

Научная статья

УДК 81'33

DOI 10.25205/1818-7935-2025-23-1-93-109

Компьютерный анализ приемов и стратегий аргументации в текстах научной коммуникации

Иван Сергеевич Пименов
Наталья Васильевна Саломатина
Мария Кирилловна Тимофеева

Институт систем информатики им. А. П. Ершова
Новосибирск, Россия

pimenov.1330@yandex.ru, <https://orcid.org/0000-0001-5946-9469>

salomatina_nv@live.ru, <https://orcid.org/0000-0001-8412-9116>

mtimof@inbox.ru, <https://orcid.org/0000-0001-8999-2330>

Аннотация

В статье приведены результаты анализа приемов аргументации, реализуемых в текстах научной коммуникации. Под приемами понимаются повторяющиеся в корпусе текстов с разметкой аргументации отдельные типы аргументов и структуры, образуемые ими. Корпус содержит тексты трех жанров: научные статьи по лингвистике и информационным технологиям из научной электронной библиотеки CyberLeninka (<https://cyberleninka.ru/>), научно-популярные статьи с форума «Хабр» (habr.com/ru, далее – статьи Habr), новости науки (poisknews.ru). Двойная разметка аргументации в текстах корпуса произведена на платформе ArgNetBank Studio пятью экспертами, специалистами в области теоретической и прикладной лингвистики. Вычисленные коэффициенты согласия между аннотаторами могут быть отнесены к разряду «существенное согласие». Моделирование аргументации соответствует стандарту Argument Interchange Format. Результатом разметки и объектом анализа являются построенные согласно стандарту графы аргументации с двумя типами вершин: информационными вершинами, содержащими аргументативные утверждения, и вершинами-схемами, определяющими для каждой связи между ними точную модель рассуждения из компендиума Уолтона. Исследуются графы, полученные из графов аргументации путем удаления всех информационных вершин и слияния входящей и выходящей дуг каждой исключенной информационной вершины в одну дугу (таким образом получаемые графы состоят только из вершин-схем), и подграфы таких графов. Частотный анализ подграфов, составляющих графы корпуса, проведен методом Frequent Subgraph Mining с учетом изоморфизма, выявляемого посредством реализации алгоритма Корделлы VF2 из библиотеки NetworkX. В результате выявлены приемы аргументации (повторяющиеся подграфы с числом вершин от 1 до 9), применяемые в текстах всех жанров (межжанровые), каждого отдельного жанра (межтекстовые), а также приемы, повторяющиеся в отдельных текстах (внутритекстовые). Установлено, что для научных статей корпуса характерны наибольшие устойчивость и разнообразие приемов аргументации, научно-популярные характеризуются активной полемической аргументацией, а приемы в научных новостях основаны на двух ключевых моделях, применяемых при компактной аргументации. Внутритекстовые приемы повторяют тенденции межтекстовых, тогда как приемы, встречающиеся в текстах всех жанров, редки. Совокупность приемов, используемых в тексте, образует применяемую в нем стратегию рассуждения. С целью выявления близких по стратегиям текстов проведена их кластеризация методами Ward и K-means. Построенные кластеры характеризуются однородностью текстов по жанру, а в кластерах с текстами одного жанра – по теме. Выявленные приемы аргументации, помимо классификации текстов, могут также применяться для оценки аргументативной составляющей текста, поиска убедительной аргументации, ее генерации и т. д. Работы по данной теме редко встречаются в исследованиях для текстов на английском языке, а для текстов на русском языке автоматический анализ закономерностей аргументации не проводился.

© Пименов И. С., Саломатина Н. В., Тимофеева М. К., 2025

ISSN 1818-7935

Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2025. Т. 23, № 1
Vestnik NSU. Series: Linguistics and Intercultural Communication, 2025, vol. 23, no. 1

Ключевые слова

вычислительная аргументация, научная коммуникация, русскоязычные тексты, приемы аргументации, стратегии аргументации

Финансирование

Исследование выполнено за счет гранта Российского научного фонда № 23-11-00261, <https://rscf.ru/project/23-11-00261/>

Для цитирования

Пименов И. С., Саломатина Н. В., Тимофеева М. К. Компьютерный анализ приемов и стратегий аргументации в текстах научной коммуникации // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2025. Т. 23, № 1. С. 93–109. DOI 10.25205/1818-7935-2025-23-1-93-109

Computer-based Analysis of Argumentation Patterns and Strategies in Scientific Communication Texts

Ivan S. Pimenov, Natalia V. Salomatina,
Mariya K. Timofeeva

A. P. Ershov Institute of Informatics Systems
Novosibirsk, Russian Federation

pimenov.1330@yandex.ru, <https://orcid.org/0000-0001-5946-9469>

salomatina_nv@live.ru, <https://orcid.org/0000-0001-8412-9116>

mtimof@inbox.ru, <https://orcid.org/0000-0001-8999-2330>

Abstract

The article presents the results of analyzing argumentation patterns that are employed in scientific communication texts. We define argumentation patterns as models of separate arguments and their composite structures repeated across corpus texts. The corpus consists of 98 texts in three different genres: 50 scientific articles from the scientific electronic library CyberLeninka (<https://cyberleninka.ru/>), 20 popular science articles from the forum habr.com/ru (hereafter – Habr articles), 28 scientific news texts from the Poisk aggregator of scientific news (poisknews.ru). Five experts in theoretical and applied linguistics have performed double annotation of the corpus texts using ArgNetBank Studio platform. Calculated inter-annotator agreement scores correspond to the “substantial agreement” range. Argumentation annotation follows the Argument Interchange Format standard. The results of annotation and the object of the analysis are argumentation graphs with two node types: information nodes (which correspond to argumentative statements) and scheme nodes (which, for each link between statements, indicate its reasoning model in accordance with Walton’s compendium). The study addresses the graphs obtained from the abovementioned graphs by excluding all information nodes and merging the incoming and outgoing links of each excluded information node into one link (thus the obtained graphs consist exclusively of scheme nodes) and the subgraphs of these obtained graphs. We perform a frequency-based analysis of subgraphs by the Frequent Subgraph Mining Method combined with Cordella VF2 algorithm implementation from the NetworkX library for isomorphism check. The analysis results in identification of argumentation patterns (repeating subgraphs from 1 to 9 node in size) that are employed in texts of all genres, one specific genre, as well as patterns repeated within separate texts. We show that scientific articles of the corpus exhibit the greatest stability and diversity of argumentation patterns, while popular science texts are marked by the active use of polemic argumentation, and patterns in scientific news are based on two reasoning models characteristic of compact argumentation. Patterns that repeat in separate texts show the same tendencies as genre-specific patterns. Using the identified patterns, we conduct clustering of the corpus texts with Ward and K-means algorithms. Resulting clusters correspond to groups of texts with similar argumentation strategies and exhibit homogeneity of texts by genre and thematic field. Identified patterns can be used not only for classifications but also for assessing the argumentative organization of texts extracting persuasive argumentation, its generation etc. Works on this topic are still underrepresented for texts in English, while for texts in Russian, such automatic analysis of argumentation patterns has not yet been performed.

Keywords

computational argumentation, scientific communication, Russian texts, argumentation patterns, argumentation strategies

Acknowledgements

The study was supported by the Russian Science Foundation grant No. 23-11-00261, <https://rscf.ru/project/23-11-00261/>

For citation

Pimenov I. S., Salomatina N. V., Timofeeva M. K. Computer-based Analysis of Argumentation Patterns and Strategies in Scientific Communication Texts. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2025, vol. 23, no. 1, pp. 93–109. DOI 10.25205/1818-7935-2025-23-1-93-109

Введение

Вычислительная аргументация (Computational Argumentation, CA) является одной из активно развивающихся подобластей автоматической обработки текста, включающей несколько направлений исследования, таких как поиск аргументов, оценка убедительности, генерация аргументации и др. Она отличается особой сложностью, поскольку требует интеграции большого объема знаний различных типов [Lauscher et al., 2022]. Знания о стратегиях аргументации важны для приложений любого из направлений. Однако в области CA задачи вычислительного анализа стратегий аргументации в текстах долго упускались из виду. Практически все работы в этой области принадлежат коллективу из Веймарского университета (Bauhaus-Universität Weimar, Germany), один из участников которого в 2019 году защитил PhD на тему „Computational Analysis of Argumentation Strategies“ [Al Khatib, 2019]. Они ограничены одним жанром – новостными редакционными статьями, а также небольшим набором стратегий, выявленных и размеченных вручную экспертами с целью применения методов машинного обучения (МО) для автоматической разметки.

Русскоязычных работ в этой области практически нет, кроме нашей работы по приемам аргументации, применяемым в коротких научных статьях [Пименов, Саломатина, Тимофеева, 2022]. В отличие от авторов работы [Al-Khatib et al., 2016] мы определяем приемы формально как повторы структур аргументации разной сложности в аннотированных текстах с аргументативной разметкой (далее для краткости – аннотациях), построенных экспертами. Однозначное соответствие между формально определенными приемами и предложенными в [Al-Khatib et al., 2016] стратегиями установить не всегда возможно. Стратегиями в нашем исследовании мы называем всю совокупность приемов в отдельном тексте. Близкие по применяемым стратегиям тексты в нашем случае могут быть выявлены формально. Материал нашего исследования существенно шире – это корпус текстов научной коммуникации (новости науки, научно-популярные статьи *Nabr*, научные статьи). Его выгодно отличает детализированная разметка аргументации, что в дальнейшем позволит точнее вычислять убедительность приемов и стратегий.

Цель работы – с использованием формализованных понятий «прием» и «стратегия» модифицировать предложенные в [Пименов, Саломатина, Тимофеева, 2022] методы поиска повторяющихся структур аргументации в аргументативных аннотациях текстов, провести эксперименты по автоматическому поиску приемов в корпусе текстов научной коммуникации с последующим анализом полученных результатов.

Полученную нами статистику использования приемов и стратегий, по крайней мере для научных текстов, можно считать вполне достоверной, поскольку в корпусе размещены только опубликованные тексты, т. е. те, что прошли рецензирование. Аналогично отобранные нами научно-популярные статьи *Nabr* характеризуются высокой пользовательской оценкой, что свидетельствует о достаточной убедительности авторской аргументации.

1. Автоматическое исследование и использование приемов аргументации

Изучение приемов аргументации. Для автоматического исследования приемов аргументации, применяемых в разных текстах, необходимы специальным образом размеченные корпуса. Русскоязычных корпусов такого сорта, кроме ArgNetBank Studio (<https://uniserv.iis.nsk.su/>

arg/corpora), нет, но и в этом корпусе указываются лишь типы отдельных аргументов, но не образуемые ими повторяющиеся структуры. Из работы [Al-Khatib et al., 2016] нам известен англоязычный корпус из 300 редакционных статей, собранных с нескольких новостных порталов, и корпус из 1000 редакционных статей, способ аннотирования которых отражает качество аргументации, понимаемое как убедительность. В корпусе «вручную» размечены 5 свойственных новостям основных элементарных стратегий (strategy) аргументации: Anecdote (личный опыт, конкретный пример, событие), Assumption (предположение, суждение, мнение автора), Common ground (общеизвестный факт), Statistics (результаты количественного анализа), Testimony (свидетельство от эксперта, организации, очевидца). Согласие Флейса между тремя аннотаторами (0,56) обозначено авторами как высокое.

В продолжение работы [Al-Khatib et al., 2016] по исследованию стратегий, применяемых в новостях, в своей следующей работе [Al-Khatib et al., 2017] авторы представили результаты анализа 29 тыс. новостных статей. Изучено применение трех стратегий аргументации (Statistics, Anecdote, Assumption) в текстах 12 различных тем. Разметка анализируемых новостей проводилась автоматически методом SVM (Support Vector Machine) со средней по всем стратегиям, включая Other (класс, не содержащий ни одной из трех стратегий), мерой $F1 = 0,62$. Анализ показал, как авторы статей рассуждают в новостных передовицах, как тема влияет на выбор и последовательность применения стратегий (может содержать до 7 стратегий). Установлена корреляция между темами и применяемыми стратегиями. Авторы считают, что полученные результаты дают ценную информацию для синтеза эффективных аргументативных текстов.

Как уже упоминалось выше, в нашей работе [Пименов, Саломатина, Тимофеева, 2022] представлено исследование приемов (типов отдельных аргументов и повторяющихся конструкций из них – цепочек и деревьев) в аргументативных аннотациях (графах аргументации) научных статей двух тематических направлений. Идентификация структурных приемов производилась автоматически путем частотного анализа подграфов, выполненного с использованием реализации алгоритма точной проверки на изоморфизм VF2. Выявленные общие и специфические структурные свойства аргументации подтвердили существование связи между темой и применяемыми типами аргументов, их структурой.

Задача определения аргументативных стратегий не сводится только к анализу структурной организации, она также предполагает изучение общих подходов к представлению доводов, проявляющихся на уровне полного текста. Дополнение логической составляющей аргументов (этоса) способами воздействия на аудиторию (манипуляцией эмоциями, апелляцией к авторитету) (пафоса) помогают вызывать доверие у людей разных психотипов и тем самым повышать убедительность всего текста. Работа [Саломатина, Пименов, Сидорова, 2022] направлена на выявление способов воздействия, а именно: этоса и пафоса, в аргументации научных текстов на русском языке классическими методами машинного обучения. Бинарная классификация аргументов этос/не-этос и пафос/не-пафос методом MNB (Multinomial Naive Bayes) показала лучший результат по F-мере – 0,65 и 0,64 в распознавании этоса и пафоса соответственно. Полученные оценки подтверждают возможность автоматического определения использованных в тексте методов воздействия.

Классификация текстов с применением аргументативной составляющей. Авторами [Anand et al., 2011] исследована возможность классификации сообщений блогов (в 457 сообщениях размечены выявленные экспертами 14 тактик (acts)) на убедительные и неубедительные. Установлено, что использование информации о тактиках позволяет классифицировать сообщения методом наивного Байеса наилучшим образом по сравнению с другими исключающими эту информацию версиями.

Попытка привлечь информацию о применяемых авторами стратегиях для выявления фейков в новостных сообщениях была предпринята в работе [Yan, Lin, and Litman, 2018]. На базе списка стратегий, выявленных в работе [Al-Khatib et al., 2016], проведено исследование, показавшее, что не все стратегии применяются в фейковых новостях. Полученная оценка F1-меры

обнаружения этих стратегий в корпусе, собранном авторами из разных источников (15 200 статей), примерно одинакова для фейковых и заслуживающих доверие текстов. Предложен алгоритм распознавания фейковых новостей, использующий информацию о принадлежности фрагментов текста к стратегиям аргументации с помощью нейросетей. К сожалению, авторы приводят лишь значения аккуратности (доли правильных ответов) и матрицу ошибок. Однако высокие значения аккуратности позволяют авторам надеяться на хорошие результаты классификации в дальнейшем. Возможно, состав используемого списка стратегий недостаточен для решения поставленной задачи.

Кластеризация текстов. Знания о стратегиях аргументации могут быть применены для автоматического определения близких, например по сложности текстов. Такая задача решалась в работе [Pimenov, Salomatina, 2022]. С помощью алгоритмов кластеризации 30 русскоязычных научных статей с детализированной разметкой аргументации в виде графа были разбиты на группы на основе совокупности сведений о наличии/отсутствии общих подграфов в текстах (приемов аргументации) и характеристик общих свойств графа (высота, ширина и др.), а также оценок этоса и пафоса в текстах. Получены интерпретируемые кластеры близких по сложности текстов, на базе которых выявлены зависимости между признаками разных типов.

Генерация текстов. В самой ранней работе по синтезу аргументации [Zukerman, McConachy, and George, 2000] предложена система генерации аргументов с учетом пяти стратегий: Premise to goal, Reduction to absurdum, Inference to the best explanation, Reasoning by cases (exclusive или non-exclusive). Графы аргументов инкорпорированы в байесовскую сеть NAG (Nice Argument Generator). В алгоритм генерации включены шаги по выдвиганию и оценке стратегий, по результатам которых проводится отбор наиболее качественных аргументов.

Еще одним примером применения риторических стратегий может служить модель синтеза [Wachsmuth et al., 2018], разработанная в соответствии с предложенными Аристотелем методами убеждения и канонами риторики. Трехкомпонентная модель генерации аргументации включает последовательность действий, задаваемых выбираемым сценарием, определяющим (1) тему (главный тезис), позицию («за», «против»), стратегию (пропорцию методов убеждения в синтезированном тексте), (2) соответствующие заявленным в (1) аргументативные единицы-предложения (Argumentative Discourse Unit, ADU), объединяемые в аргументы, (3) цепочку следования аргументов согласно выбранной стратегии. Синтез аргументации выполняется вручную 26 квалифицированными экспертами из заранее сформированного пула (200 ADU по 10 темам) аргументативных единиц. Сходство синтезированных разными экспертами текстов оценивается на уровне 50 %.

В следующей работе [El Baff et al., 2019], непосредственно связанной с предыдущей, реализован алгоритм автоматического построения аргументативных цепочек из пула ADU вне зависимости от темы. Алгоритм выполняет действия, аналогичные действиям эксперта. Объединение выбранных из пула независимых от темы, но ориентированных на механизмы воздействия (логос, пафос), ADU в единый текст (цепочку из 5 аргументов) производится на основе предобученных вероятностных моделей, составляющих из этих цепочек аргумент по ролям (тезис, аргумент «за», аргумент «против») и строящих последовательность аргументов согласно модели семантической связности. Автоматически созданные таким образом тексты лишь отчасти напоминают синтезированные вручную экспертами. Совпадение структур аргументации, построенных человеком и компьютером, составило 40–50 % для двух подряд следующих аргументов. Для порождения цепочек большей длины требуется дополнительная информация и усовершенствование алгоритма. Детализация типов аргументации могла бы повысить качество синтеза и в определенной степени сгладить различия между генерируемыми текстами.

Прогнозирование убедительности текстов. Авторы статьи [Mirzakhmedova et al., 2023] исследовали влияние упорядочения аргументов определенных типов при организации их в последовательность на убедительность дискуссии, что полезно не только для получения более глубокого понимания того, как люди взаимодействуют, но и для предсказания правильности

их действий. На материале 34 тыс. цепочек обсуждения, для которых известен тег изменения точки зрения, убедительность дискуссий была предсказана со значением меры $F1 = 75\text{--}78\%$. Для этого использовалась двунаправленная сеть LSTM (с BERT-embeddings и включением информации о последовательностях типов аргументов).

Из приведенного обзора следует, что изучение приемов аргументации в настоящее время является недостаточным и востребованным.

2. Моделирование аргументации посредством схем рассуждения

Моделирование аргументации в представленном исследовании соответствует стандарту Argument Interchange Format (AIF) [Rahwan, Reed, 2009]. AIF определяет три основных этапа аннотирования аргументации: идентификация аргументативных утверждений, обнаружение связей между ними, указание модели (схемы) рассуждения для каждой связи. Идентификация состоит в сегментации текста на аргументативные и неаргументативные фрагменты и дальнейшей сегментации аргументативных фрагментов на отдельные аргументы и их части. Выявление связей включает в себя также уточнение роли утверждений в аргументах (какие утверждения являются посылками, а какое – заключением). В результате такого трехэтапного моделирования текст преобразуется в ориентированный граф аргументации с двумя типами узлов: информационными узлами, содержащими аргументативные утверждения, и узлами-схемами, определяющими точную модель рассуждения для каждой связи между утверждениями. Связи соответствуют ребрам в графе, причем каждое ребро соединяет узлы обоих типов: стандарт не поддерживает соединение двух информационных узлов без указания схемы рассуждения, лежащей в основе их объединения. Специфика схем рассуждения предполагает их выбор из некоторой классификации, в нашем случае компендиума Уолтона [Walton, Reed, Macagno, 2008], примененного к текстам различных жанров и рекомендованного разработчиками AIF. Более подробное описание процедуры аргументативного аннотирования текста с примерами дано в [Пименов, Саломатина, Тимофеева, 2022]. Стоит отметить, что формат разметки позволяет указывать лишь тип отдельных аргументов, а приемы, состоящие из нескольких аргументов, требуют дополнительной «ручной» или автоматической разметки.

3. Определения

В данной статье приемы аргументации рассматриваются как структуры из отдельных или нескольких связанных схем аргументации, *повторяющиеся* в разных аннотациях в коллекции либо внутри одной аннотации. Формальные определения были даны нами в работе [Пименов, Саломатина, Тимофеева, 2022]. Здесь напомним основные из них, используемые далее, и введем новые, необходимые для представления набора коллекций.

1. Размеченный корпус C включает множество текстов $T = \{t_i\}$ (i – номер текста, $i = 1, \dots, I$; I – число текстов) разных жанров: $T = \cup_{g \in \{N, H, S\}} T_g$, символом N обозначены научные новости, H – научно-популярные статьи *Набр*, S – научные статьи. Каждый текст снабжен аргументативной аннотацией a_i : $C = T \cup A = \{t_i \cup a_i\}_{i=1}^I$; $A = \{a_i\}$ – множество всех аннотаций ($A = \cup_{g \in \{N, H, S\}} A_g$).

2. Согласно формату AIF a_i представляет собой граф G_i , $G_i = \langle V_i^{inf} \cup V_i^{ch}, E_i \rangle$, где V_i^{inf} – множество информационных вершин, V_i^{ch} – множество вершин-схем и E_i – множество инцидентных им ребер. Анализируемые в работе графы $\hat{G}_i = \langle V_i^{ch}, \hat{E}_i \rangle$, получаемые из графа G_i путем исключения всех информационных вершин и слияния для каждой такой вершины вхо-

дующей в нее и выходящей из нее дуг в одну дугу; тем самым, граф \hat{G}_i содержит только вершины-схемы (преобразование G_i в \hat{G}_i дано в [Пименов, Саломатина, Тимофеева, 2022]).

3. В каждом тексте i внутритекстовому приему k соответствует снабженный частотной характеристикой $F_{abs}(sg_{ik}^n)$ подграф sg_{ik}^n графа \hat{G}_i ($sg_{ik}^n \subset \hat{G}_i$) (k может изменяться от 1 до K_n , K_n – количество разных подграфов с n вершинами, $n > 0$) при условии, что $F_{abs}(sg_{ik}^n) > 1$. Совокупность всех таких приемов в тексте обозначим $\Phi_g^{In}(a_i)$ (верхний индекс In обозначает принадлежность приема к внутритекстовому типу; далее индекс Ex – к межтекстовому).

4. Межтекстовым приемам фиксированного жанра соответствует подграф sg_{ik}^n из множества $sg_{ik=1}^{n|A_g|}$ с текстовой частотой $F_{txt}(sg_{ik}^n) > 1$. Текстовая частота равна числу аннотаций из A_g , в графах которых встречается данный подграф. Множество всех межтекстовых приемов обозначим $\Phi_g^{Ex}(a_i)$.

5. Межжанровые приемы Φ^{Ex} образуют подмножество подграфов из $\Phi_g^{Ex}(a_i)$ с жанровой частотой $F_g(sg_{ik}^n) > 1$, которая определяется числом аннотаций из A , принадлежащих хотя бы двум разным жанрам, в графах которых встречается данный подграф.

6. Стратегию аргументации текста t_i определим как совокупность приемов $\Phi(a_i) = \{\Phi_1(a_i), \dots, \Phi_n(a_i), \dots, \Phi_{N_{max}}(a_i)\}$, где $\Phi_1(a_i)$ является статистикой отдельных схем ($F_{abs}(sg_{ik}^1) > 1$). Приемы из $\Phi_n(a_i) = \{sg_{ik}^n\}$ с $n > 1$ содержат статистику повторяющихся в тексте и/или подкорпусе и/или корпусе подграфов $\{sg_{ik}^n\} \subset \hat{G}_i$ с числом вершин, равным n ($n = 2, \dots, N_{max} \leq |V_i^{ch}|$, где N_{max} – максимальное число вершин в подграфах a_i , удовлетворяющих частотным ограничениям). То есть подграфы из $\Phi_n(a_i)$ могут входить как в Φ_g^{Ex} , так и в Φ^{Ex} .

4. Методы и процедура построения множества приемов аргументации

Формирование множества $\Phi(A) = \bigcup_{n,i} \Phi_n(a_i)$ сводится к выявлению структурно идентичных общих подграфов как в отдельной аннотации a_i , так и во всей коллекции A . Для решения задач такого типа применяют методы частотного анализа подграфов *FSM* (Frequent subgraph mining) [Jiang et al., 2004], включающие генерацию представляющих интерес подграфов и подсчет их частоты встречаемости в заданном наборе данных. Для подсчета частоты встречаемости необходима проверка подграфов на изоморфизм. Самые известные алгоритмы, позволяющие получить точное решение, предложены в работах [Ullmann, 1976], [Cordella et al., 2004].

В нашем случае представляющими интерес подграфами являются подграфы графов $\{\hat{G}_i\}_{i=1}^I$ с фиксированным числом вершин n . Для установления изоморфизма мы применяем реализацию алгоритма Корделлы VF2 (NetworkX) [Hagberg, 2008], которая для каждой комбинации значений индексов $i, j, k, n; i \neq j$ определяет, не содержит ли граф \hat{G}_j подграф изоморфный подграфу $\{sg_{ik}^n\} \subset \hat{G}_i$.

Процедура формирования множества $\Phi(A)$ включает извлечение из $\{\hat{G}_i\}_{i=1}^I$ 1) отдельных схем $\Phi_1(A)$ с подсчетом F_{abs}, F_{txt}, F_g ; 2) множества пар связанных вершин $\Phi_2(A)$ с вычислением всех частотных характеристик; 3) подграфов sg_{ik}^n для всех $n > 2$ с проверкой подграфов с фиксированным n на изоморфизм и вычислением частотных характеристик. Подробное описание процедуры построения $\Phi(A)$ дано в [Пименов, Саломатина, Тимофеева, 2022]. В данной модификации добавлен этап фильтрации вложенных подграфов. Фильтрация проводится по наборам приемов, встречающихся в одних и тех же текстах (наборы текстов, реализующих эти приемы, полностью совпадают), через попарные проверки вложенности меньших приемов в большие (по мере увеличения числа вершин). Вложенные графы исключаются из дальнейшего рассмотрения.

5. Задача и методы обнаружения текстов с близкими стратегиями аргументации

Постановка задачи. Для выявления текстов с похожей аргументацией решается задача кластеризации аннотаций A с целью разбиения их на группы (кластеры) с номерами из множества $R = \{r_j\}$ ($j = 1, \dots, K_r$) таким образом, чтобы внутри одного кластера оказались аннотации a_i , близкие согласно некоторой метрике, позволяющей вычислять расстояние между ними, а различающиеся a_i попадали в разные кластеры. Обучающее подмножество множества A с заранее известными метками кластеров $\{r_j\}$ отсутствует. Задача сводится к установлению множества близких по приемам аргументации кластеров с номерами из R , оценке качества кластеров, их интерпретации. В качестве признаков, характеризующих a_i , в алгоритме кластеризации используются межтекстовые приемы Φ^{Ex} .

Методы кластеризации. Выбраны два дисперсионных метода кластеризации Ward и K-means, применены их реализации на Python из библиотеки scikit-learn.

Методы вычисления качества кластеризации. Не существует наилучшего формального критерия оценки качества кластеризации. В отсутствие вариантов кластеризации, проведенной экспертами, по рассчитанным значениям метрик сложно выбрать подходящий. В данной работе критерием качества кластеризации являлась близость (по составу a_i) кластеров, полученных разными методами.

6. Результаты экспериментов

Данные. В эксперименте участвовал корпус из 98 текстов научной коммуникации: 28 новостей науки (N), 20 статей Habr (H), 50 коротких (700–2000 слов) научных статей (S) двух тематических направлений (по 25 текстов по лингвистике и информационным технологиям). Аннотирование текстов проводилось с помощью инструментов платформы ArgNetBank Studio [Сидорова и др., 2020] пятью аннотаторами: один аннотировал все три жанра, двое работали с научными статьями, а двое – с набором научных новостей и статей Habr. Всего в корпусе насчитывается более 5 тыс. аргументативных утверждений (по жанрам: 2 502 – S , 1681 – H , 859 – N) и 4 тыс. аргументов (2 155 – S , 1415 – H , 758 – N). Для оценки согласия между аннотаторами при разметке научных текстов использовалась α Криппердорфа [Krippendorff, K., 2013], для утверждений согласие равно 0,67, для схем – 0,44. Согласие при разметке новостей и блогов рассчитано с помощью коэффициента, приводимого в [Сидорова и др., 2023], оно составило в новостях 0,48 и 0,70 для утверждений и схем соответственно, в статьях Habr – 0,31 и 0,67.

Анализ приемов аргументации. В этом разделе мы рассмотрим приемы с точки зрения универсальности употребления в подкорпусах корпуса и отдельных текстах, различая приемы по сложности, определяемой числом вершин-схем (далее – схем). Статистика по применению схем в корпусе ($\Phi_1(A)$) с указанием аббревиатур их полных названий представлена в таблице. В колонках под названиями жанров указана доля применения каждой схемы от числа всех используемых в данном жанре.

Авторы научных статей демонстрируют наибольшую последовательность в использовании моделей рассуждения – для обоснования большинства тезисов они обходятся меньшим разнообразием схем (15 схем, представленных в таблице, покрывают 97 % всех применяемых аннотаторами) по сравнению с авторами двух других жанров, несмотря на самый большой размер A_s , что, видимо, является следствием существующих стандартов в написании статей. Авторы Habr, возможно, для достижения убедительности, прагматично используют неформальный стиль, что определяет большое разнообразие применяемых схем рассуждения. Научные новости характеризуются средним (по сравнению с двумя другими жанрами) разнообразием в применении схем, так как журналисты, доводя результаты исследований до широкой аудитории, в определенной степени следуют стилистическим ограничениям при написании текстов.

Статистика применения схем в аннотациях A_g

Таблица 1

Table 1

Scheme frequencies in annotations A_g

Схема	Scheme	Abr.	S	H	N
через классификацию	VerbalClassification	VC	0,13	0,03	0,01
от примера	Example	E	0,12	0,09	0,05
от практической цели	PracticalReasoning	PR	0,12	0,05	0,06
от причины к следствию	CauseToEffect	CtE	0,11	0,20	0,11
от позитивных последствий	PositiveConsequences	PC	0,10	0,03	0,07
от взаимосвязи к причине	CorrelationToCause	CtC	0,09	0,02	0,03
от части к целому	PartToWhole	PtW	0,08	0,13	0,04
от применяемого метода	AppliedMethod	AM	0,05	0,02	0,07
от знака к означаемому	Sign	Si	0,05	0,06	0,05
от негативных последствий	NegativeConsequences	NC	0,05	0,03	0,01
от экспертного мнения	ExpertOpinion	EO	0,04	0,07	0,24
через конфликт	Logical Conflict	LC	0,01	0,06	0,03
подтверждение гипотезы	EvidenceToHypothesis	EtH	0,01	0,04	0,07
от следствия к причине	EffectToCause	EtC	–	0,02	0,06
от популярного мнения	PopularOpinion	PO	0,01	0,03	0,03
	<i>Все перечисленные схемы</i>	–	<i>0,97</i>	<i>0,88</i>	<i>0,93</i>

Если понимать под межжанровостью схем примерно равную долю их применения во всех жанрах, то такой схемой является Si. Она может указывать, в частности, на статистическую обоснованность утверждений и демонстрирует почти одинаковую относительную частоту в каждом из жанров (5–6 %).

Значительные различия в частоте используемых схем также позволяют выделить характерные для каждого жанра схемы. Три преобладающие схемы в научных статьях – VC, PR, E (всего охватывают 37 % аргументов) – демонстрируют почти одинаковую интенсивность использования (13 %, 12 %, 12 %). В текстах $Nabr$ такой доминирующей схемой является CtE (20 %): вывод причинно-следственных связей по частотности почти достигает суммы двух следующих схем PtW (13 %) и E (9 %), причем они в совокупности охватывают 42 % аргументов подкорпуса.

В свою очередь, почти четверть всех аргументов в новостях опирается в обосновании исключительно на авторитет ученых (EO, 24 %) с большим отрывом от второй по частоте схемы (CtE, 11 %) и трех, разделившие третье место: 7 %, PC, AM, EtH (56 % для этих пяти схем).

Еще одно различие между неформальными статьями $Nabr$, с одной стороны, и научными статьями и новостями – с другой, заключается в относительно активном использовании в первых полемической аргументации (LC, 6 %): нападки на другие аргументы в два раза реже встречаются в новостях (3 %), обычно это происходит в случае противоречий между цитируемыми исследователями или публикациями и почти отсутствует в коротких научных статьях, ограниченный объем которых не предполагает дискуссии с другими авторами.

Межжанровые приемы. Приемы из $\Phi_n(A) \in \Phi_g^{Ex} (n > 1)$, как и отдельные схемы, характерны в большей степени для одного или пары жанров. В исследуемом корпусе обнаружено всего 88 разных межжанровых приемов и только около четверти из них применяются во всех трех жанрах. Большинство из них – цепочки из двух схем (на рис. 1 схематически представлены наиболее частотные приемы с $n = 2$, где в начальной позиции (в вершине без исходящих связей) зафиксирована схема CtE со всеми встретившимися ее продолжениями и относительными частотами приемов в трех жанрах). Большая часть приемов (17) содержит в цепочке схему причинно-следственной семантики. Частота цепочки определяется жанровыми частотами составляющих ее схем. Далее в тексте для указания связи между двумя схемами будет использоваться стрелочка «←». Так, цепочка CtE ← CtE встречается почти в трети текстов жанра *N*, в двух третях текстов *H*, более чем в трети *S*, а цепочка CtE ← PO только в трети текстов *H*, в 11 % текстов *N* и 8 % текстов *S*, у цепочки PO ← CtE частотности применения еще ниже: в 7 % текстов *N*, 2 % текстов *H*, 4 % текстов *S*.

Приемы, свойственные текстам только двух жанров, распределились следующим образом: доля общих (от всех межтекстовых) для *N* и *H* составила 15 %, для *N* и *S* – 24 %, *H* и *S* – 36 %. Ожидаемо наибольшее число приемов объединяет тексты *Nabg* и научные статьи, самые частотные из них тяготеют к одному из жанров: рассуждение CtE ← CtC присуще в большей степени научным текстам (13 % *H*, 34 % *S*), в отличие от PtW ← (PtW, PtW, PtW) (40 % *H*, 8 % *S*). В линейной записи приемов с ($n > 2$) будем использовать следующие правила: схемы, связанные в цепочку, соединяются стрелочкой «←», а схемы, присоединяемые к одной и той же (корневой) параллельно, перечисляются в скобках, скобка же связывается с корнем стрелочкой: PtW ← PtW. Такое жанровое тяготение схем обусловлено меньшей конкретикой

рассуждений в научно-популярных статьях, публикуемых онлайн, и емкостью научных ввиду ограничений на объем текстов.

Самые частотные из приемов, общих для *N* и *H* (EtH ← EO (32 % *N*, 13 % *H*) и LC ← E (7 % *N*, 33 % *H*)), также демонстрируют тяготение к жанру в отличие от менее частотных, применяемых в каждом жанре в одинаковой степени (PR ← ConflictingGoals (11 % *N*, 13 % *H*)), среди них есть дерево из трех узлов: CtE ← (CtE, EO), 11 % *N*, 13 % *H*. Жанровая специфика новостей проявляется появлением в приемах специфических схем (EtH, ConflictingGoals и др.), редко или вовсе не используемых в двух других жанрах.

Общие приемы для текстов *N* и *S* содержат больше конструкций из трех схем, чем общие для *N* и *H*. Они низкочастотны и обладают более слабой корреляцией с жанром. Среди них лишь одна цепочка (PR ← CtE ← EO (7 % *N*, 10 % *S*)), остальные конструкции – деревья (к примеру, CtE ← (PC, PC) (7% *N*, 12% *S*)). Самые частые приемы из двух схем отличает присутствие в них схем, характерных для научных текстов, например, PR, AM: PR ← CtE (14 % *N*, 56 % *S*); AM ← PR (21 % *N*, 34 % *S*).

Следует отметить, что межтекстовые приемы могут повторяться внутри одного текста ($\Phi_n(A)$ из $\Phi_g^{Ex} \cap \Phi_g^{In}, n > 1$). Таких приемов 54, из них реализующихся во всех трех жанрах 13 %. Они состоят только из пар схем (см. рис. 2). Оставшиеся 87 % приемов характерны только для пары жанров: это все, кроме одной (PR ← AM ← PC; 4 % *N*, 2 % *S*), цепочки из двух схем и два дерева из трех схем (PtW ← (PtW, PtW) 13 % *H*, 4 % *S*; CtE ← (E, E) 13 % *H*, 2 % *S*). Сравнивая частоты, указанные на рис. 1 и 2, можно заметить, что не в каждом отдельном тексте межтекстовые приемы повторяются. Так, самый распространенный прием CtE ← CtE может повторяться внутри текста только в 7 % текстов *N* (а не во всех 29 %, где встречается как межтекстовый), в 47 % *H* (из 60 %), 12 % *S* (из 34). Треть межтекстовых приемов не повторяются ни в одном из текстов.

Специфичные жанровые приемы. Всего в корпусе выявлено 767 разных межтекстовых приемов с $n > 1$, из них в текстах только одного жанра: научных новостях – 56, статьях *Nabr* – 132, научных статьях – 579. Большое разнообразие приемов в подкорпусе *S* обусловлено не только количеством текстов, но и устойчивой вариативностью аргументации, отмеченной на уровне отдельных схем (реализацией меньшего числа уникальных схем по сравнению с другими жанрами при близких частотах наиболее частотных схем).

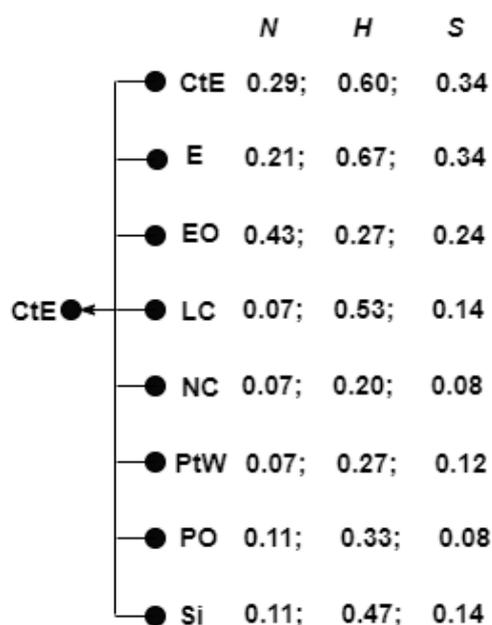


Рис. 1. Относительные частоты межтекстовых приемов ($n = 2$)
 Fig. 1. Relative frequencies of patterns repeating across texts ($n = 2$)

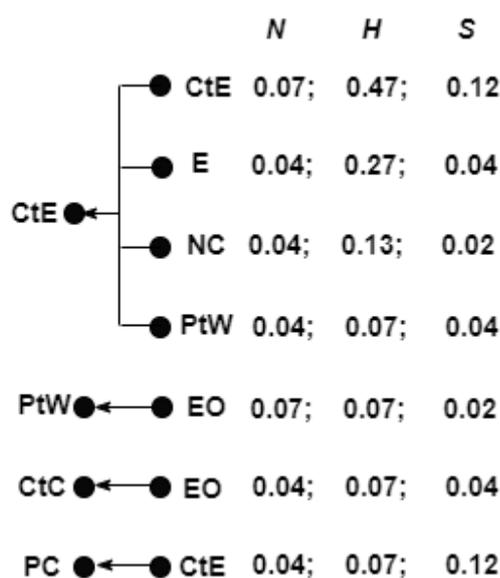


Рис. 2. Относительные частоты внутритекстовых приемов ($n = 2$)
 Fig. 2. Relative frequencies of patterns repeating in separate texts ($n = 2$)

Ограниченность влияния количества текстов проявляется при сравнении подкорпусов *N* и *H*: хотя новостных текстов больше, чем статей *Nabr* (28 против 20), приемов ($n > 1$) в них выявлено в 2 раза меньше. Разница в числе аргументов (748 и 1415) представляется менее значимой при сопоставлении *N* и *S*. В научных текстах аргументов в 1,5 раза больше (2155), чем в научно-популярных (что обусловлено большим числом текстов), однако приемов выявлено больше в 4 с лишним раза (579 и 132). Приведенная статистика отражает большую жанровую зависимость в разнообразии приемов от содержательной специфики характерной для жанра аргументации.

Указанная зависимость выявляется и при анализе специфичных приемов каждого жанра отдельно. Так, из 56 специфичных новостных приемов 21 содержит схему *EO*, 25 – схему *EtH*. Эти схемы реализуются при доказательстве тезисов через ссылку на эксперта либо обобщенное указание свидетельств некоторой гипотезы без детализации конкретных логических связей, как именно приведенные обстоятельства ее подтверждают. Как следствие, данные две модели не свойственны разветвленной сложной аргументации, ввиду чего новостные приемы ограничены в объеме: 43 % образованы двумя схемами, 45 % – тремя (против 31 % и 34 % для *H*, 13 % и 40 % для *S*).

Отличительной характеристикой приемов в статьях *Nabr* является активное использование разнообразной полемической аргументации: три наиболее частых приема с $n > 1$ содержат

разные схемы конфликтов (GeneralAcceptanceDoubt как опровержение аргументов от популярного мнения, SignFromOtherEvents как указание на нерелевантность статистики, Logical, обобщенная конфликтная схема по умолчанию, когда не подходят специализированные). Всего из 132 межтекстовых приемов в H более четверти (37, или 28 %) содержат как минимум одну конфликтную схему, причем 6 из них включают две, способные объединяться как последовательно (возражение на возражение), так и параллельно (одновременное опровержение тезиса двумя разными замечаниями, например, CtE \leftarrow (LC, LC)). Примечательна реализация в двух текстах Nabg цепочки из трех последовательных конфликтов LC \leftarrow LC \leftarrow LC, или демонстративное опровержение тезиса через приведение одного возражения с его показным отрицанием для акцентированного опровержения уже этого отрицания. Такая подчеркнутая двойная атака на тезис через его прямое опровержение вместе с опровержением возможного ответного возражения наглядно показывает актуальность экспрессивной полемической аргументации, усиливающей эмоциональную вовлеченность читателя, для статей Nabg по сравнению с двумя другими жанрами.

Научные статьи корпуса характеризуются реализацией приемов из большого числа схем, вплоть до 9: CtE \leftarrow ((PtW \leftarrow (VC, VC, VC, VC \leftarrow PR)), CtC, PtW). Такая комбинация схем, встречающаяся в двух разных текстах, не является исключением: 9 % из уникальных научных приемов образованы 5 схемами, 7 % – 6; 2,5 % – 7 и 8 (практически пятая часть из разных приемов в S образована 5 и более схемами). Повторяемость в разных научных статьях объемных логических структур обусловлена, во-первых, схожей организацией текстов через представление стандартных разделов и акцентуацию универсальных аспектов научных работ (актуальность, цель и задачи, методы, результаты, выводы, ...), а во-вторых, сопутствующим низким разнообразием схем аргументации ввиду нейтрального стиля, неуместностью применения нестандартных схем для привлечения читательского внимания.

Для жанровоспецифичных внутритекстовых приемов характерны те же тенденции, что и для межтекстовых. Так, в N из 29 внутритекстовых приемов 41 % образован схемой EO, 38 % – EtH. Активное повторение полемической аргументации свойственно для H и внутри текстов: 17 приемов из 67 (25 %) содержат схемы конфликтов (5 различных схем). Стоит отметить, что для научных статей относительно других жанров менее характерны внутритекстовые повторы, чем межтекстовые (154 и 579, по сравнению с 29 и 56, 67 и 132 для N и H). Это обусловлено лапидарностью доказательств в коротких научных статьях, нежелательностью параллельных избыточных повторов.

Стратегии аргументации. Кластеризация множества аннотаций A из размеченного корпуса S проводилась по межтекстовым приемам $\Phi^{Ex}(A)$. В компонентах вектора признаков учитывалось только присутствие/отсутствие приема. В качестве компонентов рассматривались варианты: а) $\Phi_1(A)$, б) $\Phi_2(A)$, в) $\Phi_n(A)$ ($n > 2$), г) $\Phi_n(A)$ ($n > 0$). Успешность кластеризации рассматривалась для параметра K_r (число кластеров), равного трем и пяти, исходя из количества текстов в кластеризуемом множестве. Следует отметить, что построенные методами Ward и K-means кластеры большей частью совпадают по текстовому составу, в том числе выделяются одинаковые аномальные кластеры. Такой результат позволяет считать качество проведенной кластеризации хорошим. Далее рассматриваются результаты кластеризации, полученные методом Ward.

Наилучшее разделение аннотаций по жанрам получается при использовании приемов всего спектра $\Phi_n(A)$ ($n > 0$) и $K_r = 5$. Число текстов, составляющих большинство представителей жанра в кластере, равно 89 (доля ошибок от объема корпуса – 9 %). Состав кластеров выглядит так: 0) 27(S) и 1(H), 1) 11(S), 2) 9(S), 3) 13(H), 4) 28(N), 6(H) и 3(S). Текст Nabg, попавший в 0) кластер научных статей, посвящен обзору открытых программных продуктов, используемых в решении задач макроэкономики, что сближает аргументацию в нем с применяемой в научных статьях. Разделение научных текстов на три кластера – 0), 1), 2) – вызвано применением разных приемов в текстах разной тематики: в 0) представлены преимущественно тексты по ИТ

с небольшой долей лингвистических, относящихся к компьютерной и корпусной лингвистике, в кластер 1) сгруппировались тексты главным образом по лингвистике, кластер 2) сформирован из текстов ИТ, которые отличает акцент на широко обсуждаемых передовых технологиях: облачных (3 текста), блокчейне и искусственном интеллекте (по 2 текста). В наиболее крупный кластер 4) вошли все новости, к которым присоединились 6 статей *Nabg* и 3 научные статьи с нетипично компактной аргументацией (высокой долей неаргументативного содержания). Так, статьи *Nabg* обращены не собственно к ИТ, а к смежным социальным вопросам, касающимся занятых в этой сфере работников. Соответственно, уже при кластеризации всего корпуса на формирование кластеров повлияли не только жанровые особенности текстов, но и тематические, воздействие которых на организацию аргументации может превосходить таковое от жанровой специфики.

Кластеризация текстов каждого жанра требует тщательного изучения и будет представлена в отдельной статье.

Выводы.

1. Межтекстовых приемов, реализующихся в текстах всех трех жанров, значительно меньше, чем специфичных для одного жанра. Их отличает низкая сложность (это, в основном, цепочки из двух схем). Частотные характеристики приемов указывают на тяготение большинства из них к определенному жанру.

2. Анализ приемов, выявленных в двух жанрах из трех, демонстрирует наибольшую близость аргументации в научных и научно-популярных статьях, наименьшую – в научно-популярных текстах и научных новостях.

3. Научные статьи характеризуются наибольшим разнообразием и частотностью приемов с $n > 1$, научно-популярные отличаются высокой активностью полемической аргументации, тогда как приемы в научных новостях образуются схемами, присущими компактным доказательствам (через ссылку на эксперта или обобщенное указание свидетельств), ввиду чего наиболее ограничены в объеме и вариативности.

4. Внутритекстовые приемы повторяют тенденции, установленные для межтекстовых приемов как между жанрами, так и для каждого жанра отдельно. За редким исключением в одном тексте повторяются несложные приемы.

5. Состав построенных кластеров свидетельствует об эффективности применения всей совокупности межтекстовых приемов ($n > 0$) для выявления близких по стратегиям аргументации текстов, как правило, принадлежащих одному жанру и/или одной тематической направленности.

Заключение

В статье проведен анализ приемов аргументации, автоматически извлеченных из аргументативных аннотаций русскоязычных текстов, принадлежащих трем жанрам междисциплинарной области знания – научной коммуникации (короткие научные статьи, научно-популярные статьи *Nabg* и новости науки). Тексты корпуса были размечены парами из 5 экспертов-лингвистов с согласием из диапазона «существенное» для каждого этапа разметки (аргументативных утверждений, аргументов, типов аргументов). Результат разметки – множество графов аргументации, для анализа редуцированных до графов, содержащих лишь типы аргументов. Из них методом частотного анализа подграфов (FSM), основанном на проверке подграфов с фиксированным числом вершин на изоморфизм, были извлечены повторяющиеся подграфы-приемы, для представления которых предложено формальное описание в виде спектра приемов: встречающихся только в одном тексте (для каждого текста), в корпусе текстов одного жанра, а также в разножанровых корпусах. Такое представление упростило анализ внутритекстовых, межтекстовых и межжанровых приемов и позволило рассмотреть стратегии написания текста в целом. Установлено, что приемов аргументации, свойственных всем трем

жанрам, значительно меньше, чем реализующихся в текстах двух или одного жанра. Показано, что на уровне используемых приемов наиболее схожа аргументация в научных и научно-популярных статьях, наименее – в научных текстах и научных новостях. Подробно представлена специфика приемов аргументации, характерных для каждого жанра отдельно. Эффективность применения межтекстовых приемов в совокупности (от схем до приемов-повторов максимальной сложности) для установления жанра текста подтвердила проведенная кластеризация корпуса, которая позволила с незначительной долей ошибок распределить тексты по кластерам для каждого жанра.

Список литературы

- Пименов И. С., Саломатина Н. В., Тимофеева М. К.** Формальное выявление приемов аргументации в научных текстах // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2022. Т. 20. № 1. С. 21–36. DOI 10.25205/1818-7935-2022-20-1-21-36
- Саломатина Н. В., Пименов И. С., Сидорова Е. А.** Распознавание в научных текстах аргументов с этосом и пафосом // Тез. Междунар. конф. «Марчуковские чтения – 2022», 3–7 октября 2022, Академгородок, Новосибирск, Россия. 2022. С. 140.
- Сидорова Е. А., Ахмадеева И. Р., Загорюлько Ю. А., Серый А. С., Шестаков В. К.** Платформа для исследования аргументации в научно-популярном дискурсе // Онтология проектирования. 2020. Т. 10, № 4(38). С. 489–502.
- Сидорова Е. А., Ахмадеева И. Р., Загорюлько Ю. А., Кононенко И. С., Серый А. С., Чагина П. М., Шестаков В. К.** Комплексный подход к анализу аргументативных отношений в текстах научной коммуникации // Онтология проектирования. 2023. Т. 13, № 4(50). С. 562–579.
- Al-Khatib K., Wachsmuth H., Kiesel J., Hagen M., Stein B.** A News Editorial Corpus for Mining Argumentation Strategies // Proc. of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016. P. 3433–3443.
- Al Khatib K.** Computational Analysis of Argumentation Strategies // Dissertation to obtain the academic degree of Dr. rer. nat. Faculty of Media Bauhaus-Universität Weimar, Germany, September 2019.
- Al-Khatib K., Wachsmuth H., Hagen M., Stein B.** Patterns of Argumentation Strategies across Topics // Proc. of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, September 7–11, 2017. P. 1351–1357.
- Anand P., King J., Boyd-Graber J., Wagner E., Martell C., Oard D., Resnik P.** Believe Me—We Can Do This Annotating Persuasive Acts in Blog Text // Proc. of the conf. Computational Models of Natural Argument, San Francisco, California, USA, August 7, 2011.
- Cordella L. P., Foggia P., Sansone C., Vento M.** A (Sub) Graph Isomorphism Algorithm for Matching Large Graphs // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004. Vol. 26, no. 10. P. 1367–1372.
- El Baff R., Wachsmuth H., Al Khatib K., Stede M., Stein B.** Computational Argumentation Synthesis as a Language Modeling Task // Proc. of the 12th International Conference on Natural Language Generation. Tokyo, Japan, 2019. P. 54–64.
- Hagberg A. A., Schult D. A., Swart P. J.** Exploring network structure, dynamics, and function using NetworkX // Proc. of the 7th Python in Science Conference (SciPy2008), Pasadena, CA USA, Aug 2008. P. 11–15.
- Jiang C., Coenen F., Zito M.** A Survey of Frequent Subgraph Mining Algorithms. // The Knowledge Engineering Review. 2004. Vol. 1. P. 1–31.
- Krippendorff K.** Content analysis: An introduction to its methodology, 3rd edition. Thousand Oaks, CA: Sage, 2013.
- Lauscher A., Wachsmuth H., Gurevych I, and Glavaš G.** Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation, arXiv: 2107.00281v3 [cs.CL] 8 Nov 2022.

- Mirzakhmedova N., Kiesel J., Al-Khatib K., Stein B.** Unveiling the Power of Argument Arrangement in Online Persuasive Discussions // Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2023) – Singapore, Singapore, 6–10 Dec, 2023. P. 15659–15671.
- Pimenov I. S., Salomatina N. V.** Identification of Scientific Texts with Similar Argumentation Complexity // 2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). 2022. P. 880–875.
- Rahwan I., Reed C.** The argument interchange format. Argumentation in artificial intelligence, ed. Rahwan I. and Simari G., Springer, 2009. P. 383–402.
- Ullmann J. R.** An Algorithm for Subgraph Isomorphism // Journal of the ACM. 1976. Vol. 23(1). P. 31–42.
- Wachsmuth H., Stede M., El Baff R., Al-Khatib K., Skeppstedt M., Stein B.** Argumentation Synthesis following Rhetorical Strategies // Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, New Mexico, USA, August 20–26, 2018. P. 3753–3765.
- Walton D., Reed C., Macagno F.** Argumentation schemes Fundamentals of critical argumentation. New York: Cambridge University Press, 2008. 443 p.
- Yan M., Lin Y., Litman D.** Argumentatively Phony? Detecting Misinformation via Argument Mining // ACM Trans. Graph. 2018. Vol. 37 (4). Article 111. <https://doi.org/10.1145/1122445.1122456>.
- Zukerman I., McConachy R., George S.** Using argumentation strategies in automated argument generation // In INLG'2000 Proc. of the First International Conference on Natural Language Generation. Mitzpe Ramon, Israel. Association for Computational Linguistics. 2000. P. 55–62.

References

- Al-Khatib K., Wachsmuth H., Kiesel J., Hagen M., Stein B.** A News Editorial Corpus for Mining Argumentation Strategies. *Proc. of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 3433–3443.
- Al Khatib K.** Computational Analysis of Argumentation Strategies. *Dissertation to obtain the academic degree of Dr. rer. nat. Faculty of Media Bauhaus-Universität Weimar, Germany, September 2019.*
- Al-Khatib K., Wachsmuth H., Kiesel J., Hagen M., Stein B.** Patterns of Argumentation Strategies across Topics. *Proc. of the 2017 Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark, September 7–11, 2017, pp. 1351–1357.
- Anand P., King J., Boyd-Graber J., Wagner E., Martell C., Oard D., Resnik P.** Believe Me—We Can Do This Annotating Persuasive Acts in Blog Text. *Proc. of the conf. Computational Models of Natural Argument, San Francisco, California, USA, August 7, 2011.*
- Cordella L. P., Foggia P., Sansone C., Vento M.** A (Sub) Graph Isomorphism Algorithm for Matching Large Graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, vol. 26, no. 10, pp. 1367–1372.
- El Baff R., Wachsmuth H., Al Khatib K., Stede M., Stein B.** Computational Argumentation Synthesis as a Language Modeling Task. In: *Proc. of the 12th International Conference on Natural Language Generation*. Tokyo, Japan, 2019, pp. 54–64.
- Hagberg A. A., Schult D. A. and Swart P. J.** Exploring network structure, dynamics, and function using NetworkX. *Proc. of the 7th Python in Science Conference (SciPy2008)*, Pasadena, CA USA, pp. 11–15, Aug 2008.
- Jiang C., Coenen F., Zito M.** A Survey of Frequent Subgraph Mining Algorithms. *The Knowledge Engineering Review*, 2004, vol. 000(1), pp. 1–31
- Krippendorff K.** Content analysis: An introduction to its methodology, 3rd edition. Thousand Oaks, CA: Sage, 2013.
- Lauscher A., Wachsmuth H., Gurevych I., Glavaš G.** Scientia Potentia Est – On the Role of Knowledge in Computational Argumentation, arXiv: 2107.00281v3 [cs.CL] 8 Nov 2022.

- Mirzakhmedova N., Kiesel J., Al-Khatib K., Stein B.** Unveiling the Power of Argument Arrangement in Online Persuasive Discussions. *Proc. of the conf 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)* – Singapore, Singapore, 6–10 Dec, 2023. pp. 15659–15671.
- Pimenov I. S., Salomatina N. V., Timofeeva M. K.** Formal Identification of Argumentation Patterns in Scientific Texts. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2022, vol. 20, no. 1, pp. 21–36. (in Russ.) DOI 10.25205/1818-7935-2022-20-1-21-36.
- Pimenov I. S., Salomatina N. V.** Identification of Scientific Texts with Similar Argumentation Complexity. *2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*, 2022, pp. 880–875.
- Rahwan I., Reed C.** The argument interchange format. *Argumentation in artificial intelligence*, ed. Rahwan I. and Simari G., Springer, 2009, pp. 383–402.
- Salomatina N. V., Pimenov I. S., Sidorova E. A.** Automatic identification of arguments with ethos and pathos in scientific articles. *Thesis of the International conf. Marchuk Scientific Reading – 2022, 3–7 October 2022*, Academgorodok, Novosibirsk, Russia, 2022. (in Russ.)
- Sidorova E. A., Akhmadeeva I. R., Zagorulko Yu. A., Kononenko I. S., Sery A. S., Chagina P. M., Shestakov V. K.** An integrated approach to the analysis of argumentative relationships in scientific communication texts. *Ontology of deSiing*, 2023, vol. 13, no. 4(50), pp. 562–579. (in Russ.)
- Sidorova E. A., Akhmadeeva I. R., Zagorulko Yu. A., Sery A. S., Shestakov V. K.** Research platform for the study of argumentation in popular science discourse. *Ontology deSiing*, 2020, vol. 10, no. 4(38), pp. 489–502. (in Russ.)
- Ullmann J. R.** An Algorithm for Subgraph Isomorphism. *Journal of the ACM*, 1976, vol. 23(1), pp. 31–42.
- Wachsmuth H., Stede M., El Baff R., Al-Khatib K., Skeppstedt M., Stein B.** Argumentation Synthesis following Rhetorical Strategies. *Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, New Mexico, USA, August 20–26, 2018*. pp. 3753–3765.
- Walton D., Reed C., Macagno F.** *Argumentation schemes Fundamentals of critical argumentation*. New York: Cambridge University Press, 2008, 443 p.
- Yan M., Lin Y., Litman D.** Argumentatively Phony? Detecting Misinformation via Argument Mining. *ACM Trans. Graph.* 37, 4, Article 111 (August 2018), 6 pages. <https://doi.org/10.1145/1122445.1122456>.
- Zukerman I., McConachy R., George S.** Using argumentation strategies in automated argument generation. In *INLG'2000 Proc. of the First International Conference on Natural Language Generation. Mitzpe Ramon, Israel. Association for Computational Linguistics*, 2000, pp. 55–62.

Информация об авторах

Иван Сергеевич Пименов, программист ИСИ СО РАН

Саломатина Наталья Васильевна, кандидат физико-математических наук, старший научный сотрудник ИСИ СО РАН

Тимофеева Мария Кирилловна, доктор филологических наук, ведущий научный сотрудник ИСИ СО РАН

Information about the Authors

Ivan S. Pimenov, Programmer, A. P. Ershov Institute of Informatics Systems

Natalia V. Salomatina, Candidate of Sciences (Physics & Maths), Senior Researcher, A. P. Ershov Institute of Informatics Systems

Mariya K. Timofeeva, Doctor of Sciences (Philology), Leading Researcher, A. P. Ershov Institute of Informatics Systems

*Статья поступила в редакцию 30.08.2024;
одобрена после рецензирования 04.12.2024; принята к публикации 06.12.2024*

*The article was submitted 30.08.2024;
approved after reviewing 04.12.2024; accepted for publication 06.12.2024*