегий аргументации в коллекции текстов научных статей, подтверждают положения, выносимые на защиту.

1. Подготовлен корпус с многоуровневой аргументационной разметкой научных статей на русском языке, включающий 100 текстов из двух предметных областей, где для каждого текста представлена двойная разметка разными аннотаторами. Вклад соискателя заключается а) в первоначальном отборе текстов для разметки; б) в разметке 92 статей, в которых выявлено 5297 утверждений и 4563 аргументов; в) в создании методологии многоаспектной разметки аргументации в научных статьях, регулирующей согласованную разметку текстов разными аннотаторами; г) в разработке программных скриптов для автоматической модификации разметки и устранению типовых расхождений между аннотаторами. Моделирование аргументационных структур данных текстов осуществлено в соответствии с дискурсивным подходом и стандартом Argument Interchange Format. Разметка аргументации включает как выявление тезисов в оформлении языковыми средствами, так и указание структурных связей между ними с обозначением абстрактных моделей рассуждения в основе этих связей (согласно классификации аргументационных схем Д. Уолтона). Корпус может применяться для решения различных задач в области компьютерной обработки аргументации, как показано в проведённых в ходе исследования экспериментах. Сочетание языковых средств оформления аргументов и их абстрактно-логических моделей при формальном представлении аргументации обеспечивает возможность многоаспектного аргументационного анализа.

2. Показано, что аргументационные структуры в обработанных текстах образованы небольшим числом базовых моделей рассуждения (15 типовых аргументационных схем покрывают 96% всех аргументов корпуса). Отмечается регулярное использование авторами разных текстов одних и тех же составных приёмов аргументации, которые соответствуют конфигурациям из нескольких отдельных моделей (элементарных приёмов). Выявлена зависимость в реализации как элементарных, так и составных приёмов аргументации от жанра и предметной области, к которым относится текст.

3. Отмечено, что прикладной характер статей корпуса по информационным технологиям влечёт активное употребление в них оценочной лексики и модальных конструкций долженствования при оформлении логических доказательств языковыми средствами. Такие показатели непрямого эмоционального

воздействия практически не выражены в трудах по лингвистике, которым ввиду теоретического характера свойственно более активное обращение к авторитету иных исследователей.

4. Достигнуты приемлемые показатели качества при компьютерном распознавании элементов аргументации (тезисов, аргументов по методам убеждения, отдельных типов аргументов, связей в их основе) посредством реализованного программного комплекса из различных блоков компьютерной обработки аргументации, что свидетельствует о возможности исследования приемов в заранее неразмеченных текстах. На этапе распознавания тезисов достигнута F-мера, равная 69%, для выявления аргументов с пафосом и этосом – 66% и 65%, для аргументов отдельных шести типов – от 54% до 68%.

5. Выявлена явная жанровая специфика составных приёмов аргументации в текстах разных жанров научной коммуникации. Научным статьям корпуса свойственна устойчивая вариативность доказательств при ограниченном числе уникальных схем, научно-популярные тексты выделяются частотой полемических приёмов, научные новости отличаются низким разнообразием приёмов ввиду непродуктивности их преобладающих схем в образовании регулярных составных конструкций. Посредством кластеризации текстов, представленных по совокупности реализованных в них приёмов, показана высокая различительная способность приёмов аргументации в передаче жанровой и тематической специфики текстов научной коммуникации.

6. На основе интеграции формальных характеристик аргументации в разных аспектах (обработанных в ходе отдельных предварительных экспериментов) реализован метод автоматической кластеризации научных статей со сходной организацией аргументации. Продемонстрировано, что компьютерная обработка аргументационных структур по их формальным показателям обладает достаточной различительной способностью для разграничения групп текстов с наглядной передачей многоаспектной специфики доказательств внутри каждой группы. Эффективность формальных аргументационных структур в передаче содержательной специфики текстов продемонстрирована и посредством эксперимента по жанровой атрибуции. Это обеспечивает возможность компьютерной характеризации стратегий аргументации, реализуемых на уровне полных текстов, по структурной и языковой специфике излагаемых в них аргументов.

Обозначены перспективы дальнейших исследований в области формального анализа аргументации: расширение корпуса размеченных текстов, разработка новых методов компьютерной обработки аргументации для решения широкого спектра прикладных задач.

**Публикации по теме диссертации**

*Статьи в изданиях из перечня российских рецензируемых научных журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата наук*

1. Пименов, И. С. Формальное выявление приемов аргументации в научных текстах / И. С. Пименов, Н. В. Саломатина, М. К. Тимофеева // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. – 2022. – Т. 20, № 1. – С. 21–36. – DOI: 10.25205/1818-7935-2022-20-1-21-36. (RSCI) (1,2 а.л.; авторский вклад – 0,4 а.л.)

2. Пименов, И. С. Сочетаемость аргументов разных функциональных групп в научных текстах / И. С. Пименов // Филологические науки. Вопросы теории и практики. – Тамбов: Грамота. – 2022. – № 11. – С. 3672–3680. – DOI: 10.30853/phil20220581. (РИНЦ) (1 а.л.)

3. Пименов, И. С. Анализ расхождений в аргументационной разметке научных статей на русском языке / И. С. Пименов // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. – 2023. – Т. 21, № 2. – С. 89–104. – DOI: 10.25205/1818-7935-2023-21-2-89-104. (RSCI) (1,3 а.л.)

4. Пименов, И. С. Компьютерный анализ приемов и стратегий аргументации в текстах научной коммуникации / И. С. Пименов, Н. В. Саломатина, М. К. Тимофеева // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. – 2025. – Т. 23, № 1. – С. 93–109. – DOI: 10.25205/1818-7935-2025-23-1-93-109. (Белый список, Q1) (1,4 а.л.; авторский вклад – 0,47 а.л.)

5. Salomatina, N. V. Identification of connected arguments based on reasoning schemes “from expert opinion” / N. V. Salomatina, I. S. Kononenko, E. A. Sidorova, I. S. Pimenov // Journal of Physics: Conference Series, Volume 1715, International Conference «Marchuk Scientific Readings 2020» (MSR-2020). – 2020. – Russia, Novosibirsk. – Pp. 1–10. – DOI: 10.1088/1742-6596/1715/1/012013. (Scopus) (1 а.л.; авторский вклад – 0,25 а.л.)

6. Salomatina, N. V. Identification of argumentative sentences in Russian scientific and popular science texts / N. V. Salomatina, E. A. Sidorova, I. S. Pimenov // Journal of Physics: Conference Series, Volume 2099, International Conference «Marchuk Scientific Readings 2021» (MSR-2021). – 2021. – Russia, Novosibirsk. – Pp. 1–8. – DOI: 10.1088/1742-6596/2099/1/012025. (Scopus) (0,75 а.л.; авторский вклад – 0,25 а.л.)

7. Pimenov, I. S. The Quantitative Evaluation of the Pathos to Ethos Ratio in Scientific Texts / I. S. Pimenov, N. V. Salomatina, M. K. Timofeeva // 2022 IEEE 23rd International Conference of Young Professionals in Electron Devices and Materials (EDM). – 2022. – Russia, Altai. – Pp. 312–317. – DOI: 10.1109/EDM55285.2022.9855111. (Scopus) (0,75 а.л.; авторский вклад – 0,25 а.л.)

8. Zasypkin, A. S. The Combined Approach to Identifying Argumentation Structures in Short Scientific Papers // A. S. Zasypkin, I. S. Pimenov, N. V. Salomatina // 2023 IEEE 24rd International Conference of Young Professionals in Electron Devices and Materials (EDM). – 2023. – Russia, Altai. – Pp. 1800–1805. – DOI: 10.1109/EDM58354.2023.10225223. (Scopus) (0,9 а.л.; авторский вклад – 0,3 а.л.)

9. Pimenov, I. S. Evaluating the Influence of Argumentation Markers on the Identification of Reasoning Models / I. S. Pimenov, N. V. Salomatina // Communications in Computer and Information Science. – Vol. 2086. – 2024. – Pp. 282–297. – DOI: 10.1007/978-3-031-67826-4_21 (Scopus) (1,1 а.л.; авторский вклад – 0,37 а.л.)

10. Pimenov, I. S. An Automatic Method for Standartizing Argumentative Annotations across Annotators Genres / I. S. Pimenov, N. V. Salomatina // 2024 IEEE 25th International Conference of Young Professionals in Electron Devices and Materials (EDM). – 2024. – Pp. 2260–2265. – DOI: 10.1109/EDM61683.2024.10615176. (Scopus) (0,9 а.л.; авторский вклад – 0,45 а.л.)

11. Pimenov, I. S. Productivity-Based Analysis of Argumentation Patterns across Texts of Different Genres / I. S. Pimenov, N. V. Salomatina // 2024 IEEE 25th International Conference of Young Professionals in Electron Devices and Materials (EDM) – 2024. – Pp. 2270–2275. – DOI: 10.1109/EDM61683.2024.10615197. (Scopus) (0,75 а.л.; авторский вклад – 0,37 а.л.)

12. Pimenov, I. Employing Argumentation Patterns for Genre Classification of Scientific Communication Texts / I. Pimenov, N. Salomatina // 2025 IEEE 26rd International Conference of Young Professionals in Electron Devices and Materials (EDM). – 2025. – Russia, Altai. – Pp. 1780–1785. – DOI: 10.1109/EDM65517.2025.11096648. (Scopus) (0,9 а.л.; авторский вклад – 0,45 а.л.)

*Статьи в прочих научных изданиях*

13. Саломатина, Н. В. Распознавание аргументативных связей в научно-популярных текстах / Н. В. Саломатина, И. С. Кононенко, Е. А. Сидорова, И. С. Пименов // Системная информатика. – № 16. – 2020. – С. 149–164. – DOI: 10.31144/si.2307-6410.2020.n16.p149-164. (РИНЦ) (0,82 а.л.; авторский вклад – 0,21 а.л.)

14. Пименов, И. С. Специфика аргументационного аннотирования научных и научно-популярных текстов / И. С. Пименов // Труды международной конференции «Корпусная лингвистика-2021». – Санкт-Петербург, Скифия-принт. – 2021. – С. 330–337. (РИНЦ) (0,32 а.л.)

15. Pimenov, I. S. Identification of Scientific Texts with Similar Argumentation Complexity / I. S. Pimenov, N. V. Salomatina // 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). – 2022.

– Pp. 870–875. – DOI: 10.1109/SIBIRCON56155.2022.10017119. (Scopus) (0,82 а.л.; авторский вклад – 0,41 а.л.)

16. Пименов, И. С. Маркерный подход к автоматизации извлечения аргументативных структур из научных статей / И. С. Пименов, Н. В. Саломатина, А. С. Засыпкин // Труды международной конференции «Корпусная лингвистика – 2023». – Санкт-Петербург. – 2024. – С. 184–191. (РИНЦ) (0,32 а.л.; авторский вклад – 0,1 а.л.)

17. Саломатина, Н. В. Применение методов машинного обучения для выявления аргументативных связей в текстах научной коммуникации / Н. В. Саломатина, Е. А. Сидорова, И. С. Пименов // Онтология проектирования. – № 14 (51). – 2024. – С. 82–93. – DOI: 10.18287/2223-9537-2024-14-1-82-93. (RSCI) (1,1 а.л.; авторский вклад – 0,37 а.л.)

18. Саломатина, Н. В. Распознавание в научных текстах аргументов с этосом и пафосом / Н. В. Саломатина, И. С. Пименов, Е. А. Сидорова // Тезисы Международной научной конференции «Марчуковские научные чтения – 2022». – Новосибирск. – 2022. – С. 140. (РИНЦ) (0,03 а.л.; авторский вклад – 0,01 а.л.)

19. Саломатина, Н. В. Автоматическое обнаружение главного и локальных базовых аргументов в научном тексте на русском языке / Н. В. Саломатина, И. С. Пименов // Тезисы Международной научной конференции «Марчуковские научные чтения 2024». – Новосибирск. – 2024. – С. 128–129. (РИНЦ) (0,05 а.л.; авторский вклад – 0,02 а.л.)

20. Пименов, И. С. Анализ приёмов аргументации в текстах научной коммуникации / И. С. Пименов // Обработка естественного языка для лингвистики, IT, бизнеса: сборник материалов Международного научно-практического семинара (отв. ред. З. И. Резанова). – Томск: Издательство Томского государственного университета. – 2025. – С. 42–44. (РИНЦ) (0,07 а.л.)