

Научная статья

УДК 81'322.2

DOI 10.25205/1818-7935-2024-22-1-27-49

**Аргументативная разметка корпуса текстов
научной интернет-коммуникации:
жанровый анализ и исследование типовых моделей рассуждения
с помощью платформы ArgNetBank Studio**

**Мария Кирилловна Тимофеева¹, Дарья Владимировна Ильина²
Ирина Семёновна Кононенко³**

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

¹mtimof@inbox.ru, <https://orcid.org/0000-0001-8999-2330>

²dviljina@gmail.com, <https://orcid.org/0000-0001-6191-5632>

³irina_k@cn.ru, <https://orcid.org/0000-0001-5057-6807>

Аннотация

В статье представлен корпус текстов с аргументативной разметкой, относящихся к различным типам научной интернет-коммуникации. Разметка корпуса проведена на платформе ArgNetBank Studio. Обсуждаются жанры и функциональные характеристики текстов научной интернет-коммуникации. Сформулированы цель создания корпуса и принципы отбора включаемых в него данных, определены типы отбираемых текстов. Рассмотрены особенности аргументативной разметки научных и научно-популярных статей с комментариями, которые, в отличие от остальных типов включенных в корпус текстов, представляют собой интернет-диалоги. Это обуславливает специфику их аргументативной структуры. Обсуждается также проблема идентификации аргументативных схем, возникающая при разметке текстов разных типов. Проведенный пробный эксперимент показал, что для ряда пар схем выбор между ними затруднен. Предложено учесть сложности такого рода при разработке методики разметки, определив семантические различительные признаки схем аргументации. Данный подход проиллюстрирован на примере одной из рассмотренных пар схем. В ходе разметки аналитических текстов возникла необходимость добавления новых схем рассуждения и дополнения некоторых из уже имеющихся. В статье приведены формализованные описания таких схем и иллюстративные примеры из проанализированных текстов корпуса.

Ключевые слова

аргументативная разметка, жанр, корпус, научная интернет-коммуникация, модель аргументации

Финансирование

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

Для цитирования

Тимофеева М. К., Ильина Д. В., Кононенко И. С. Аргументативная разметка корпуса текстов научной интернет-коммуникации: жанровый анализ и исследование типовых моделей рассуждения с помощью платформы ArgNetBank Studio // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2024. Т. 22, № 1. С. 27–49. DOI 10.25205/1818-7935-2024-22-1-27-49

© Тимофеева М. К., Ильина Д. В., Кононенко И. С., 2024

Argumentative Annotation of the Scientific Internet-Communication Corpus: Genre Analysis and Study of Typical Reasoning Models based on the ArgNetBank Studio Platform

Mariya K. Timofeeva¹, Daria V. Ilina²,
Irina S. Kononenko³

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

¹mtimof@inbox.ru, <https://orcid.org/0000-0001-8999-2330>

²dviljina@gmail.com, <https://orcid.org/0000-0001-6191-5632>

³rina_k@cn.ru, <https://orcid.org/0000-0001-5057-6807>

Abstract

The article presents a corpus of texts with argumentative annotation pertaining to different types of scientific Internet communication. The corpus is annotated using the ArgNetBank Studio platform. The genres and functional characteristics of scientific Internet communication texts are discussed. The purpose of creating the corpus and the principles of selecting the data to be included are formulated, the types of selected texts are defined. The peculiarities of argumentative annotation of scientific and popular science articles with commentaries, which, unlike other types of texts included in the corpus, are internet-dialogues, are considered. This determines the specificity of their argumentative structure. We also discuss the issue of identifying argumentative schemes that arises within the process of annotating the texts of different types. A trial experiment has shown that for a number of pairs of schemes the choice is difficult. We propose taking these difficulties into account while developing a methodology of annotation and defining distinguishing semantic features of argumentative schemes; this approach is illustrated by analyzing the schemes from one of the considered pairs. In the course of annotating the analytical texts, the necessity of adding new reasoning schemes and supplementing some of the existing ones has arisen. The paper provides formalized descriptions of such schemes and illustrative examples from the analyzed corpus texts.

Keywords

Argumentative annotation, genres, corpus, scientific Internet communication, argumentation model

Funding

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

For citation

Timofeeva M. K., Ilina D. V., Kononenko I. S. Argumentative annotation of the scientific internet-communication corpus: genre analysis and study of typical reasoning models based on the ArgNetBank Studio platform. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2024, vol. 22, no. 1, pp. 27–49. (in Russ.) DOI 10.25205/1818-7935-2024-22-1-27-49

Введение

Современное общество характеризуют как информационное, существенную роль в нем играет информация и операции с ней. В частности, цифровая среда стала неотъемлемым средством поддержки, распространения и развития научных знаний. Создаются оперативные формы публикаций и открытые архивы научных статей, средства быстрого информационного обмена, происходит развитие информационной инфраструктуры науки, используются разнообразные виды электронной научной коммуникации.

В таких условиях особую значимость приобретает развитие средств формализации и автоматизации процессов работы с текстами научной интернет-коммуникации, прежде всего – работы с содержаниями текстов. Соответственно актуально развитие языковых моделей семантико-прагматического уровня, так как именно они связаны с пониманием текста.

Один из вариантов анализа, ориентированный на решение ряда задач компьютерной лингвистики (автоматизация процессов реферирования, семантического поиска, классификации текстов и т. д.) – это анализ рассуждения, реализуемого в тексте, т. е. выявление траекторий

движения мысли автора, направленных на адекватную поставленным целям передачу определенного содержания. Современные методы построения схем рассуждения базируются на исследованиях по анализу аргументации, представленных, например, в [Perelman, Olbrechts-Tyteca, 1971; Walton, 2008; Еемеерен, 2021]. Одним из инструментов компьютерной реализации данного подхода является платформа для обработки текстов на русском языке ArgNetBank Studio (<https://uniserv.iis.nsk.su/arg/>) [Сидорова и др., 2020].

Цель данной статьи состоит в обсуждении опыта построения и анализа схем рассуждения на материале текстов научной интернет-коммуникации, осуществляемой на русском языке. Особое внимание уделено научно-популярным статьям, размещенным на форуме «Хабр» (habr.com/ru) общим объемом 57 356 словоупотреблений. Для создания корпуса текстов с аргументативной разметкой использовалась указанная платформа.

Дальнейшее изложение имеет следующую структуру. В разделе 2 обсуждается создание корпуса текстов с аргументативной разметкой. Рассматривается область интернет-коммуникации и ее жанровое деление, требования к включаемым в создаваемый корпус текстам, структура этого корпуса и его количественные характеристики. Проанализированы результаты разметки двух типов статей: научных статей с комментариями рецензента и научно-популярных статей с комментариями. В разделе 3 описаны новые схемы рассуждения, включенные в используемый инструмент разметки. В заключении кратко формулируются итоги проделанной работы и намечаются дальнейшие планы.

1. Создание корпуса

1.1. Понятие интернет-коммуникации

Единого определения понятия «интернет-коммуникация» нет, более того, применительно к такому виду взаимодействия используются также другие наименования, устоявшиеся разграничения между которыми отсутствуют: компьютерно-опосредованная коммуникация, виртуальный дискурс, электронная / виртуальная / компьютерная / сетевая коммуникация, компьютерный / сетевой / виртуальный / компьютерно-медийный / электронный дискурс, интернет-дискурс. Причем варианты наименований со словом «коммуникация» предлагается использовать как общие названия этого вида взаимодействия (существующего наряду с другими видами: устной, письменной, печатной формами), а варианты со словом «дискурс» – применительно к определенной конкретной сфере, не обязательно компьютерно-опосредованной (например, персональный, научный, деловой дискурс) [Щипицына, 2009].

Во многих публикациях приводится определение И. Н. Розиной. Она предлагает следующее толкование термина *компьютерно-опосредованная коммуникация*: «...новое прикладное направление, в котором исследуется использование людьми электронных сообщений (чаще мультимедийных) для формирования знания и взаимопонимания в разнообразных средах, контекстах и культурах» [Розина, 2007. С. 233].

Подходы к классификации речевых жанров интернет-коммуникации пока не привели к построению общепринятой концепции. Отчасти это объясняется отсутствием единства и в определении понятия *жанр* применительно к традиционным (не компьютерным) видам коммуникации. К числу наиболее признаваемых определений можно отнести формулировки М. М. Бахтина и Т. Г. Винокура: «...каждая сфера использования языка вырабатывает свои относительно устойчивые типы таких высказываний, которые мы и называем речевыми жанрами» [Бахтин, 1996. С. 237]; каждый жанр «характеризуется типическими ситуациями, формальными и содержательно-тематическими особенностями общения» [Винокур, 1980. С. 117].

Для научной интернет-коммуникации также пока нет сложившейся классификации жанровых моделей. Однако публикации на эту тему имеются. Например, в работе [Данкова, Дубров-

ская, 2019] анализируются профессиональные сообщества, существующие в сети ВКонтакте, рассматриваются состав и жанровые формы актуализации действующих в них аксиологических ориентиров. В результате выделены следующие жанры: самопрезентация, обмен мнениями, просьба, обращение за советом, совет, благодарность, осуждение, обвинение, оскорбление, одобрение. Уже по данному перечню понятно, что типология жанров в разных видах научной интернет-коммуникации различна. Например, в проанализированных научно-популярных статьях, а также в комментариях-дискуссиях по поводу их содержания не встретились такие жанры из приведенного перечня, как обвинение и оскорбление, что, по-видимому, объясняется более строгими принципами модерирования по сравнению с сетью ВКонтакте. То же самое можно сказать и о закономерностях, которые считаются присущими языку вообще, на данном этапе его функционирования, и в разной степени проявляются в современной интернет-коммуникации [Интернет-коммуникация, 2018. С. 6–8]: тенденция к усилению диалогичности; тенденция к усилению коллоквиализации (разговорности) общения; тенденция к экспрессивизации общения; тенденция к глобальному снижению и огрублению речи.

1.2. Жанры интернет-коммуникации и характеристика создаваемого корпуса

Любой текст, если его коммуникативная задача не сводится к чистому информированию (объяснительный текст), несет наряду с информированием функцию убеждения аудитории в справедливости высказываемых в нем положений и/или необходимости совершения определенных действий (аргументативный текст). Именно такими являются тексты, относящиеся к научной сфере общения, или функциональному стилю – Ф-стилю, в соответствии с терминологией [Кибрик, 2009].

Объектом исследования предлагаемой работы являются интернет-тексты, которые относятся к нескольким жанрам научного Ф-стиля и собраны в качестве жанровых подкорпусов единого корпуса научной интернет-коммуникации, представленного на платформе ArgNetBank Studio. В работе [Кибрик, 2009] жанровая проблематика рассматривается в тесной связи с классификацией функциональных стилей, отражающих сферы человеческой деятельности. В рамках научного Ф-стиля можно выделить два субстиля: собственно научная и научно-популярная коммуникация. К целевой аудитории текста научно-популярного субстиля относится самая широкая массовая публика, которую автор текста считает необходимым убедить в истинности изложенной в тексте информации. В предлагаемой работе научно-популярная коммуникация представлена новостными статьями (научные новости) и статьями с сайта habr.com/ru (далее – *habr-статьи*). К собственно научному субстилю отнесены короткие научные статьи, научные рецензии и научные статьи с комментариями рецензента. Их целевая аудитория существенно уже – это специалисты в научной области, к которой относится тематика текстов. Коммуникативная цель автора – убедить коллег в справедливости излагаемых им результатов научных исследований и/или высказать свое мнение, оценку этих результатов и способа их изложения.

Коллекция текстов научной интернет-коммуникации распределилась в корпусе по указанным субстилям и жанровым категориям следующим образом:

- научно-популярная коммуникация:
 - научные новости (30 текстов, средний объем текста – 506 словоупотреблений),
 - *habr-статьи* (30 статей с комментариями, средний объем статьи без комментариев – 1912 словоупотреблений, с комментариями – 16 482);
- научная коммуникация:
 - короткие научные статьи (50 статей, средний объем статьи – 1053 словоупотребления),
 - научные рецензии (30 текстов, средний объем рецензии – 346 словоупотреблений),
 - научные статьи с комментариями рецензентов (10 текстов, средний объем статьи без комментариев – 3571 словоупотребление, с комментариями – 4848).

Аннотирование текстов проводилось на основе модели Д. Волтона [Walton et al., 2008]. Эта работа содержит компендиум схем аргументов, которые, по мнению авторов, являются наиболее полезным и широко признанным инструментом, разработанным к настоящему времени в теории аргументации. Под схемой аргумента понимается модель, с помощью которой аргументирующий «стремится так или иначе добиться “переноса приемлемости” с эксплицитной посылки на свою точку зрения, ... чтобы слушатель или читатель принял эту точку зрения...», «конвенциональные способы выражения связи между тем, что утверждается в посылке, и тем, что утверждается в тезисе» [Еемерен, 2006. С. 25]. Схемы аргументов отражают стереотипные модели мышления, которые в общем случае не являются дедуктивными и монотонными. Поэтому их также называют моделями правдоподобного вывода, в том смысле, что если посылки аргумента верны, то предположительно верно и заключение. Схемы описываются в онтологии как набор утверждений-посылок, утверждение-вывод (тезис) и (опционально) набор презумпций и исключений. Интерфейс платформы представлен на рис. 1.

Роль	Тип утверждения	Описание утверждения
Conclusion	PresumedCause_Statement	В истинно в этой ситуации
Conclusion	Scheme_Application	
ExistenceOfSign_Premise	ExistenceOfSign_Statement	А истинно в данной ситуации
IndicationBySign_Premise	IndicationBySign_Statement	Как правило, истинность знака А, обозначающего В, указывает на истинность В
SignEventCorrelation_Presumption	SignEventCorrelation_Statement	Есть строгая корреляция между знаком и обозначаемым событием
SignFromOtherEvents_Exception	SignFromOtherEvents_Statement	Имеются другие события, более надёжно объясняющие знак

Первоисточник	Проект	Аргумент	Контекст
Виртуальное моделиро...	AZ-1	A19. Sign	S19. В таблице 1 ... S20. соотношение кейсов ...
Зачем люди придумыва...	БорзиковАлексей ...	От знака	Логлан — это язык, абсолютно лишенный двус...
Зачем люди придумыва...	БорзиковАлексей ...	От знака	Язык ифкуль для увеличения (применим) ин...
Зачем люди придумыва...	БорзиковАлексей ...	От знака	Подобного рода объединения существуют по ...
Зачем люди придумыва...	БорзиковАлексей ...	От знака	большинство слов имеют очень широкий спек...

Рис. 1. Схемы аргументации на платформе ArgNetBank Studio
Fig. 1. Argumentation schemes on the platform ArgNetBank Studio

Процесс аргументативной разметки текста включает следующие этапы: 1) разделение текста на текстовые сегменты, представляющие собой единицы рассуждения; 2) исключение из рассмотрения фрагментов текста, не вносящих вклад в ход рассуждения; 3) выявление смысловых связей между сегментами; 4) установление типов выявленных связей между текстовыми сегментами. Единицей рассуждения обычно является клауза, предложение, абзац. Перечисленные этапы, как правило, осуществляются не последовательно, а параллельно, при значительном объеме анализируемого текста возможен итеративный процесс. Взаимосвязи сегментов представляются в виде ориентированного графа с двумя типами вершин: 1) текстовые вершины с пометкой в виде текстового фрагмента (имеют вид прямоугольника) и 2) вершины, помеченные типами отношений (имеют вид овала). Внешний вид графа показан на рис. 2–5. Соответствующие типам отношений схемы аргументации описаны в онтологии (см. рис. 1), там же можно найти примеры их использования при разметке текстов корпуса.

Состав корпуса обусловлен целями его создателей, одной из которых было максимальное (по возможности) жанровое разнообразие текстов, которые будут подвергаться аргументативному аннотированию и в дальнейшем использоваться для нейросетевого анализа аргументации.

Второй принцип при сборе коллекции заключался в том, чтобы жанровые категории были максимально тематически нейтральными, что можно обеспечить многообразием тем в рамках каждого жанра. В связи с этим собраны новости из разных тематик, при отборе *habr*-статей старались также не ограничиваться преобладающей на сайте тематикой IT.

Тексты каждой категории характеризуются собственными жанровыми особенностями. Так, короткие научные статьи, обладая относительно небольшим размером (в среднем 1053 словоупотребления), как правило, имеют один главный тезис в конце текста и не имеют «длинных» связей (т. е. связанные аргументативными отношениями утверждения располагаются недалеко друг от друга), в них практически отсутствуют повторы, кроме того, многие тексты характеризуются эксплицитной жанровой структурой (введение, обзор, цель работы, методы, заключение, выводы), в других жанровая структура выявляется с помощью маркеров-индикаторов структуры [Саломатина, 2006]. В рецензиях главный тезис тоже находится в конце, он сводится к трем вариантам: принять рецензируемую статью к публикации, принять после доработки или отклонить. Аргументативные отношения образуют кустовую структуру, аргументами первого порядка являются достоинства и/или недостатки статьи. Главный тезис научных новостей, напротив, располагается в начале (лиде) новостной заметки, для заметки характерны повторы разной степени общности, имеется четкая жанровая структура с выделением хэдвайла, лида и бэкграунда [Кононенко, 2016]. Более сложные по структуре *habr*-статьи и статьи с комментариями могут вообще не иметь главного тезиса, а комментарии к ним привносят в изложение черты диалога.

Исходя из анализа текстов научной коммуникации данного корпуса, можно говорить о нескольких факторах определения жанра:

- 1) коммуникативная задача/цель;
- 2) жанровая структура;
- 3) наличие главного тезиса;
- 4) наличие повторов;
- 5) наличие комментариев;
- 6) статус участников коммуникации.

А. А. Кибрик пишет о разных аспектах определения жанра: «Первый из таких способов можно назвать **экстралингвистическим**. Он был предложен одним из наиболее известных современных специалистов по жанру – британско-американским лингвистом Дж. Суэйлсом. По определению Суэйлса [Swales 1990]¹, жанры – это принадлежность дискурсивных сообществ. По выражению Суэйлса, сообщества владеют жанрами. Дискурсы одного жанра имеют общий набор **коммуникативных целей**, признаваемых дискурсивным сообществом» [Кибрик, 2009. С. 10]. При анализе факторов, имеющих значение для идентификации жанра текста, выясняется, что структура коммуникативных задач, описанная в статье [Кононенко, 2014] как фактор жанровой классификации интернет-сайтов, актуальна и для выделения жанров интернет-текстов.

Коммуникативная задача описывается на основе параметров коммуникативной ситуации: текст можно рассматривать как коммуникативный акт (далее – КА), осуществляемый автором (далее – А) в рамках объемлющей деятельности (далее – Д), субъектом которой является А, контрагентом – целевая аудитория (далее – ЦА). Автор передает ЦА определенную информацию, которая является отражением тематики текста (далее – Т) и онтологии деятельности (да-

¹ Данный источник процитирован в работе А. А. Кибрика, работа не включена в список литературы к настоящей статье.

лее – Д). Текст нацелен на выполнение определенных *коммуникативных задач* (далее – КЗ), которым соответствуют *реактивные задачи* (далее – РЗ) для ЦА.

Таблица 1

Иерархия коммуникативных задач

Table 1

Hierarchy of communicative tasks

<i>Коммуникативные задачи</i>		<i>Реактивные задачи</i>	
1. Фатическая КЗ	наладить общение с ЦА	1. Фатическая РЗ	ЦА становится участником общения (вступает в / поддерживает общение)
2. Иллокутивная КЗ	передать ЦА информацию о Т и/или Д (в частности, с целью получить от ЦА информацию, необходимую для реализации Д)	2. Когнитивная РЗ	
		2.1. ЦА усваивает информацию о Т и/или Д, следствием чего может быть выполнение Реактивных задач 2.2–4	
		2.2. ЦА активизирует или вырабатывает определенное отношение к Т и/или Д (при наличии прагматической задачи А стремится к позитивной оценке ЦА относительно Д)	
		3. Иллокутивная РЗ	
3. Перлокутивная КЗ	обеспечить РЗ, т. е. желаемую реакцию ЦА	3.1. ЦА передает А информацию о Т, необходимую для реализации КЗ	ЦА совершает некоммуникативное действие Д', связанное с Т и встроенное в структуру Д
		3.2. ЦА передает А отношение к Т и/или оценку Д	
4. Прагматическая РЗ			

КЗ и РЗ выстраиваются в иерархию, в которой реализация задачи более низкого уровня является предпосылкой выполнения задачи более высокого уровня. Первые две (низкоуровневые) задачи характерны для любого КА. В зависимости от структуры задач, определяющей конечные цели и мотивы, рассматриваемые нами тексты по их назначению можно разделить на три жанровые группы.

Информационный текст: структура КЗ включает РЗ1–РЗ2. Нацелен на информирование, изменение знаний, мнений ЦА по тематике Т, исходит из предварительной общей гипотезы об интересе к Т у ЦА; корректировки представлений о ЦА не происходит, что является характерной чертой монологического изложения. Не предполагает немедленных реакций ЦА. К этому типу в корпусе относятся короткие научные статьи и научные новости.

Диалогический текст: структура КЗ включает РЗ1–РЗ3. Нацелен не только на информирование ЦА, но и на получение обратной связи. Это позволяет собеседникам в реальном времени достраивать и корректировать свои знания о Т и ЦА. Размещается на сайтах, ориентированных на организацию общения, что типично для блогов и форумов. Технологически такие сайты в наибольшей степени обеспечивают возможность диалогического или полилогического общения (в зависимости от числа коммуникантов) в стиле устной коммуникации. В частности,

возможны различные виды обмена репликами. Компонентный диалог представляет собой обмен стимулирующей и реагирующей репликами. Каскадный диалог представляет собой обмен несколькими репликами, каждая из которых реагирует на предыдущую. При этом возможны смены участников общения. К типу диалога в корпусе относятся *habr*-статьи, примеры и анализ которых будут приведены в разделе 2.4.

Деловой текст: структура КЗ включает Р31–Р34. Ориентирован на организацию совместной деятельности. Занимает промежуточное положение в отношении адекватности представлений об адресате. Обеспечивает обратную связь с ЦА для уточнения некоторых аспектов, необходимых для успешной реализации целей деятельности Д, которую обслуживает данный текст. Как и в первом случае, образ ЦА весьма обобщен, вследствие чего коммуникативный процесс в целом характеризуется высокой степенью стереотипности. Особенностью делового общения является нацеленность не просто на информирование, но на создание условий для реализации прагматической задачи Р34. Примерами текстов данного типа являются рецензии и научные статьи с комментариями рецензента, реализующими компонентный диалог.

1.3. Разметка текстов научных статей с комментариями рецензента

В группе деловых текстов выделяется жанр научных статей с комментариями. В них контрагентом автора текста является рецензент, который редко вербализует согласие, но регулярно эксплицирует отрицательную оценку и рекомендации по исправлению исходного текста. По объему комментарии составляют существенно меньшую часть текста, чем сама статья. Учитывая деятельностный характер текстов этого жанра, очевидно, что побудительные ходы у рецензента преобладают и встречаются очень часто, существенно чаще, чем в жанре аналитических статей. Рецензент выделяет фрагмент текста, на который он хочет высказать реакцию. На платформе это выглядит как повтор фрагмента авторского текста (используется отношение эквивалентности), который представляет часть аргументации автора и связывается с другими частями его аргументации. В целом в этом жанре наблюдается много побудительных реакций, что связано с прагматической задачей «совершенствования текста» и статусом «доминирования» рецензента над автором. Среди побудительных реакций очень много реакций типа «косвенное побуждение», которые описываются в рамках схемы «К практическому выводу» (PracticalReasoning)² (см. раздел 2.1): *ссылки; повтор; ссылки на всех; много отсылок к тому, что будет сказано*. Нередко постановка прагматической цели рецензентом аргументируется. На рис. 2 представлен фрагмент, где имеется несколько побудительных комментариев рецензента и выделен один, который сопровождается кратким объяснением позиции рецензента.

Как видим, рецензент показал стандартность («невычурность») текста с помощью рассуждения по аналогии (Analogy) и предложил поставить и выполнить цель «удалить данное замечание» (*данное замечание излишне*) в рамках все той же схемы «К практическому выводу».

Рассмотрим более простой случай из того же текста. Львиную долю имеющихся комментариев в нем составляют реплики рецензента «ссылки». В рамках той же схемы «К практическому выводу» автору предлагается сделать следующее: если есть план написать данный фрагмент, то следует поставить цель «дать ссылки» как способ реализации плана и реализовать эту цель действием «дать ссылки». Таким образом, данная цель и соответствующее действие представлены косвенно: «ссылки».

Иногда рецензент вступает в прямой конфликт с аргументацией автора текста.

²Здесь и далее при первом упоминании аргументативной схемы в скобках дается ее название на английском языке, представленное на платформе.



Рис. 2. Компонентный диалог в статье с комментариями рецензента
Fig. 2. Component dialogue in the article with reviewer comments

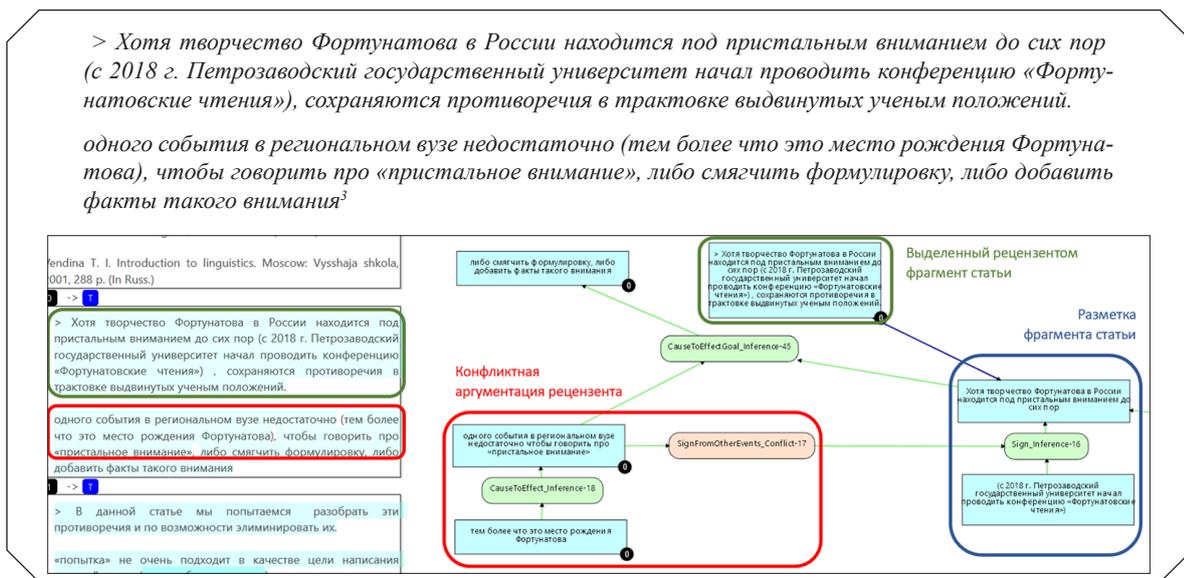


Рис. 3. Конфликтная аргументация в статье с комментариями рецензента
Fig. 3. Conflict argumentation in the article with reviewer comments

Разметка представлена на рис. 3.

Автор признаком своего утверждения о наличии в России особого внимания к творчеству Фортунатова посчитал конференцию в ПГУ (схема «От знака» (Sign)), а рецензент противопоставил этой аргументации конфликт типа «Знак других событий» (SignFromOtherEvents) и предложил поставить и реализовать одну из двух целей по корректировке текста. Проблема представления конфликтов состоит в том, что аргументация автора, видимо должна быть повторена в дублие фрагмента текста, так как конфликтная схема напрямую связывается с аргументом, которому она противопоставляется.

³Цитаты приведены с сохранением орфографии и пунктуации авторов.

1.4. Разметка текстов научно-популярных статей с комментариями

Вторая из выделенных выше трех жанровых групп научных текстов (диалогическая) заслуживает отдельного рассмотрения, так как отнесенные к ней тексты статей с комментариями (habr-статьи) существенно превышают остальные тексты корпуса по объему, а также по количеству участников и разнообразию их профессиональной принадлежности (фактически не имеющих ограничений). Далее обсуждаются результаты аргументативной разметки этих статей.

Аргументативная разметка текстов осуществлялась на платформе ArgNetBank Studio. Приводимые далее иллюстративные примеры взяты из статьи «Самый научный гайд по сну» и ее обсуждения на веб-сайте habr.com/ru. Объем всего текста 10 192 словоупотребления, из них 3017 словоупотреблений – это сама научно-популярная статья, 7175 словоупотреблений – комментарии к ней (133 комментария), в построенном графе 486 текстовых сегментов. Далее в скобках перед комментарием указывается его номер в публикации, размещенной на указанном выше сайте. Поскольку комментарии также подвергались аргументативному анализу, приведенный в примере текст может соответствовать части комментария с указанным номером.

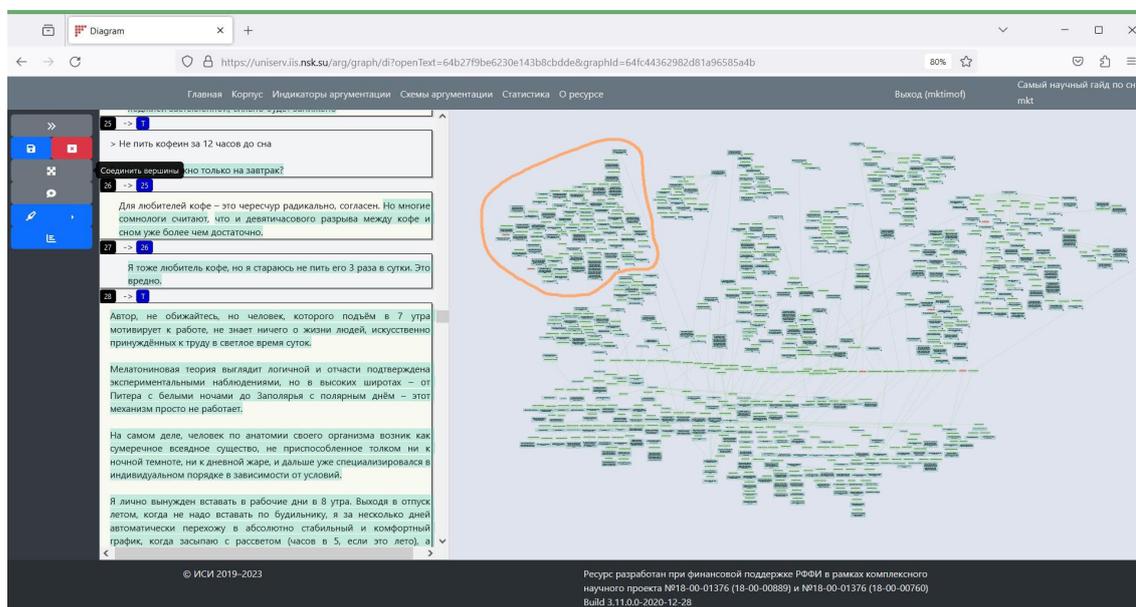


Рис. 4. Граф для текста «Самый научный гайд по сну»
Fig. 4. Graph for the text “The most scientific sleep guide”

Граф, представляющий модели рассуждения в указанном тексте, изображен на рис. 4. Этот граф является связным, т. е. не содержит изолированных текстовых вершин. Фрагмент графа, соответствующий статье, находится в нижней части изображения, в верхней его части – поддережья, соответствующие комментариям. Между ними – выстроенные в линию аргументативные отношения, связывающие комментарии с определенными утверждениями статьи. Комментарии образуют связные подграфы, каждый из которых присоединен к основному тексту статьи через то или иное отношение. По размеру подграфов можно увидеть, какие утверждения статьи вызвали наиболее активные обсуждения комментаторов. Для примера выделен один из таких подграфов, соответствующий дискуссии комментаторов по поводу вопроса «во сколько ложиться спать и вставать?». Каждый такой подграф имеет главную вершину, которая не является аргументом никакого другого утверждения комментария, а непосредственно связана с утверждениями самой статьи. В выделенном подграфе это отношение «От предвзято-

сти» (Bias), т. е. эмоционально-оценочное указание комментатора на то, что автор высказывает в статье свое предвзятое мнение по обсуждаемому вопросу, не учитывая опыта других людей. Чаще всего главная вершина поддерева комментариев связана со статьей через модели аргументации «От знающего» (PositionToKnow, 8 случаев); использовались также другие схемы: «Опровержение гипотезы» (FalsificationOfHypothesis, 6 случаев), «От неясности вербальной классификации» (VagueVerbalClassification, 5 случаев), «От исключения» (ExceptionalCase, 4 случая), «От применяемого метода» (AppliedMethod, 3 случая), по 1–2 раза встретились отношения «От незнания» (Ignorance), «От части к целому» (PartToWhole), «Логический конфликт» (LogicalConflict), «От предвзятости», «От следствия к причине» (EffectToCause), «К методу» (ToMethod), «От примера» (Example), по правилу гипотетического силлогизма (ModusPonens), «От негативных следствий» (NegativeConsequences), «К практическому выводу», «От знака».

Тексты жанра научно-популярной статьи с комментариями обладают рядом особенностей по сравнению с научными и другими научно-популярными текстами. В комментариях встречаются такие типы составляющих, которые либо отсутствуют в основном тексте, либо обладают меньшей частотностью, иными характеристиками содержания и функциональности. Эти особенности текста комментариев можно сгруппировать в пять классов.

1. Выражение отношения комментатора к какому-либо компоненту рассуждения или автору текста. Это может быть благодарность, оценка, эмоция, ирония.

В научном тексте может быть выражена благодарность, но она обычно относится к источнику финансовой поддержки, который лишь косвенно можно трактовать как аргумент «От эксперта» (ExpertOpinion), высказанный в поддержку представленной в статье работы. В комментариях к научно-популярному тексту благодарность, высказанная комментатором, может выполнять оценочную функцию, выступая как аргумент «От знающего»: (88) *Полезная статья. Спасибо.* Здесь комментарий выполняет функцию поддержки следующего тезиса, которым автор статьи заключил свой текст: *Надеюсь, я сделал вашу жизнь чуточку лучше.* Аналогичная реплика (42) *спасибо, приятная статья* поддерживает аргументацию другого комментатора.

Если сопоставить две похожие схемы рассуждения «От эксперта» и «От знающего», то можно заметить, что второй вид схем более характерен для комментария, чем для основного текста. По своему содержанию эти две схемы различаются тем, что первая из них апеллирует к признанному эксперту в обсуждаемой области знания, а вторая – к человеку, который рассуждает на основе своего собственного субъективного опыта, подтверждая обсуждаемые идеи или опровергая их. Типичный пример реализации такого отношения – комментарий (50) *Я долго пытался найти способ лучше высыпаться.* Этот комментарий предваряет аргумент комментатора, использующего свой личный опыт для обсуждения определенного тезиса статьи.

Оценка может совмещаться не только с благодарностью, но и с выражением эмоции, усиливая аргументативную схему, в реализации которой они участвуют. Например, фрагмент (84) *А первый раз на нем помните? О, Боже, это был мрак,* характеризующий личный опыт комментатора, выполняет аргументативную функцию в составе схемы типа «От негативных следствий»; наряду с данной схемой в поддержку того же тезиса реализована другая схема «От позитивных следствий», также имеющая весьма эмоциональное языковое воплощение: (83) *Неистово плюсюю!!!!*

Эмоции могут присоединяться и к другим видам отношений, например, в схеме типа «Опровержение гипотезы»:

(102) *Нету у меня 24-часового циркадного ритма. Нету. Сколько же еще можно спекулировать на эту тему?*

Данный комментарий к основному тексту фальсифицирует высказанный автором статьи тезис *Циркадные ритмы составляют 24 часа, и для хорошего сна нужно жить в соответствии с ними.*

По сравнению со схемами рассуждения, имеющими эмоционально-оценочную окраску, иронично-шутливые комментарии встречаются значительно реже, аргументативно значимых среди них не оказалось. Например, комментарий (128) *Какие смешные эти ПТУшники... САМЫЙ НАУЧНЫЙ гайд по сну... В Академию наук еще не позвали?* не включен в общий граф рассуждения, так как выполняет метаязыковую, а не аргументативную функцию. В этом фрагменте можно усмотреть пейоративную оценочность, тем не менее неясно, относится ли она к содержанию основного текста (и тогда данный фрагмент надо было бы признать аргументативным) или является чисто метаязыковой.

2. Выражение мысли в форме вопроса. В комментариях вопрос нередко используется при реализации разных схем рассуждения, в частности, для указания на неполноту знания, на нечеткость понятий, наличие исключений, т. е. в схемах рассуждения типа «От незнания», «От неясности вербальной классификации», «От исключения». Например:

(131) *Есть ли идеи как бороться короткое просыпание которое на графике обозначено в р-не 6 утра?* В этом случае вопрос направлен против безальтернативно сформулированного в статье тезиса о необходимом для всех людей режиме сна. Реализуя схему «От незнания», автор комментария указывает на незнание автором статьи определенной информации.

(133) *Чем замерять качество сна? Всяческие часы/браслеты с пульсометром реально работают или нет?* Данный вопрос реализует схему «От неясности вербальной классификации», выявляя нечеткость тезиса о проблемах со сном, так как для обнаружения проблемы надо иметь средства оценки качества сна.

(129) *Мне вот одно интересно – как многие люди, и известные, и не очень – умудряются спать по 4 часа? И при этом вполне «шурупать мозгами» и вполне доживают до старости.* Вопрос здесь реализует схему «От исключения» и направлен против тезиса о том, что взрослым нужно спать по 7–8 часов, подросткам – около 9.

Другие виды схем рассуждения также могут реализовываться в форме вопроса.

3. Использование дискурсивных анафор, отсылающих к фрагментам статьи или комментариев, в частности, своих собственных, высказанных ранее. Отсылки к предыдущим частям текста характерны не только для текста самой статьи, но и для комментариев к ней. В приведенных ниже примерах знаком > («больше») обозначены цитаты, на которые отвечает комментатор, при этом предложение или реплика может цитироваться полностью или частично. В двойных квадратных скобках указываются гиперссылки.

(11) *Лично у меня, по субъективным ощущениям, весь эффект кофеина полностью проходит часов через 6–7.*

В следующем комментарии данная фраза цитируется частично:

(12) > *эффект кофеина*

А как его почувствовать? Ни растворимый, ни зерновой кофе не вызывает у меня никакого эффекта. Так, вкусенький напиток.

В примере из комментария (40) цитируется полная фраза из комментария (33):

(33) *Общеизвестно, что волк – ночной хищник.*

(40) > *Общеизвестно, что волк – ночной хищник. серьезно?*

Вместо цитирования в дискурсивной анафоре может использоваться ссылка, например:

(56) *Выше [[задал]] этот вопрос в формате предположения о сдвиге режима...*

Возможны отсылки к частям текста без указания точной локализации. В такой форме обычно высказываются возражения на определенную авторскую мысль, которую комментатор предлагает реконструировать самостоятельно. Такова, например, отсылка к тексту в приведенном выше комментарии (102). Еще один пример – высказывание комментатора (50) *Выдержи-*

вать строго график отхода ко сну и подъема невыполнимо и поэтому ненужно. Для нахождения фрагмента статьи, на который среагировал данный комментатор, необходим смысловой анализ, что не всегда является тривиальной задачей, особенно при большом объеме текста. В данном случае комментатор использовал схему типа «Опровержение гипотезы», реагируя на рекомендацию автора статьи *Установите себе будильник и обозначьте время засыпания и подъема. Взрослым нужно спать по 7–8 часов, подросткам [[около 9]]*.

4. Функциональная нестабильность. В одном и том же тексте разные люди могут усматривать разные смысловые составляющие и приписывать этим составляющим разные функции.

Для авторского текста характерна целостная логика изложения, в которой каждая часть текста имеет определенное назначение, выполняет определенную функцию, в частности, входя в состав темы или ремы. Кроме того, менее значимые или менее новые мысли при их выражении в авторском тексте могут оказаться в менее заметной позиции. Например, могут войти в содержание пресуппозиции, явно не выражаемой в тексте, или реализоваться в виде свернутой пропозиции.

У текста с комментариями много независимых авторов. То, что одному из них виделось известным или очевидным, может не показаться таким другим. Например, комментатор может сегментировать текст и ранжировать его составляющие не так, как это предполагал автор статьи или другие комментаторы. Поэтому возражения комментатора могут быть направлены как на рему, так и на тему высказывания; а менее явно выраженные мысли могут выйти из тени и попасть в фокус внимания. В результате при разметке текста невозможно заранее предугадать, какая из частей текста вызовет возражение или поддержку комментатора. В качестве примера можно привести указанный выше комментарий (12), в котором цитируется словосочетание «эффект кофеина». Такая ситуация обуславливает сложность сегментации текста. Возникает вопрос: рассматривать ли подобные случаи как причину для более дробной сегментации? В данной работе было решено этого не делать, в частности, фрагмент (12) был связан отношением «От знающего» со всей фразой (11). Основанием для этого является то, что вне контекста неясно, на какой аспект смысла словосочетания «эффект кофеина» направлен интерес комментатора.

Задача упрощается, если часть реплики, привлекающая внимание комментатора, выделена индикаторами, например, скобками как в следующем примере. В этом случае заключенный в скобки текст самодостаточен и образует отдельный текстовый фрагмент (рис. 5):

(92) *Тожже на Mattle недавно перешел. Не идеальна (мне натирает уши ляпка) но затемняет действительно прекрасно.*

(93) *> натирает уши ляпка*

Да, точно! В конце концов наловчился удерживать ляпку на затылке.

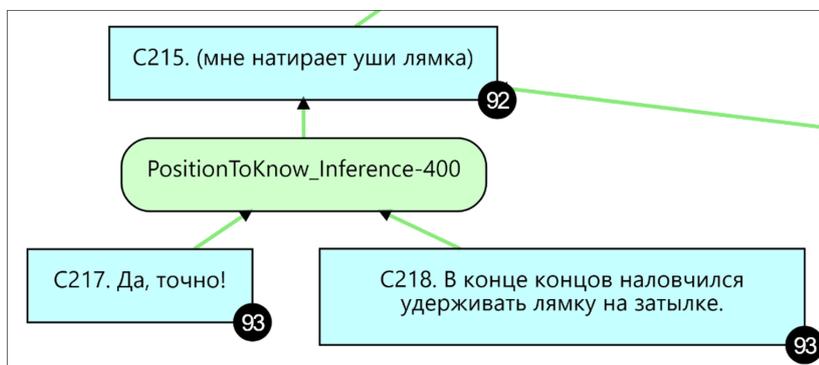


Рис. 5. Фрагмент графа для статьи «Самый научный гайд по сну»
Fig. 5. The fragment from the graph for the article “The most scientific sleep guide”

5. Для комментариев характерно использование специфических индикаторов. Так, завершение предложения может обозначаться разными средствами выражения эмоции: двоеточие и закрывающая скобка :) или двоеточие, дефис и закрывающая скобка :-), а также несколько закрывающих скобок))))) или несколько восклицательных знаков !!!!. В качестве признака начала предложения используется аббревиатура от латинского *post scriptum* (ps или PS).

Для интернет-коммуникации свойственны нарушения правил орфографии и пунктуации, что надо учитывать при автоматической обработке текста. В комментариях самые частые нарушения – написание первого слова предложения с маленькой буквы как в примере (40).

Результат разметки текста с комментариями может быть использован для выявления тех частей статьи, которые попали в фокус внимания комментаторов. Анализ количества стрелок, ведущих от комментариев к определенной вершине основного графа, построенного по тексту статьи, показывает, сколько комментаторов сфокусировались на мысли, соответствующей данной вершине. В данном случае такими оказались следующие утверждения автора статьи:

Если ваши соседи не понимают того, как важен для вас сон, купите беруши.

Установите себе будильник и обозначьте время засыпания и подъема. Взрослым нужно спать по 7–8 часов, подросткам [[около 9]].

Вот самые эффективные способы это сделать.

В первых двух случаях комментаторы делятся своим личным опытом, в третьем случае комментаторы опровергают, подтверждают или дополняют предлагаемые автором способы наладить сон, связь осуществляется через схемы «Опровержение гипотезы», «От знающего», «От части к целому».

К числу стандартных проблем разметки относится вопрос о выборе типа схемы для конкретных текстовых фрагментов. Для более четкого понимания проблемы был проведен эксперимент.

В опросе участвовали бакалавры и магистранты направления <название направления, название университета> (всего 21 человек, из них 13 студентов 4-го курса бакалавриата и 8 магистрантов). Всем участникам предварительно, в рамках лекции по лингвистике текста, включенной в соответствующие учебные программы, были даны общие сведения об аргументативном анализе текста. Однако опыта проведения аргументативной разметки студенты не имели и руководствовались только пояснениями в тексте задания, здравым смыслом, собственным опытом изучения лингвистики и логики. Можно сказать, что это были типичные представители начинающих аннотаторов, которые достаточно подготовлены (имеют базовое образование по лингвистике и математике, включая математическую логику), но не имеют опыта аргументативного аннотирования.

Каждому участнику было предложено пять пар, в каждой паре надо было выбрать то отношение, которое видится наиболее подходящим для характеристики связи между заданными фрагментами текста. В комментарии при необходимости давались краткие сведения о содержании текста, касающемся этих фрагментов. Текст задания выглядел следующим образом.

Ниже приведены пары связанных суждений. Нужно определить, каким отношением связаны суждения в каждой паре. Предлагается в каждом случае выбрать одно из двух указанных отношений.

Обозначения: А – ситуация, описанная первым суждением, В – ситуация, описанная вторым суждением.

EffectToCause или BestExplanation?

Какое отношение: «От следствия – к причине» или «от факта – к наилучшему его объяснению»?

Выберите для интерпретации выражения А→В один из двух вариантов:

- а) стрелка означает, что А есть следствие В (т. е. имеет место отношение EffectToCause)
б) стрелка означает, что А есть наилучшее объяснение для В

А [Там Па Линг играет ключевую роль в истории современной миграции людей через Азию]

В [Наши человеческие предки пробирались через материковую часть Юго-Восточной Азии по пути в Австралию около 86 000 лет назад]

Комментарий: в тексте описаны раскопки в пещере Там Па Линг, где найдены окаменелости возраста около 86 000 лет.

Результаты выбора схем показаны в табл. 2. Первая пара схем далее анализируется более подробно, поэтому для нее приведен фрагмент контекста, жирным шрифтом выделены части, обозначенные через А и В. Для пары «От корреляции к причине» (CorrelationToCause) или «От знака» были рассмотрены случаи реализации этих схем в двух разных текстах: первый – по археологии, второй – по экономике.

Таблица 2

Результаты выбора схем

Table 2

Scheme selection results

Варианты выбора типа отношения	Отношение	Количество ответов
1. «От следствия к причине» или «От лучшего объяснения»?	«От следствия к причине»	12
	«От лучшего объяснения»	9
2. «От знака» или «От факта к гипотезе»?	«От знака»	14
	«От факта к гипотезе»	7
3. «От причины к следствию» или «От позитивных следствий»?	«От причины к следствию»	6
	«От позитивных следствий»	15
4. «От корреляции к причине» или «От знака»? (текст по археологии)	«От корреляции к причине»	6
	«От знака»	15
5. «От корреляции к причине» или «От знака»? (текст по экономике)	«От корреляции к причине»	5
	«От знака»	16

В недрах лаосской пещеры ученые обнаружили самые ранние известные свидетельства того, что **наши человеческие предки пробирались через материковую часть Юго-Восточной Азии по пути в Австралию около 86 000 лет назад (В)**. ... Пещера находится более чем в 300 километрах от моря... **«Там Па Линг играет ключевую роль в истории современной миграции людей через Азию (А)**, но ее значение и ценность только сейчас признаются», — [отмечает] палеоантрополог Копенгагенского университета Фабрис Деметер...

Во всех случаях решения не были однозначными. Из этого следует, что данные пары отношений нуждаются в уточнении различий между смыслами двух схем. Это полностью не устранило расхождения между решениями аннотаторов, так как текстам естественного языка свойственна неоднозначность. Однако, эксплицировав смысловые различия, можно сделать выбор схем более объяснимым.

Для примера рассмотрим первую пару схем и ее использование при разметке статьи «В лавосской пещере найдена кость человека возрастом 86 тыс. лет» с сайта <https://poisknews.ru/>. По приведенному в табл. 2 фрагменту видно, что сегменты А и В расположены в тексте в обратном порядке и не находятся рядом. Первое сложное предложение начинает статью, в нем – в придаточном предложении (описывающем ситуацию В) – формулируется ее основной тезис, значимость которого подчеркивает далее палеоантрополог, говоря о ситуации А.

Слова палеоантрополога Деметера можно рассматривать с двух точек зрения: либо как речевое действие (произнесение определенного высказывания), либо как введение в рассмотрение описываемой ситуации А (референциальная трактовка).

Если мы выбираем первую трактовку, то установление причинно-следственной связи представляется оправданным: указанное речевое действие – это реакция Деметера на находку в пещере Там Па Линг.

Вторая трактовка, скорее, ведет к иному решению. Человек не способен непосредственно воспринимать причинно-следственные связи, такого рода рецепторы у него отсутствуют. Эти связи устанавливаются на основе научных и общих энциклопедических знаний, признаваемых в данное время в данном месте, или на основе анализа встречаемости событий. Единичный факт обнаружения останков в пещере, вообще говоря, не достаточен для интерпретации ее роли в миграционных процессах. Кроме того, в данном случае устанавливается связь между фактом (В) и его интерпретацией (А), а любой факт теоретически имеет более одной интерпретации, и указание только одной из них означает, что она в каком-то отношении лучше других. В статье не обсуждаются количество найденных останков и возможные версии попадания их в пещеру; мы не знаем, насколько подобные сомнения сняты специалистами и насколько надежна высказанная гипотеза. В таком информационном контексте уместнее схема «От лучшего объяснения», так как данные для выбора более сильной связи – причинно-следственной – читателю статьи не предоставлены.

Анализ данного примера позволяет сформулировать следующие рекомендации.

1. Если текстовый блок – это прямая или косвенная речь, то возможны две его трактовки: как речевого действия или как указания на определенную ситуацию (экстенционал). Выбор одной из этих трактовок необходимо сделать до выбора аргументативной схемы, в которую входит данный блок.

2. Если текстовые блоки – это описания ситуаций, не являющихся речевыми действиями, то основания для установления между этими блоками причинно-следственной связи дают следующие условия: а) в тексте есть индикатор, явное указывающий на такую связь; б) имеются научные или энциклопедические знания, теории или методы, поддерживающие установление такой связи. В противном случае уместнее другая схема.

3. При установлении связи между текстовыми блоками, один из которых описывает факт, а другой – его интерпретацию, надо учитывать, что, вообще говоря, возможны и другие интерпретации.

2. Новые схемы рассуждения

Опыт разметки представленных жанров показал, что некоторые способы рассуждения, которые оказалось невозможно интерпретировать однозначно в рамках имеющейся онтологии, потребовали введения новых схем: «От цели к причине» (CauseToEffectGoal), «Существую-

шая практика» (PracticalRealization), «К методу» (ToMethod), «Неактуальность утверждения» (IrrelevantStatement), «Ложная дилемма» (FalseDilemma). Примеры реализации этих схем взяты из размеченных habr-статей.

Регулярность использования этих способов аргументации в анализируемых текстах, на наш взгляд, связана и с тематическими, и с жанровыми особенностями текстов. Научное и наукоориентированное рассуждение требует, во-первых, точности в характеристике фактов, событий, их интерпретаций и отношений между ними и, во-вторых, соответствия апробированным способам решения задач (подходам, методам, приемам). Что касается жанровых особенностей, то наличие компонентного диалога (в научных и habr-статьях с комментариями) и перлокутивных коммуникативных задач (в рецензиях) предполагает критику и обсуждение отдельных утверждений и мнений с точки зрения их актуальности и соответствия целям и сложившимся практикам, а в комментариях к habr-статьям – еще и личному опыту комментаторов.

2.1. Поддерживающие схемы (схемы, направленные на доказательство тезиса)

2.1.1. От цели к причине

Схема является разновидностью аргумента «К практическому выводу» и содержит причинную посылку схемы «От причины к следствию».

Этот аргумент используют при обосновании необходимости действия *A* причинно-следственными отношениями между действием *A* и желаемым положением дел *B*, см., например, рис. 6.



Рис. 6. Пример реализации аргументативной схемы «От цели к причине»
Fig. 6. Representation of the argumentation scheme CauseToEffectGoal

2.1.2. Существующая практика

Данная схема является разновидностью схемы «К практическому выводу», имевшейся в онтологии изначально. Отличие новой схемы состоит в том, что ее вывод носит не практический («Следует совершить действие *B*»), а теоретический характер. Описание и пример реализации схемы представлены на рис. 7.

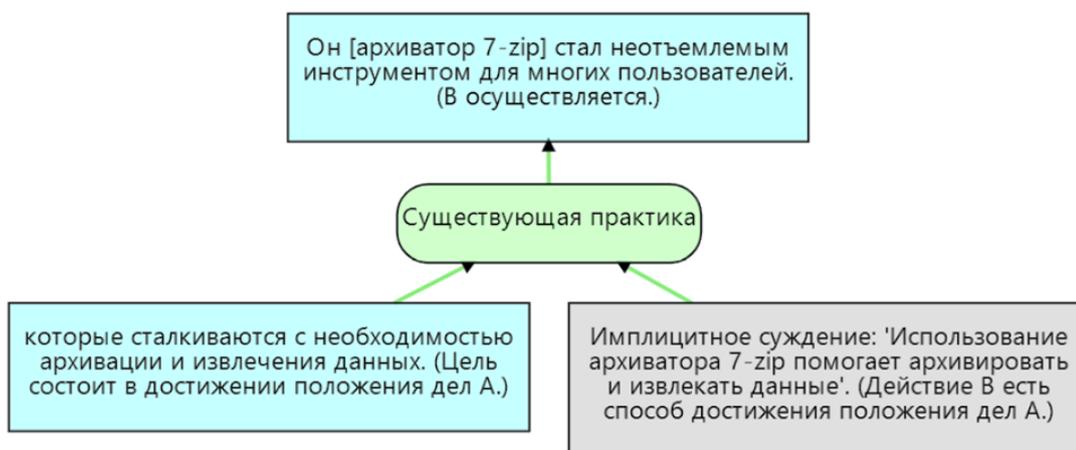


Рис. 7. Пример реализации аргументативной схемы «Существующая практика»⁴
 Fig. 7. Representation of the argumentation scheme PracticalRealization

2.1.3. К методу

Этот тип аргумента является усложнением схемы «К практическому выводу». Она дополняется информацией о методе, который предлагается использовать в данной ситуации на основании его конвенциональности в той или иной предметной области. Описание и пример реализации данного типа аргумента представлены на рис. 8.



Рис. 8. Пример реализации схемы «К методу»
 Fig. 8. Representation of the argumentation scheme ToMethod

2.2. Конфликтные (атакующие) схемы (схемы, направленные на опровержение тезиса)

2.2.1. Неактуальность утверждения

В анализируемых статьях точка зрения нередко вводится на фоне других точек зрения или в сопоставлении с ними, например: [Точка зрения₁] *Еще совсем недавно, многие экономи-*

⁴На рис. 7 и 8 в прямоугольниках серого цвета для наглядности представления структуры аргумента помещены имплицитные утверждения. Однако в разметке этот функционал не использовался, и имплицитные послышки и выводы не реконструировались.

сты предрекали Китаю и китайской экономике большое будущее. <...> [Точка зрения₂] но сейчас, похоже, что эра этого «чуда» подходит к концу. Точка зрения₂, подкрепленная ссылкой на экспертов (многие экономисты), считалась справедливой некоторое время назад, но в настоящее время она, по утверждению автора, не является актуальной. В этом фрагменте точки зрения противопоставлены по отношению к актуальности на время написания текста.

Аналогичные случаи встречаются при обзоре обновлений в рассматриваемой области, при рассмотрении истории вопроса, при описании динамики личных воззрений, при противопоставлении более обоснованной точки зрения менее обоснованной.

Для аннотирования таких рассуждений в онтологию была введена конфликтная схема, в которой аргумент, представляющий неактуальную (устаревшая, неточная, не учитывающая всех факторов и пр.) точку зрения эксперта или носителя «всеобщего» мнения, атакуется более надежным утверждением. Таким образом, схема «Ненадежность утверждения» имеет следующие роли:

- атакуемые элементы: аргумент «От эксперта», аргумент «От всеобщего мнения» (PopularOpinion);
- исключение: *A* – неактуальное утверждение (под *A* понимается утверждение, которое в атакуемой схеме считается истинным).

2.2.2. Ложная дилемма

Анализ полилога в комментариях к аналитическим статьям позволил выявить еще одну схему атаки – «Ложная дилемма», которая может быть применима к текстам этого и других жанров со сложными аргументативными структурами, в том числе диалогических или полемических жанров. Она состоит в демонстрации того (или указании на то), что некоторое противопоставление, считаемое дилеммой, таковой не является. Пример реализации схемы представлен на рис. 9.

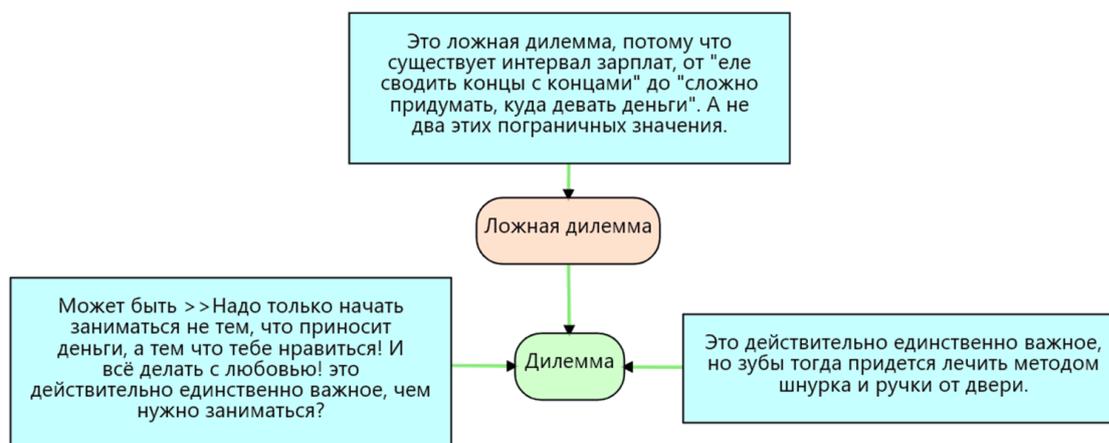


Рис. 9. Пример реализации схемы «Ложная дилемма»
Fig. 9. Representation of the argumentation scheme FalseDilemma

2.3. Перспективы пополнения онтологии

Обнаружились и другие способы рассуждения, описание которых отсутствует в онтологии, но которые еще находятся в процессе осмысления и формирования в качестве отдельных схем аргументации.

2.3.1. Аргументация от качества

Анализ различных тем, представленных в аннотированных статьях, содержит качественные характеристики тех или иных идей, выступающие в роли аргумента в пользу или в опровержение того или иного тезиса. Большая часть таких характеристик интерпретируется с помощью схем аргументов «От лучшего объяснения», «От позитивных последствий» и «От негативных последствий». Однако ряд случаев не укладываются в эти три типа аргументов. Так, убедительность тезиса о необходимости применения той или иной разработки и т. п. может доказываться с помощью плана дальнейшей работы над ней, например: *Так или иначе, мы планируем дальше расширять когнитивные возможности нашей модели со следующими релизами* – или с помощью ее сопоставления с аналогами, например: *На этом корпусе обучалось большинство наших моделей, и он довольно чист, но в современных реалиях уже не кажется большим. Для сравнения, размер хорошего английского корпуса [[The Pile]] составляет более 800 Гб.* В настоящий момент еще неясно, предпочтительно ли вводить в онтологию одну общую качественную схему аргумента или отдельные, более конкретные – для каждого типа случаев.

2.3.2. Условно-следственные отношения

При использовании схемы «От причины к следствию» обнаружились рассуждения, в которых от собственно причинно-следственных отношений (когда факта, события или действия *A* достаточно для появления или осуществления *B*) следует отделить условно-следственные (без достаточности условия). Например, в рассказе об ипотечном кризисе 2008 года его всемирный характер объясняется двумя факторами: [Фактор₁] *Пузырь лопнул – заемщики стали массово отказываться от ипотеки, а кредитные рейтинги лучших секьюритизированных ипотечных бумаг быстро стали мусорными.* [Фактор₂] *В американские ипотечные бумаги инвестировало огромное количество фондов со всего мира.* [Следствие] *Поэтому крах рынка кредитования недвижимости в США ударил по многим странам.* Не представляется возможным объединить описание этих факторов в одно утверждение, поскольку первое из них поддерживается другими аргументами. Разрешение ситуаций такого типа возможно двумя способами: введением схемы «От условия к следствию» или добавлением функционала совместной поддержки тезиса (возможности указания на то, что группа аргументов только в своей совокупности обеспечивает поддержку вывода). Второй способ позволит объединять не только условно-следственные, но и другие отношения, однако не даст возможности указать на недостаточность (частичный характер) одного аргумента при отсутствии других.

Заключение

В статье описаны некоторые результаты, основанные на опыте аннотирования аргументации в интернет-текстах научного функционального стиля, выявлены сложности анализа определенных моделей рассуждения, характерных для жанра научно-популярной статьи с комментариями.

В связи с этим затронуты базовые вопросы жанрового анализа интернет-текстов и, более конкретно, жанровой классификации текстов на основе целевой составляющей коммуникативной ситуации. Это позволило выделить подгруппы научных интернет-текстов (информационные, деловые и диалогические) и описать некоторые их жанровые особенности. Несколько более подробно описан компонентный диалог, характерный для научных статей с рецензиями, отнесенных к жанровой группе деловых текстов. Жанровое исследование аргументации в интернет-текстах имеет множество перспектив: интересно не только развить намеченную в статье многоаспектную классификацию, но и посмотреть, какие модели рассуждения преимущественно применяются в текстах полученных жанровых классов. Для этого необходимо

провести статистический анализ использования схем рассуждения по корпусу в целом и по отдельным менее изученным жанровым классам.

На основе анализа моделей рассуждения и способов их языкового воплощения был выявлен ряд особенностей научно-популярных статей с комментариями. Так, аргументативная функция часто усиливается путем добавления эмоционально окрашенных или оценочных составляющих, этот способ придания аргументу дополнительного веса применяется к разным типам аргументов. Особенностью языковой формы реализации различных типов аргументативной связи является использование вопросов. Разметка текста усложняется динамичностью его сегментации: автор статьи и комментатор могут выделять в качестве аргументативно значимых разные фрагменты текста. Дополнительные сложности вносит также использование дискурсивных анафор, в которых антецедент анафоры нужно искать не по его языковой реализации (как в случае цитирования), а по выраженной в нем мысли. Для проверки различимости некоторых аргументативных отношений был проведен пробный эксперимент, подтвердивший сложность их различения. Повышение объяснимости различий между такими отношениями намечено как одна из перспектив развития рекомендаций по аннотированию текстов.

Отмечены перспективы пополнения онтологии: в процессе анализа и аннотирования текстов обнаруживаются способы рассуждения, которые не укладываются в текущую онтологию и требуют оформления в виде отдельных схем аргументации. Эти аргументы выражают причинно-целевые, условно-следственные и качественные отношения, отношения цели и средства, идеи актуальности/неактуальности и контрастности/контрадикторности утверждений. Полагаем, что эти схемы специфичны для наукоориентированных и других текстов со сложным содержанием, в которых важна точная характеристика отношений между фактами, событиями, явлениями и мнениями, а также для текстов с диалогом в комментариях и текстов, предполагающих перлокутивный эффект уточнения и корректировки контрагентом его текста.

Список литературы

- Бахтин М. М.** Проблема речевых жанров // Собр. соч. в 7 т. М.: Русские словари, 1996. Т. 5. С. 159–207.
- Винокур Т. Г.** Закономерности стилистического использования языковых единиц. М.: Наука, 1980. 237 с.
- Гаузенбласс К.** О характеристике и классификации речевых произведений // Новое в зарубежной лингвистике. Вып. VIII: Лингвистика текста. М.: Прогресс, 1978. С. 57–77.
- Еемерен Ф. Х. ван.** Современное состояние теории аргументации // Важнейшие концепции теории аргументации. СПб.: Филологический факультет СПбГУ, 2006. С. 14–33.
- Еемерен Ф. Х. ван, Гроотендорст Р.** Систематическая теория аргументации: прагма-диалектический подход. М., 2021. 264 с.
- Интернет-коммуникация как новая речевая формация: коллективная монография / Науч. ред. Т.Н. Колокольцева, О. В. Лутовинова. М.: Флинта, 2018. 328 с.
- Кибрик, А. А.** Модус, жанр и другие параметры классификации дискурсов // Вопросы языкознания. 2009. № 2. С. 3–21.
- Кононенко И. С., Сидорова Е. А., Веремьянина А. О.** Подход к извлечению информации о событиях в энергетике (на материале новостных сообщений информагентств) // Тр. XX Байкальской Всерос. конф. «Информационные и математические технологии в науке и управлении». Ч. 3. Иркутск: ИСЭМ СО РАН, 2016. № 3. С. 126–136.
- Розина И. Н.** Технологии исследования и продвижения компьютерно-опосредованной коммуникации // Educational Technology & Society. 2007. Т. 10, № 2. С. 230–245.

- Саломатина Н. В., Гусев В. Д.** Автоматизация формирования индикаторных словарей и возможности их использования // Тр. Междунар. конф. Диалог-2006 «Компьютерная лингвистика и интеллектуальные технологии». М.: Наука, 2006. С. 121–125.
- Сидорова Е. А., Ахмадеева И. Р., Загорулько 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. DOI: 10.18287/2223-9537-2020-10-4-489-502
- Щипицина Л. Ю.** Классификация жанров компьютерно-опосредованной коммуникации по их функции // Изв. Рос. гос. пед. ун-та им. А. И. Герцена. СПб., 2009. № 114. С. 171–178.
- Kononenko I. S.** Pragmatic Aspects of Internet Communication: Towards Websites Genre Models // Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue” (2014). Issue 13 (20). Vol. 1. Moscow: RGGU, 2014. P. 251–260.
- Walton D., Reed C., Macagno F.** Argumentation schemes. Fundamentals of critical argumentation. New York: Cambridge Univ. Press, 2008. 443 p.

References

- Bakhtin M. M.** The Problem of Speech Genres. In: *Bakhtin, M. M. Collected Works: in 7 vols.* Moscow, Russian Dictionaries, 1996, vol. 5, pp. 159–207. (in Russ.)
- Eemeren F. N., Grootendorst P.** Systematic theory of argumentation: pragma-dialectical approach. Moscow, 2021, 264 p. (in Russ.)
- Eemeren F. N.** The State of the Art in Argumentation Theory. In: *Crucial concepts in argumentation theory.* St. Petersburg, Faculty of Philology, SPbSU publ., 2006, pp. 14–33. (in Russ.)
- Hausenblaz K.** On the characterisation and classification of speech works. In: *The new in foreign linguistics. Issue VIII: Linguistics of the text.* Moscow, Progress publ., 1978, pp. 57–77. (in Russ.)
- Internet-communication as a new speech formation: collective monograph / ed. by T. N. Kolokoltseva, O. V. Lutovinova. Moscow, FLINTA publ., 2018. 328 c. (in Russ.)
- Kibrik A. A.** Modus, genre and other parameters of discourse classification. *Voprosy Jazykoznanija*, 2009, no. 2, pp. 3–21. (in Russ.)
- Kononenko I. S.** Pragmatic Aspects of Internet Communication: Towards Websites Genre Models. Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue”, 2016, iss. 13 (20), vol. 1, Moscow, RGGU, 2014, pp. 251–260.
- Kononenko I. S., Sidorova E. A., Veremianina A. O.** Approach to Extracting Information on Energetics from News Agency Reports. Information and mathematical technologies in science and management. Proceedings of the XX Baikal All-Russian Conference. Irkutsk, 2016, no. 3, pp. 126–136. (in Russ.)
- Rozina I. N.** Technologies of Research and Advancement of Computer-Mediated Communication. *Educational Technology & Society*, 2007, vol. 10, iss. 2, pp. 230–245. (in Russ.)
- Salomatina N. V., Gusev V. D.** Automation of CUE Dictionaries Formation and Their Applications. In: Computational Linguistics and Intellectual Technologies, Moscow, Nauka publ., 2006, pp. 121–125. (in Russ.)
- Shchipitsina L. Yu.** Functional Classification of Computer-Mediated Genres. *Izvestia: Herzen University Journal of Humanities & Sciences*, 2009, no. 114, pp. 171–178. (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 of designing*, 2020, vol. 10, no. 4, pp. 489–502. (in Russ.) DOI: 10.18287/2223-9537-2020-10-4-489-502.
- Vinokur T. G.** Regularities of stylistic use of linguistic units. Moscow, Nauka publ., 1980, 237 p. (in Russ.)
- Walton D., Reed C., Macagno F.** Argumentation schemes. Fundamentals of critical argumentation. New York, Cambridge Univ. Press, 2008, 443 p.

Информация об авторах

Тимофеева Мария Кирилловна, доктор филологических наук, доцент, ведущий научный сотрудник Института систем информатики им. А. П. Ершова СО РАН

Ильина Дарья Владимировна, кандидат филологических наук, программист 2-й категории Института систем информатики им. А. П. Ершова СО РАН

Кононенко Ирина Семёновна, научный сотрудник Института систем информатики им. А. П. Ершова СО РАН

Information about the Authors

Mariya K. Timofeeva, Doctor of Sciences (Philology), Associate Professor, leading researcher of the A. P. Ershov Institute of Informatics Systems SB RAS.

Daria V. Iina, Candidate of Sciences (Philology), 2nd category programmer of the A. P. Ershov Institute of Informatics Systems SB RAS.

Irina S. Kononenko, Researcher of the A. P. Ershov Institute of Informatics Systems SB RAS.

*Статья поступила в редакцию 23.11.2023;
одобрена после рецензирования 01.12.2023; принята к публикации 17.01.2024*

*The article was submitted 23.11.2023;
approved after reviewing 01.12.2023; accepted for publication 17.01.2024*