

Методология многоаспектной разметки аргументации в научных статьях

Содержание

Введение	2
1. Формальная модель для представления аргументационных структур	2
2. Методика аннотирования текста.....	2
2.1. Сбор и предобработка текстов коллекции	3
2.2. Предварительный анализ текста	3
2.2.1. Правила определения главного тезиса	4
2.2.2. Выявление аргументативных высказываний	5
2.2.3. Правила определения границ компонентов аргумента	6
2.2.4. Разграничение структурных единиц текста.....	6
2.3. Глубинный анализ текста.....	7
2.3.1. Построение подграфов аргументации отдельных структурных единиц	7
2.3.1.1. Актуальные схемы аргументации	10
2.3.1.2. Уточнение типов связи	18
2.3.2. Объединение подграфов в связный граф.....	21
2.3.3. Проверка разметки, нумерация утверждений и схем аргументации	21
3. Выбор и подсчет коэффициентов согласия	21
4. Инструкция пользователя	22
Список литературы.....	23

Введение

Термин «методология» в данной работе используется как синоним к термину «методика», т.е. рассматривается в «узком» смысле как совокупность методов, приемов, опробованных и изученных для выполнения работы по разметке аргументации в научных статьях. Методика предназначена для людей мало знакомых с тематикой, но профессионально заинтересованных в изучении особенностей создания корпусов с разметкой аргументации для использования в машинном обучении (МО).

Под аргументом понимается совокупность связанных утверждений, одно из которых (заключение) доказывается объединением остальных (посылок). Утверждение представляет собой сформулированное на естественном языке высказывание, включенное в структуру аргументации. Многоаспектность разметки заключается в моделировании аргументационных структур на трёх уровнях:

- 1) выделения утверждений, образующих аргументы;
- 2) построения связей между утверждениями внутри пересекающихся аргументов (что подразумевает определение ролей утверждений как посылок / заключений);
- 3) детализации типов связи (моделей рассуждения, или аргументационных схем) между утверждениями в аргументе (посылками и заключением), что является важной отличительной особенностью создаваемого на основе методики корпуса.

Целью разработанной методики являлось создание такого алгоритма разметки текстов, который бы обеспечивал согласованную разметку, производимую разными аннотаторами. Формальной оценкой согласованности служат различные коэффициенты согласия, основанные на количественной и качественной оценке расхождений между аннотаторами на каждом уровне моделирования аргументационных структур.

1. Формальная модель для представления аргументационных структур

Используемый стандарт для формального представления аргументации – Argument Interchange Format (AIF) [Rahwan I., Reed C. 2009] – позволяет анализировать лингвистическое оформление утверждений языковыми средствами и их формальную структурную организацию на уровне абстрактных (логических) схем аргументации. AIF обеспечивает возможность поэтапной разметки аргументации согласно условиям многоаспектности 1) – 3) (см. Введение). В результате моделирования текст преобразуется в ориентированный граф аргументации с двумя типами узлов: информационными узлами, содержащими утверждения, и узлами-схемами, определяющими точную модель рассуждения для каждой связи между утверждениями. Связи соответствуют ребрам в графе, причем 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 данном случае из компендиума Уолтона [Walton et al, 2008], применимого к текстам различных жанров и рекомендованного разработчиками AIF. Важно отметить, что реализация какой-либо схемы из классификации Уолтона не требует эксплицитного выражения в тексте всех поддерживаемых ею посылок, поскольку некоторые из них могут только подразумеваться. Подробное описание Уолтоном принципов выявления аргументационных схем способствует единообразию в моделировании аргументационных структур разными аннотаторами.

2. Методика аннотирования текста

Данными для аргументационной разметки выступают неструктурированные тексты на естественном языке. Итогом аннотирования являются ориентированные корневые

ациклические графы аргументации. Инструкция пользователя, содержащая технические аспекты разметки, размещена в разделе 4.

2.1. Сбор и предобработка текстов коллекции

Общие требования к текстовой коллекции: однородность, репрезентативность.

Однородность коллекции обеспечивается введением следующих ограничений:

- 1) включением в коллекцию текстов одного жанра;
- 2) ограничением числа предметных областей, к которым относятся тексты;
- 3) соответствием объема каждого текста характерному для жанра.

Репрезентативность коллекции достигается за счет:

- 1) политематичности коллекции,
- 2) разнообразия авторов текстов, изданий, в которых они публикуются.

Работоспособность созданных в данном исследовании методов продемонстрирована на текстах жанра короткой научной статьи (500–2000 слов) двух тематик (лингвистика, ИТ). Источник публикаций – on-line библиотеки. Методика может быть применена к научным текстам большего объема.

Специфические требования учитывают особенности жанра. Необходимо, чтобы короткие научные статьи обладали следующими свойствами:

- 1) явным выражением главного тезиса (маркирование либо чёткое формулирование);
- 2) явной структуризацией разделов автором (необязательно, полезно для выбора одного текста из нескольких, удовлетворяющих всем иным критериям);
- 3) минимумом таблиц и рисунков в тексте (их содержание часто замещает словесную формулировку доводов).
- 4) ограниченным числом символов некириллического алфавита, отсутствие идеографических знаков (китайских, японских и пр.).

Рекомендации.

1. Использовать on-line ресурсы со стабильным доступом к текстам.
2. Подбирать статьи с применением поисковых запросов по ключевым словам, с фильтрацией по предметной области, по дате публикации и пр.
3. Перед разметкой проводить предобработку текстов: удалять фрагменты, не содержащие аргументацию (ключевые слова, список литературы, благодарности, сноски, сведения об авторах), аннотацию (как сжатый вариант текста), а также переносы в слова; восстанавливать авторское абзацное членение; унифицировать индикаторы списков при использовании нетипичных графических символов; удалять цифровые ссылки на сноски.
4. Включать в разметку метаинформацию: название статьи, ФИО авторов, предметную область, тематический раздел, ссылку на источник.
5. Предпочтительно привлекать к разметке лингвистов, знакомых с тематикой текстов.
6. Решение о включении текста в коллекцию принимать по результатам «ручного» анализа текстов аннотаторами.

2.2. Предварительный анализ текста

Цель предварительного анализа – знакомство с содержанием текста и выбор главного доказываемого тезиса – утверждения, выражающего основную идею. Кроме того, в ходе ознакомительного прочтения текста аннотатору следует обращать внимание на полезные для принятия решений во время разметки особенности:

- а) явленную структуру, отражающую логику изложения (введение, обзор литературы, применяемые методы, эксперимент: данные и результаты, выводы, заключение, абзацы);
- б) характерные аспекты содержания текста (для научных статей: актуальность, цель работы, новизну, методы, ..., итог);

в) маркеры разных типов, сигнализирующие об аспектах содержания, типе/схеме рассуждения и пр.

г) отдельные легко узнаваемые (как правило, частотные либо маркированные) схемы рассуждения (известные из компендиума Уолтона);

2.2.1. Правила определения главного тезиса

Главный тезис (Main Thesis, МТ, корневая вершина графа) определяет фокус внимания и дальнейшие действия разметчика по установлению и связыванию аргументативных высказываний. Поэтому при множественной разметке текста МТ должен быть согласован аннотаторами перед построением графов аргументации. Пример расхождений в разметке вследствие разногласия в выборе главного тезиса доступен по ссылке: https://uniserv.iis.nsk.su/arg/corpora#text_6171228c64d56970646766be.

Ключевые признаки главного тезиса:

- 1) *информативность* (семантическая точность и полнота): МТ максимально точно и полно (по сравнению с другими утверждениями статьи) отражает результирующий вывод по излагаемому материалу и ключевую доказываемую идею;
- 2) *прагматическая замкнутость*: главный тезис функционирует как законченное самостоятельное утверждение, которое представляется в тексте для выражения *именно своего содержания* (а не направлено на раскрытие иного вывода, по отношению к которому является вспомогательным).

Факультативные признаки (возможный ориентир в определении главного тезиса):

- 3) *позиционный*: главный тезис позиционно тяготеет к заключительной части статьи (Выводы, Заключение, последний \ предпоследний абзац);
- 4) *тематический*: близок по смыслу к заголовку статьи, может рассматриваться как его дополнение или расширенная формулировка. Исключение – заголовки, формулировка которых является сужением или расширением рассматриваемой в статье проблематики, либо по каким-то причинам перемещает акценты относительно материала исследования. Пример такого текста доступен по ссылке: https://uniserv.iis.nsk.su/arg/corpora#text_64297_ac1d17da2686dd44caa. Здесь главный тезис, представляющий основной вывод, соотносится с заголовком не так явно, как вступительные тезисы о влиянии интернета на процессы заимствования;
- 5) *лексический*: снабжен характерным маркером. Исключение: маркерное выражение является стилистическим приемом автора как, например, в статье «Семантика-1», маркер *Таким образом*,: «*Таким образом*, пропозиция спортивного события...», «*Таким образом*, стандартная модель...». В данном случае как маркер оно не может быть использовано.

При выборе МТ сначала анализируется финальная часть текста – Выводы, Заключение и/или последние абзацы, из нее выбирается высказывание, удовлетворяющее как минимум ключевым критериям. Если такого не обнаружено, анализируется остальная часть текста. Пример такого текста: Корпусная-2 (https://uniserv.iis.nsk.su/arg/corpora#text_63f0a776d29f_90dd6eb68bc1), где МТ представлен в начале статьи, не дублируется в её конце (где от частных примеров авторы переходят к применению описываемых корпусов и дальнейшим перспективам исследований). Если претендентов на МТ обнаружилось больше одного, то на роль МТ выбирается максимально точная и полная формулировка вывода с проявлением наибольшего числа признаков. Исключение: в отдельных случаях, несмотря на наличие позиционного признака и маркера, главным тезисом может быть выбрано другое утверждение как наиболее соответствующее выводу, который можно сделать из содержания текста. Примером такого текста является «Заимствования-4», доступным по ссылке:

https://uniserv.iis.nsk.su/arg/corpora#text_64ba6af121cfd05dcde7a723.

Если не нашлось ни одного высказывания, удовлетворяющего всем критериям, то главным тезисом считается заголовок статьи (см. пример текста: https://uniserv.iis.nsk.su/arg/corpora#text_6171227835f876ece0f0bb3e).

Согласно 2) информационный узел, соответствующий МТ, не имеет исходящих путей (он не является посылкой ни для одного другого утверждения). Потенциальные расширения МТ (дополнительные комментарии к основному заявленному выводу) рассматриваются как поддерживающие его посылки. Например, указание возможных направлений дальнейшей работы, которые интерпретируются как практическая польза от достижения результата исследования, в научных статьях вводятся в аргументационную структуру как посылки к МТ по схеме «от практической цели» (*Practical Reasoning*).

2.2.2. Выявление аргументативных высказываний

В коротких научных статьях аргументативное содержание преобладает над неаргументативным. Поэтому в граф аргументации включаются все рассуждения текста, непосредственно или опосредованно связанные с доказательством главного тезиса.

Критерии для определения *аргументативных* высказываний (утверждений):

а) *аргументативной связности* – в полной структуре текста утверждения реализуются в составе аргументов (или конфликтов), при пропуске такого утверждения будет разорвана связь между смежными с ним (его посылками или заключениями для аргументов, атакуемыми или атакующими утверждениями в случае конфликтов).

б) *содержательной полноты* – в утверждениях, как правило, выражаются целостные суждения (каждое утверждение *сообщает законченную мысль*, не теряет осмысленности при его изолированном прочтении). Исключения: в составе аргументов с отдельными схемами рассуждения (*Expert Opinion, Verbal Classification,...*) могут выделяться семантически неполные посылки (например, названия перечисляемых классов).

в) *явности выражения* – каждая информационная вершина в графе аргументации должна соответствовать конкретному фрагменту аннотируемого текста. Неявные утверждения не отражаются в структуре аргументации.

К неаргументативному содержанию относятся высказывания следующего типа:

- 1) высказывания, не вводящие информативного содержания, а лишь отсылающие к другим частям текста и выступающие в роли риторических связок (вида «а теперь рассмотрим X», где X подробно описывается в следующих предложениях);
- 2) информационные, для которых не находится способа их связывания с утверждениями;
- 3) названия (и содержание) таблиц, рисунков, определения обозначений к ним;
- 4) заголовки списков, дублируемые содержанием других утверждений (см. Корпусная-3, https://uniserv.iis.nsk.su/arg/corpora#text_64ba6b2d23fce50c6ee0a330: заголовок списка «...следует отметить, по каким критериям мы вправе производить сопоставление между подкорпусами сопоставимого корпуса – переводным подкорпусом и оригинальным подкорпусом:» повторяется расширенным тезисом «При учете данных критериев при построении электронных корпусов становится возможным осуществление сопоставительного анализа между подкорпусами сопоставимого корпуса», который дополняет указание на классификацию практическим основанием её применения);
- 5) содержание списков большого объема, которое рациональнее следовало бы представить в виде отдельной таблицы (см. «Классификация-4», где автор перечислил все 22 признака для векторного представления объектов в виде построчного списка https://uniserv.iis.nsk.su/arg/corpora#text_64f9962ce2c61efe4a570545).

Критерии, позволяющие выявить *неаргументативные* высказывания в тексте:

а) *риторической добавочности* (применяется для риторического упорядочения материала) – высказывание является неаргументативным, если оно не содержит уникальных сведений (не сообщаемых нигде далее по тексту), а только указывает на их сообщение в другой части текста или применяется строго для риторической организации

других тезисов в линейной структуре текста как последовательности предложений (например, «результаты анализа приведены в таблице» «обратимся к проблеме», «рассмотрим случай»...).

б) *дублирование отдельных тезисов* – содержание одного высказывания, претендующего на аргументативность, полностью или частично повторяет содержание другого, тогда в граф аргументации включается *только одна вершина* с наиболее содержательной формулировкой в высказывании.

На используемой для разметки платформе имеется словарь индикаторов аргументации, которые автоматически выделяются в загруженном в корпус тексте полужирным шрифтом и позволяют аннотатору обратить внимание на фрагмент, содержащий такой индикатор как на потенциально аргументативный. Есть возможность создать «авторский» индикатор и внести его в список потенциальных индикаторов (см. также (МетодологияРазметки_v1(2024)mod, Приложение, таблица 1, [https://uniserv.iis.nsk.su/arg_files/МетодологияРазметки_v1\(2024\)mod.pdf](https://uniserv.iis.nsk.su/arg_files/МетодологияРазметки_v1(2024)mod.pdf)). Маркеры включаются в состав утверждений в целях упрощения компьютерной обработки.

2.2.3. Правила определения границ компонентов аргумента

Основной тип сегмента в корпусе коротких научных статей – предложение. Исключение – списки, где каждый элемент списка является объемным и представлен отдельным абзацем. Элемент списка такого сорта считается отдельным утверждением. Менее объемные списки-перечисления разбиваются на отдельные сегменты, если элемент списка (отдельное утверждение) является выводом аргумента или есть необходимость показать множественный характер аргументации (число посылок больше единицы). Исключением также является редкий случай, когда объединяются два предложения, если роль посылки либо заключения эти предложения могут выполнять только совместно (указание любого одного из них при пропуске другого приведёт к семантической неполноте связи в общей структуре). Такие зависимые друг от друга по смыслу предложения могут соединяться анафорической ссылкой.

Недостатками такой сегментации являются, в частности, пропуск аргумента, содержащегося в сложном предложении, отсутствие выделения однородных членов предложения в отдельные утверждения, если они поддерживаются различными посылками. В последнем случае все посылки будут присоединяться к узлу в графе, содержащему соответствующие выводы. Однако наиболее существенные рассуждения чаще всего формулируются авторами развернуто: компоненты аргумента представлены отдельными предложениями.

2.2.4. Разграничение структурных единиц текста

Подграфы графа аргументации обычно формируются по отдельным структурным единицам текста и аспектам содержания, так называемым блокам (см. п. 2.2 а), б)). Отдельные аспекты содержания часто снабжаются характерными маркерами (МетодологияРазметки_v1(2024)mod, Приложение, таблица 2, [https://uniserv.iis.nsk.su/arg_files/МетодологияРазметки_v1\(2024\)mod.pdf](https://uniserv.iis.nsk.su/arg_files/МетодологияРазметки_v1(2024)mod.pdf)). Прослеживается связь аспектов содержания текста с применением конкретных схем. Вершины подграфов – главные тезисы блоков – часто связаны между собой или с главным тезисом текста.

Рекомендуется в ходе построения графа определять базовые (главные) утверждения каждого абзаца, которые характеризуют основные пути доказательства главного тезиса текста. Именно они часто являются утверждениями, позиционно разнесенными в тексте, но образующими аргумент. Основу хорошо формализуемых правил выявления главных утверждений абзаца составляет фильтрация неглавных утверждений абзаца.

- *Условия фильтрации неглавных утверждений:*

- 1) Если в абзаце *больше 3 утверждений* и *позиция утверждения не первая*, то утверждение не главное.
- 2) Если в абзаце *есть имя собственное* (персона), то утверждение не главное.
- 3) Если в абзаце *есть длинная цитата* (из 4+ слов), то утверждение не главное.

Выполнение любого одного условия означает, что утверждение не является главным. Условия упорядочены по убыванию качества фильтрации. Проведение компьютерного эксперимента по применению построенных правил, показало, что их усложнение позиционной части условия 1) ненамного повышает точность, но значительно понижает полноту. Условие 2): наличие имени в утверждении указывает на отсылку к другому исследователю (что редко для главных тезисов абзацев). Условие 3): цитируемые фразы редко появляются в главных тезисах.

2.3. Глубинный анализ текста

Глубинный анализ базируется на результатах первичного прочтения текста: выделенных блоках и выявленном главном тезисе. Его итогом является *связный* граф аргументации. Последовательный проход по блокам текста (от начальных к конечным или наоборот) или выборочный по отдельным блокам и/или их частям через разграничение и анализ каждой), зависит от структуры конкретной статьи и стиля аннотатора.

2.3.1. Построение подграфов аргументации отдельных структурных единиц

Анализ явно выделенного или неявного логического блока проводится с целью связывания утверждений блока в аргументы с помощью подходящих аргументативных отношений (схем из компендиума Уолтона).

При определении связей между утверждениями предлагается ориентироваться на следующие критерии:

а) *аргументационной общности* (основной критерий при выявлении связи между утверждениями) – если между двумя (или более) утверждениями можно осуществить смысловой переход согласно некоторой модели рассуждения, то соответствующая связь между ними фиксируется в графе: информационные вершины, соответствующие связанным утверждениям, соединяются путем, в составе которого имеется схема, инцидентные ей и каждому утверждению ребра. Ориентация ребер: от посылки к схеме, от схемы к утверждению-заключению (см. Инструкцию разметки).

б) *пространственной близости* – поиск для данного утверждения связанных с ним утверждений следует проводить, начиная с ближайшей окрестности в последовательности предложений блока текста (редко вплоть до границ всего текста).

в) *семантической полноты аргументов* – ряд моделей рассуждения допускает поддержку заключения несколькими дополняющими друг друга посылками с разными смысловыми ролями. Соответственно, в случае обнаружения между двумя утверждениями аргументативной связи такого типа, в схеме рассуждения которой допускается включение числа посылок больше единицы, необходимо провести поиск *всех утверждений, которые могут соответствовать этим посылкам*. Связность аргументационного графа *не предполагает* полного заполнения всех допустимых посылок: значительная часть смысловых компонентов в аргументах может оставаться пустой ввиду неявного выражения соответствующих утверждений.

При построении связей между утверждениями в одном блоке также желательно соблюдать его смысловую целостность: сначала определить главный тезис для этого блока, затем связывать утверждения из данного блока, не добавляя утверждения из других. Разметку производить от главного тезиса блока к его посылкам.

Способы связывания утверждений [Lawrence, J., Reed, C. 2019]:

- *Последовательная аргументация* (Sequential Arguments) – это “цепочка” связанных утверждений, когда заключение одного аргумента является посылкой другого.
- *Множественная конвергентная аргументация* (Convergent Arguments) – одно и то же утверждение-заключение поддерживается совокупностью утверждений-посылок.
- *Множественная дивергентная аргументация* (Divergent Arguments) – одна и та же посылка поддерживает несколько выводов.

На рис.1 представлены основные способы связывания утверждений: Convergent Arguments (S33, S35 → S32), Divergent Arguments (S24 → S34, S35), Sequential Arguments (S24 → S35 → S32). Вершины-утверждения изображены прямоугольниками, вершины-схемы – эллипсами. Утверждения и схемы обозначены символами S и A соответственно и снабжены номерами (согласно порядку утверждений в тексте). Порядок утверждений в тексте отличается от порядка в аргументационной структуре: посылки могут встречаться в тексте как до, так и после заключения, а между утверждением и доводами в его поддержку может приводиться обоснование этих доводов.

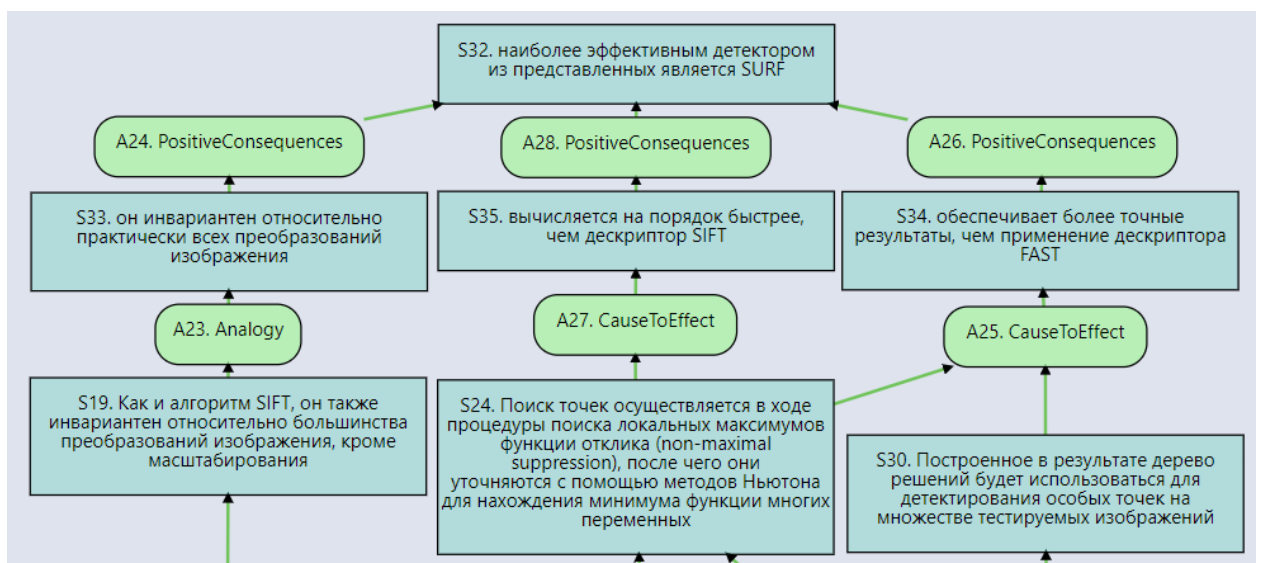


Рис.1. Аргументационная разметка фрагмента текста.

Применение разных способов связывания (Convergent Arguments и Sequential Arguments) является причиной расхождений при множественной разметке текста. Но в большей части случаев оба варианта допустимы и не противоречат логике доказательства. В примере на рисунке 2 разметчики соглашаются в выделении трех утверждений из одного абзаца (S41, S42 и S44 в варианте слева, и S24, S25, S26 в варианте справа), причем оба предполагают реализацию между ними модели «Part to Whole». Однако один аннотатор добавляет промежуточное утверждение S53 из другого абзаца (который непосредственно вводит пример перевода) и присоединяет к нему две посылки S42 и S44 параллельно, тогда как другой разметчик предполагает между ними последовательную связь (не считает их функционально равноправными, а предполагает зависимость одной посылки от другой). Как следствие, хотя разметчики воспринимают фрагмент текста достаточно близким образом (оба выделяют три тезиса, выявляют среди них общий вывод на основе двух посылок), автоматическая проверка отмечает расхождение во всех пяти связях.



Рис. 2. Пример расхождения при построении связей.

Общая рекомендация: если при построении связей внутри блока взаимные отношения утверждений, поддерживающих главный тезис блока, представляются не совсем ясными, желательно присоединить эти утверждения к главному посредством параллельных связей. Такой стандартизирующий принцип позволяет повысить единообразие разметки и избежать частых расхождений в конфигурации связей.

Типы связей:

- *поддержка утверждения*,
- *конфликты* (для анализируемого в рамках корпуса жанра короткой научной статьи конфликты в аргументации не свойственны ввиду монологического характера и ограниченного объёма статьи):
 - логический конфликт – прямой конфликт между двумя утверждениями, когда одно (атакующее) входит в явное противоречие с другим (атакуемым);
 - атака на аргумент – тип конфликта, когда атакующим утверждением оспаривается не некое конкретное утверждение, а переход от одного утверждения к другому (оспаривается корректность связи между посылкой и заключением в атакуемом аргументе, хотя посылка, обосновывающая заключение, может рассматриваться корректной сама по себе и не оспариваться).

Пример разграничения двух типов конфликтов представлен на рисунке 3.

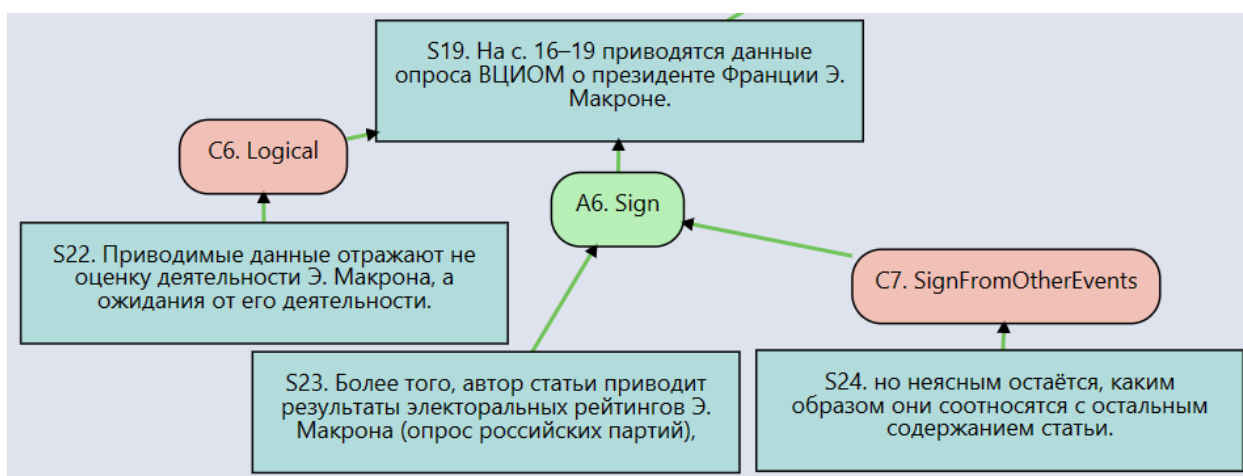


Рис. 3. Два типа конфликтов в аргументации.

Конфликт, обозначенный С6, отражает прямое возражение, выраженное в утверждении S22, на утверждение S19: в атакуемом утверждении заявляется приведение данных об отношении к президенту Франции, тогда как в атакующем подчёркивается, что указанные данные характеризуют не оценку деятельности президента, а лишь ожидания от его деятельности. Конфликт под индексом С7 вводит возражение в утверждении S24 на переход от утверждения S23 к S19: хотя содержание S23 не оспаривается, сомнению подвергается его релевантность исследованию. Пример на рисунке взят из текста рецензии на научную статью (не принятую к публикации в журнале).

2.3.1.1. Актуальные схемы аргументации

Используемый для разметки инструмент платформы ArgNetBank Studio поддерживает поиск аргументов с заданными схемами (такая возможность облегчает обучение разметчиков на совокупности примеров из уже размеченного корпуса). Для жанра короткой научной статьи не все схемы, доступные разметчикам, в равной степени актуальны. Принятая ориентация на создание корпуса для машинного обучения также вносит определенные ограничения на разнообразие применяемых схем в разметке, в том числе на учет \ не учет отдельных деталей в формулировке аргумента, поскольку излишняя детализация (как следствие – увеличение числа классов с близким содержанием) может затруднить последующую автоматическую классификацию аргументов. В таблице 1 представлены часто реализуемые в научных текстах схемы, покрывающие около 98% всех аргументов в размеченном корпусе. В примерах и формализованном описании одним и тем же цветом помечены фрагменты текста и пропозициональные переменные формализованного описания, что облегчает понимание сути использованного в аргументе типа рассуждения и его идентификацию в последующей разметке.

Таблица 1. Актуальные для коротких научных статей схемы аргументации

Схема, формализованное описание	Неформальное описание	Пример
<p><i>1. Analogy</i> <u>Premise</u>: А истинно (ложно) в случае С1. <u>Premise</u>: В общем, случай С1 похож на случай С2. <u>Conclusion</u>: А истинно (ложно) в случае С2.</p>	<p>– вводит тезис, корректность которого обосновывается через его сходство с некоторым другим заявленным тезисом.</p>	<p>Pr: Подобный принцип моделирования был использован при построении лексико-семантической карты славянских языков (по коэффициенту близости больших параметрических ядер лексики) в работе А.А. Кретьова и И.А. Меркуловой [16].</p> <p>Con: При помощи корреляционных коэффициентов была построена карта имени <i>coronavirus</i> по данным корпуса NOW (рис. 3), где длина линии, соединяющей все идиомы, указывает на</p>

		<p>тесноту их связи: чем короче линия, тем выше коэффициент корреляции между ними и теснее связь [12], [13]</p>
<p><i>2. Applied Method</i> <u>Premise:</u> метод E применяется в области D, содержащей пропозицию A. <u>Premise:</u> A было получено методом E. <u>Conclusion:</u> A истинно (ложно).</p>	<p>Обосновывает корректность описываемых результатов через указание особенностей метода, с помощью которого эти результаты получены. Под методом в данном случае могут пониматься как конкретные алгоритмы, так и наборы техник из анализируемой предметной области, так и данные, на которых получен заявленный результат.</p>	<p>Pr: Пропозициональный анализ избран нами в качестве продуктивного метода выявления семантики 50 текстов газетных спортивных новостей, отобранных из британского национального корпуса.</p> <p>Pr: Качественный и количественный анализ исследуемых текстов установил ситуации, которые могут являться экстралингвистической основой ядерного элемента суперструктуры текстов спортивных новостей – основного спортивного события.</p> <p>Con: Наиболее частотным типом ситуации, актуализируемой основным событием в исследуемых текстах, является итог соревновательной деятельности, отражающий победу, поражение, ничью, получение медалей и т.п.</p>
<p><i>3. Cause to Effect</i> <u>Premise:</u> A является причиной B. <u>Premise:</u> В этом случае A имеет место. <u>Conclusion:</u> В этом случае B будет иметь место.</p>	<p>самая общая причинно-следственная связь, применяемая в следующих целях:</p> <ul style="list-style-type: none"> – для указания на последовательность процедурных этапов (шагов алгоритма); – для соединения выполняемых процедур и технического описания их результатов (без оценочного толкования); – для указания на 	<p>Pr: Стремительное развитие информационных технологий в корне изменило отношение к традиционному процессу перевода.</p> <p>Pr: Сама же история исследований и разработок в данной сфере насчитывает уже более пятидесяти лет.</p>

	<p>последовательность событий во времени;</p> <ul style="list-style-type: none"> – для организации абстрактного рассуждения (если для аргументативно связанных тезисов не определяется иная схема аргументации); – для развития оценочных блоков или блоков практического рассуждения (если посылка причины уже содержит положительный или отрицательный оценочный компонент / компонент необходимого действия (т.е. является выводом для схем Positive/Negative Consequences или Practical Reasoning), а вывод к данной посылке более подробно характеризует такой компонент). 	<p>Соп: С середины семидесятых годов двадцатого века наблюдается устойчивый рост интереса к машинному переводу.</p>
<p><i>4. Effect To Cause</i> <u>Premise:</u> A является причиной B. <u>Premise:</u> В этом случае B будет иметь место. <u>Conclusion:</u> В этом случае A имеет место.</p>	<ul style="list-style-type: none"> – выражает причинно-следственную связь в её обратном рассмотрении (от следствий к причинам); – посылка обозначает явление, причина которого выражается в выводе; – схема обозначает переход от более поверхностной идеи к глубинной (поверхностная идея формулируется первой для перехода к глубинной на её примере). 	<p>Pr: Для грамматического оформления системы понадобились специальные маркеры (классные показатели).</p> <p>Pr: В даргинском языке 1 класс обслуживает маркер в, 2 класс обслуживает маркер р, 3 класс обслуживают маркеры б, д.</p> <p>Соп: В современном даргинском языке представлены 3 класса: 1 класс – класс мужчин (лица, обозначающие мужской пол), 2 класс – класс женщин (лица, обозначающие женский пол), 3 класс – всех остальных.</p>
<p><i>5. Example</i> <u>Premise:</u> В этом случае индивид A имеет свойство F, а также свойство G. <u>Premise:</u> а типично для вещей, имеющих F и имеющих или не имеющих G. <u>Conclusion:</u> Как правило, если</p>	<p>схема объединяет конкретный пример и иллюстрируемый им общий тезис, обозначаемый выводом. Посылка может включать следующие элементы:</p> <ul style="list-style-type: none"> – указание на частный случай некоего общего явления; – конкретное описание 	<p>Pr: Так, например, всем известны такие «магические» числа, как «три», «семь», «девять», «двенадцать», «тринадцать».</p> <p>Pr: Эти числа отражают</p>

<p>x имеет свойство F, то (обычно, вероятно, типично) x также имеет свойство G.</p>	<p>алгоритма; – представление используемой формулы; – указание на иллюстративный материал (таблицы, рисунки) без явного словесного комментария, интерпретации. В большинстве случаев посылки являются листьями графа (за исключением блоков подробных примеров с собственной аргументационной структурой)</p>	<p>особенности мышления, мироощущения, религиозные суеверия людей с давних времен. Соп: Издавна числа наделялись людьми некоей магией.</p>
<p><i>6. Expert Opinion</i> <u>Premise</u>: Источник E это эксперт в области D, к которой относится пропозиция A. <u>Premise</u>: E утверждает, что A истинно (ложно). <u>Conclusion</u>: A истинно (ложно).</p>	<p>– вводит тезис за авторством явно указываемого авторитета; – вводит тезис с явным указанием на авторитетность его источника; – вводит тезис в формулировке из цитируемого источника</p>	<p>Pr: Наблюдение относительно воздействующих особенностей рекламы подтверждается мнением авторитетных специалистов. Pr: «Рекламное воздействие по определению является манипулятивным, так как изменение структуры сознания покупателя происходит [...] в результате принятия на веру эмоционально, экспрессивно вводимой не прямой и не исчерпывающей информации» [1, с.67]. Соп: Актуализация тех или иных мотивов в рекламе относится к приемам подсознательного стимулирования, когда отношение аудитории к рекламируемому объекту формируется с помощью различных представлений (стереотипов, мифов, имиджей), автоматически вызывающих в массовом сознании положительную реакцию.</p>

<p>7. <i>Logical</i></p> <p>Схема поддерживает две роли компонентов:</p> <p><u>Conflicted statement</u> – опровергаемый тезис. <u>Conflicting statement</u> – опровергающий тезис.</p>	<p>– общая схема для случаев опровержения (а не поддержки) одного тезиса другим.</p> <p>– вводит тезис через указание противоречия между этим тезисом и неким иным, введённым отдельно (через контрастирование этих тезисов).</p> <p>– Ввиду жанровой специфики коротких научных статей (монологический характер, ограниченный объём) в них редко реализуются конфликты между тезисами.</p>	<p><u>Conflicted statement:</u> Нельзя отрицать, что нынешние относительно устойчивые часто употребляемые словосочетания можно назвать современными фразеологизмами.</p> <p><u>Conflicting statement:</u> Одним из трёх важных параметров, характеризующих рассматриваемые выражения, является устойчивость (а это прямая ассоциативная параллель с историей).</p>
<p>8. <i>Negative Consequences</i></p> <p><u>Premise:</u> если осуществится A, то, вероятно, возникнут плохие последствия. <u>Conclusion:</u> A не следует осуществлять.</p>	<p>обозначает причинно-следственную связь, где</p> <p>– вывод либо посылка явно содержит отрицательный оценочный компонент;</p> <p>– вывод описывает отрицательный результат действия/ситуации в посылке.</p>	<p>Pr: Но при этом, все вышеперечисленные методики имеют один общий недостаток: качественную оценку (экспертную оценку) определения угроз безопасности информации.</p> <p>Con: На основании вышеизложенного, можно сделать вывод о том, что применение только качественных методик оценки угроз безопасности информации недостаточно.</p>
<p>9. <i>Part to Whole</i></p> <p><u>Premise:</u> B подвид (часть) A. <u>Premise:</u> B имеет свойство G. <u>Conclusion:</u> A имеет свойство G.</p>	<p>– обозначает меронимическую связь (от части к целому).</p> <p>– реализуется при анализе составных сущностей через анализ их составных частей.</p>	<p>Pr: Чалдини дает читателям уникальный взгляд на психологические процессы, лежащие в основе наших решений и способность убеждать.</p> <p>Pr: В работе выделены шесть фундаментальных принципов, которые оформляют наши повседневные решения.</p> <p>Con: Глубокое</p>

		погружение в мир психологии убеждения представлено в работе Роберта Чалдини «Influence: The Psychology of Persuasion».
<p><i>10. Popular Opinion</i> <u>Premise:</u> общепринято, что A истинно. <u>Conclusion:</u> A истинно.</p>	<p>– вводит тезис с указанием (возможно неявным) на его распространённость ("известно, что" / "как известно" / "давно известно, что"). – вводит тезис за авторством неопределённого источника.</p>	<p>Pr: Совершенно очевидно, что в настоящее время когнитивное направление в лингвистике заявило себя как новая научная парадигма, имеющая свой объект и инструментальный анализа.</p> <p>Con: Это направление предполагает исследование языковых явлений в совершенно новом аспекте.</p>
<p><i>11. Popular Practice</i> <u>Premise:</u> Если A – общепринятая практика среди знакомых с тем, что допустимо или нет в отношении A, то это даёт повод думать, что A допустимо. <u>Premise:</u> A это общепринятая практика среди тех, кто знаком с тем, что приемлемо или неприемлемо в отношении A <u>Conclusion:</u> A – допустимое действие.</p>	<p>– вводит тезис с предписанием некоего действия (выбором конкретного алгоритма, исследования определённой проблемы) ввиду его частого совершения иными агентами.</p> <p>– посылка первого типа (о том, что общепринятость A влечёт его допустимость) практически не выражается явно в текстах выбранного жанра.</p>	<p>Pr: Модель «программное обеспечение как услуга» (SaaS) подразумевающая использование программного обеспечения, установленное на сервере, в настоящее время наиболее популярна среди пользователей.</p> <p>Con: Для организации образовательного процесса по информатике и другим дисциплинам хорошо подходят публичные, общественные или гибридные сервисы облачных технологий.</p>
<p><i>12. Position to Know</i> <u>Premise:</u> E утверждает, что A истинно (ложно). <u>Premise:</u> E известно об истинности A. <u>Conclusion:</u> A истинно (ложно).</p>	<p>– вводит тезис за авторством явно указываемого источника без эксплицитного указания на его экспертность (например, обращение к носителям языка, не являющимся лингвистами, в лингвистических работах); – вводит собственный тезис автора статьи, который подаётся как предположение.</p>	<p>Pr: На наш взгляд, для получения прозрачной картины необходимо применять тексты разных переводчиков.</p> <p>Con: Важным условием создания корпусов для переводоведческих задач является использование</p>

	– посылка второго типа (Е известно об истинности А) практически не выражается явно в текстах выбранного жанра.	переводов разных переводчиков.
<p><i>13. Positive Consequences</i></p> <p><u>Premise</u>: Если А осуществится, то это вероятно, приведет к хорошим последствиям.</p> <p><u>Conclusion</u>: А следует осуществить.</p>	<p>обозначает причинно-следственную связь, где</p> <ul style="list-style-type: none"> – вывод либо посылка явно содержит положительный оценочный компонент; – вывод описывает положительный результат действия/ситуации в посылке. 	<p>Pr: Подобный анализ позволит учитывать изменения в ответах пользователя.</p> <p>Con: Таким образом, необходимо проводить не только нейросетевой анализ тональности каждого сообщения, но и анализ групп сообщений.</p>
<p><i>14. Practical Reasoning</i></p> <p><u>Premise</u>: Осуществление В есть способ осуществить А.</p> <p><u>Premise</u>: Цель состоит в осуществлении А.</p> <p><u>Conclusion</u>: В следует осуществить.</p>	<p>обозначает причинно-следственную связь, где</p> <ul style="list-style-type: none"> – вывод обозначает некое действие (или процедуру, или метод) с прагматическим компонентом должностования / предпочтительности этого действия. – посылка указывает цель или способ достижения цели без включения модального компонента практичности 	<p>Pr: Читателю требуются определенные навыки прочтения публицистического текста, включая умение согласовать содержание с названием для получения подтекста [6].</p> <p>Pr: Для формирования у школьников и студентов такого навыка следует типологизировать модели экспликации в заголовке смыслов для рассмотрения на уроках анализа текста [7]</p> <p>Con: Предпринятое исследование преследовало цель выявить продуктивные средства иллюстрации в заголовках.</p>
<p><i>15. Sign</i></p> <p><u>Premise</u>: А истинно в данной ситуации</p> <p><u>Premise</u>: как правило, истинность знака А, обозначающего В, указывает на истинность В.</p> <p><u>Conclusion</u>: В истинно в этой ситуации.</p>	<ul style="list-style-type: none"> – доказываемый тезис и свидетельство его правильности оперируют сущностями разного уровня, дополняют друг друга как означающее и означаемое (между ними нет причинно-следственных, гипонимических, меронимических связей). – обосновывает корректность тезиса через приведение фактов 	<p>Pr: Одно из самых весомых событий мирового масштаба произошло в 1997 году, когда компьютер "IBM Deep Blue" победил действующего чемпиона мира по шахматам Гарри Каспарова.</p> <p>Pr; Победа машины над человеком ознаменовала</p>

	(например, статистических сведений) в его поддержку.	<p>важный поворотный момент для ИИ.</p> <p>Соп: Текущий период можно охарактеризовать как время "зарождающихся технологических прорывов в самом широком спектре областей", включая искусственный интеллект.</p>
<p><i>16. Verbal Classification</i></p> <p><u>Premise:</u> Для всех x, если x имеет свойство F, то x можно отнести к имеющим свойство G.</p> <p><u>Premise:</u> A имеет свойство F</p> <p><u>Conclusion:</u> A имеет свойство G.</p>	<p>– посылки содержат общие тезисы, основание классификации и/или свойство принадлежности к классу;</p> <p>– вывод – обозначение общего класса;</p> <p>– как правило, связь используется при определении нескольких классов;</p> <p>– в случае параллельного использования схемы побочные выводы (классы вне рассмотрения, приводимые для контраста с анализируемым классом) могут не вести к основному тезису;</p>	<p>Pr: Каждый уровень напрямую зависит от элементов содержания, обеспечивающих достижение эквивалентности.</p> <p>Pr: Он выделил пять уровней (типов) эквивалентности: цели коммуникации, описания ситуации, высказывания, сообщения и языковых знаков.</p> <p>Соп: Только при достижении идентичности на всех уровнях содержания текста оригинала и текста перевода, перевод можно считать эквивалентным.</p>

Выбор схемы осуществляется с учётом факторов трёх типов:

- *пропозиционального содержания* утверждений в аргументе (семантика схемы должна соотноситься с семантикой связанных тезисов, отражать логику перехода от посылок к заключению в основе их связи);
- *контекста* в выстраиваемой аргументационной структуре текста (для утверждений, входящих в состав анализируемого аргумента, следует учитывать их предполагаемые связи и схемы в других их включающих аргументах, поскольку такие внешние схемы могут влиять на формулировку утверждений в анализируемом аргументе, например, вводить в их состав маркеры, не относящиеся к данному аргументу);
- *маркированности*: наличия маркеров конкретных схем с учётом их однозначности (маркеры схемы могут различаться по прочности связи с этой схемой, некоторые маркеры могут соответствовать нескольким схемам либо не всегда соотноситься с аргументативным содержанием) и позиционной специфики (некоторые маркеры для одной схемы могут встречаться преимущественно в посылках либо в заключении, но не в обоих типах утверждений).

В случаях, когда обозначенные факторы указывают на разные схемы, следует основывать выбор на семантическом содержании утверждений, затем на контексте в аргументационной структуре, лишь затем учитывать маркеры. Причиной этому является неоднозначность маркеров и возможность их нетипичного использования (так, утверждение вида «например, X заявляет о» может оказаться посылкой по схеме *Expert Opinion*, а не *Example*). Однако если контекст и маркеры указывают на одну схему, а пропозициональное содержание может быть отнесено и к другой, предпочтительнее выбирать первую.

2.3.1.2. Уточнение типов связи

Для уточнения типов связи используют следующие приемы:

1. Дерево вопросов.

2. Правила, построенные на основе функциональной классификации схем.

1) Правила, уточняющие семантику схем из разных функциональных групп.

2) Правила, уточняющие семантику схем из одной функциональной группы.

1. Дерево вопросов для выбора одной схемы из пятнадцати самых частотных, указанных в таблице 1 (за исключением конфликтной схемы *Logical Conflict*). Номера после ответов указывают на дальнейшие уточняющие вопросы).

(1) Выражает ли аргументативная связь причинно-следственное отношение?

Да: (2). Нет: (7).

(2) Содержит ли аргументативная связь компонент практического содержания?

Да: (3). Нет: (6).

(3) В рассуждении выражена положительная/отрицательная оценка результатов?

Да: (4). Нет: (5).

(4) Эксплицированная в рассуждении оценка является положительной?

Да: **Positive Consequence**. Нет: **Negative Consequences**.

(5) Практический компонент сосредоточен в представлении метода, которым достигнуты результаты (да), или проявляется в сочетании поставленной цели и способа её достижения (нет)?

Да: **Applied Method**. Нет: **Practical Reasoning**.

(6) Рассуждение строится от наблюдаемых явлений к их причинам?

Да: **Effect to Cause**. Нет: **Cause to Effect**.

(7) Основано ли рассуждение на обращении к некоторой точке зрения?

Да: (8). Нет: (10).

(8) Обозначен ли конкретный источник, предлагающий эту точку зрения?

Да: (9). Нет: (**Popular Opinion**).

(9) Является ли источник приводимой точки зрения авторитетом?

Да: **Expert Opinion**. Нет: **Position to Know**.

(10) Содержит ли заключение указание действия, корректность которого обосновывается распространённой практикой в посылке?

Да: **Popular Practice**. Нет: (11).

11) Обращается ли рассуждение к изучаемым объектам через их классификацию?

Да: **Verbal Classification**. Нет: (12).

(12) Основана ли связь на иллюстрации заключения через рассмотрение частного случая?

Да: (13). Нет: (15)

(13) Описывает ли заключение более общую ситуацию по отношению к случаю в посылке (да), или же в рассуждении акцентируется сходство частных ситуаций (нет)?

Да: (14). Нет: **Analogy**.

(14) Приводится ли указанный частный случай как конкретный пример?

Да: **Example**. Нет: **Sign**.

15) Возможна ли интерпретация сущности в посылке как составной части сущности в заключении?

Да: **Part To Whole**. Нет: **Схема не из списка**.

2. Правила, построенные на основе функциональной классификации схем.

Расхождения между разметчиками на уровне схем часто связаны с разноаспектным восприятием типа связи. Пример такого расхождения представлен на рисунке 4: один разметчик охарактеризовал связи между утверждениями как классификацию (по формальному представлению списка) и указал схему *Verbal Classification*, тогда как другой интерпретировал эти связи согласно смысловому соотношению посылок и заключения, а потому использовал модель *Negative Consequences*.

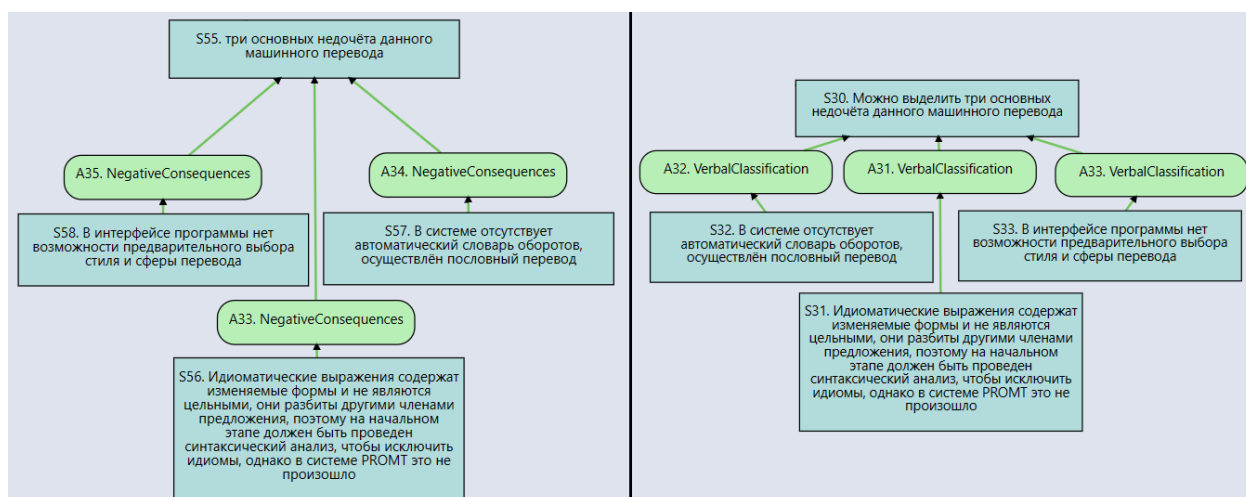


Рис. 4. Пример расхождения разметчиков в выборе схем.

Ввиду разноаспектного восприятия связей разметчиками часто смешиваются не функционально близкие схемы, а друг от друга далёкие. Поэтому, зафиксировав функциональную классификацию схем (см. таблицу 2), можно *проверить схему на потенциальную возможность реализации в каждой функциональной группе* [Пименов И.С. 2022]. Учёт всех функциональных групп для отдельных связей обеспечит их многоаспектное восприятие, а группировка схем уменьшит количество вариантов для проверки.

Таблица 2. Состав функциональных аргументационных групп из детализированных схем:

Функциональная аргументативная группа	Схемы аргументации
От авторитета (From Authority)	<i>Expert Opinion, Popular Opinion, Popular Practice, Position to Know</i>
От практической значимости (From Practical Value)	<i>Applied Method, Practical Reasoning, Positive / Negative Consequences</i>
Причинно-следственные переходы (Causal Analysis)	<i>Cause to Effect, Effect to Cause</i>
Через детализацию (Through Elaboration)	<i>Analogy, Example, Part to Whole, Sign, Verbal Classification</i>

Правила, уточняющие семантику схем из разных функциональных групп. Выбор схемы для систематически возникающих разногласий осуществляется согласно следующим правилам:

1) *Effect to Cause* или *Example*?

Отличие схемы *Effect to Cause* от схемы *Example* состоит в том, что в первом случае при реализации перехода от поверхностной идеи к глубинной поверхностная идея

формулируется для перехода к глубинной как её частное проявление; (для схемы *Example*, наоборот, некоторая идея иллюстрируется конкретным частным примером).

2) *Cause to Effect* или *Positive Consequences* или *Practical Reasoning*?

Схему *Cause to Effect* отличает от её частных аналогов (*Positive Consequences*, *Practical Reasoning*) нейтральность посылки, отсутствие в выводе оценочного комментария / модальности долженствования, иначе связь характеризуется не схемой *Cause to Effect*, а её аналогами. Маркеры тональности могут быть полезны в этом случае.

3) *Part to Whole* или схема из другой функциональной группы.

При выборе между схемой *Part to Whole* и другой схемой, уместной для анализируемого рассуждения, следует предпочесть вторую схему ввиду её большей конкретности (модель с обоснованием некоторого свойства целого по свойству его части характеризуется наибольшей общностью значения: основания для её указания могут быть выявлены для широкого ряда случаев).

Правила, уточняющие семантику схем из одной функциональной группы.

1. From Authority

Различия между схемами *Expert Opinion*, *Position to Know* и *Popular Opinion* передаются противопоставлением по двум основаниям: явность указания на конкретный источник ((*Expert Opinion* или *Position to Know*) vs *Popular Opinion*) и авторитетность источника (*Expert Opinion* vs (*Position to Know* или *Popular Opinion*)).

2. From Practical Value

Схемы с практическим компонентом разграничиваются по авторской расстановке акцентов в представлении некоторого действия: подчёркиваются ли его последствия (позитивные или негативные), обосновывается ли достоверность полученных результатов надёжностью способа их достижения (*Applied Method*), выводится ли в фокус внимания конкретная цель, определяющая способ (*Practical Reasoning*).

3. Causal Analysis

Cause to Effect или *Effect to Cause*

Выбор между двумя причинно-следственными схемами определяется акцентуацией в авторском построении рассуждения: требуется понять, ориентировано ли доказательство на обоснование причины через представление её следствия либо же наоборот. В этом случае особенно значимым выступает рассмотрение широкого контекста (аргументов, смежных с анализируемым, общей цепочки рассуждения в макроструктуре аргументации).

4. Through Elaboration

Различие между схемами *Example*, *Part to Whole* и *Verbal Classification* может быть представлено в терминах онтологий через разграничение экземпляров и классов. Указание некоторых подклассов заданного класса соответствует схеме *Verbal Classification*, тогда как приведение экземпляров класса соответствует *PartToWhole*, если таких экземпляров потенциально ограниченное число, и *Example*, если нет явного ограничения их возможного числа.

Example или *Verbal Classification*: схема *Verbal Classification* отличается от *Example* контекстной спецификой: во-первых, посылки при параллельном использовании схемы *Verbal Classification* могут развиваться иными аргументационными схемами, в то время как посылки примеров часто являются листьями в графе либо служат выводом к таким листьям, комментариям к примерам; во-вторых, для классификаций типично параллельное указание нескольких классов через большое количество однотипных посылок, тогда как для примеров в жанре короткой научной статьи более характерно приведение только одного примера к доказываемому тезису.

Для разграничения схем *Example* и *Sign* важно оценить смысловое соотношение ситуаций в посылке и заключении. Первая модель покрывает случаи частных примеров к более общим явлениям, тогда как вторая специализирована на содержательной импликации

(одна ситуация *подразумевает* другую). Для схемы *Analogy* значима акцентуация на сходстве двух ситуаций без их противопоставления как разноуровневых.

2.3.2. Объединение подграфов в связный граф

При объединении подграфов блоков в связный граф следует опираться на формальную организацию текста, выстроенную авторами. Подграф заключения (заключительного абзаца) часто является прикорневой частью связного графа, а корнем – одно из утверждений этого блока. Посылки к корневой вершине (главному тезису) – обычно выводы из других блоков текста.

Объединение подграфов отдельных блоков может производиться либо через исходящую связь от главного тезиса блока (не являющегося посылкой ни для одного утверждения в этом блоке), либо через входящую связь к одному из листьев подграфа блока (утверждению, не являющемуся заключением для других утверждений в этом смысловом блоке). Внешних связей, соединяющих данный блок с остальным текстом, может быть несколько. Не следует нарушать логическую связность абзаца через вставку внешних утверждений (из иных абзацев) между его тезисами. Исключением являются случаи, когда ввиду авторской организации текста такое построение связей является единственным возможным вариантом разметки.

Поскольку расхождения между разметчиками становятся сильнее с увеличением расстояния между тезисами в тексте, желательно избегать частого построения связей между позиционно удалёнными утверждениями (на расстоянии трети и более объема текста). Каждую такую связь следует проверять на возможность её разбиения на несколько связей меньшей длины через указание промежуточного утверждения. Однако если такое утверждение в тексте отсутствует, либо не получается определить явную модель рассуждения для одной из двух новых связей (от исходной посылки к промежуточному утверждению, от него к исходному заключению), то следует сохранить связь между отдалёнными утверждениями. В частности, дальние связи возможны при присоединении к главному тезису, выраженному в заключении, логического блока, рассматриваемого в самом начале текста, например, формулировки задачи исследования и обоснования её актуальности.

2.3.3. Проверка разметки, нумерация утверждений и схем аргументации

Проверка разметки текста является действием полезным, но не обязательным. В ходе проверки производится нумерация утверждений и схем (отдельно). Наличие нумерации позволяет оценить объем аргументации в тексте, насколько последовательно она организована в отдельных логических блоках, как строится аргументация в блоках: от формулировки тезиса к посылкам или наоборот. Наличие нумерации также упрощает автоматическую обработку аргументационных графов, поскольку явно передаёт сведения о линейном порядке утверждений в тексте и может использоваться, например, для выявления текстов с непоследовательным представлением тезисов.

3. Выбор и подсчет коэффициентов согласия

Стандартно количественно измерять степень согласия между аннотаторами в Argument Mining принято с помощью следующих коэффициентов согласия: каппа Коэна, альфа Криппендорфа, каппа Флейсса [M Teruel, C Cardellino, F Cardellino. 2018]. Важными параметрами критерия является число разметчиков, полнота данных (все ли тексты размечены всеми участвующими в разметке аннотаторами). Достоинствами коэффициента альфа Криппендорфа [Krippendorff, Klaus. 2013] являются следующие: применимость к любому количеству аннотаторов, неполным (отсутствующим) данным, кроме того, он

подстраивается под небольшие размеры выборки данных. Принято вычислять коэффициенты согласия отдельно для результатов разметки на каждом из трех этапов: 1) выбора утверждений, 2) построения связей между утверждениями, 3) установления типа связи. Интерпретация значений коэффициентов предлагается в работе [Landis, J. et al. 1977]: значения между 0.01 и 0.20 – небольшое согласие, 0.21 и 0.40 – удовлетворительное, 0.41 и 0.60 – умеренное, 0.61 и 0.80 – существенное, а 0.81 и 1 почти идеальное согласие.

Дискуссионным является и вопрос о необходимости учета случайного согласия. Для анализа расхождений в разметке нами был разработан и применен алгоритм вычисления коэффициента аргументативного сходства (КАС), не учитывающий случайные совпадения [Пименов И.С. 2023].

4. Инструкция пользователя

Детализированная разметка текстов на русском языке доступна на платформе ArgNetBank Studio (<https://uniserv.iis.nsk.su/arg/corpora>) [Сидорова Е.А. и др., 2020]. Инструментарий платформы позволяет: 1) создавать коллекции текстов, 2) производить аргументационную разметку текстов, 3) визуализировать размеченные структуры аргументации в виде графов, 4) экспортировать и импортировать размеченные структуры аргументации в машиночитаемом формате .json и пр.

1) Для создания коллекций текстов в режиме просмотра корпусов доступна функция «Создать подкорпус», представленная отдельной кнопкой. Название корпуса может быть свободно изменено в любой момент через функцию «Переименовать» во всплывающем меню при нажатии правой кнопкой мыши по созданной папке (соответствующей корпусу). Это всплывающее меню содержит и функцию «Загрузить текст» для добавления текстов коллекции в виде текстовых данных (иллюстрации и сложные объекты вроде формул не поддерживаются). Загруженные тексты не могут быть изменены (ограничение обусловлено связью текстов с созданной разметкой). Если требуется внести изменения в текст, следует загрузить текст повторно. Простой вариант текста может быть перемещен в другую папку через нажатие на его название с удерживанием кнопки мыши.

2) Аргументационная разметка текста, загруженного в коллекцию, предполагает переход в режим проекта из режима просмотра корпусов. В первую очередь следует создать проект, который будет соответствовать варианту разметки текста, по нажатию на зеленую кнопку «Создать проект». При переходе в режим проекта по умолчанию открывается графический режим, поддерживающий построение аргументационного графа, на уровне просмотра. Для внесения изменений в граф следует нажать на кнопку «Редактировать проект».

2.1) Создание аргументационных утверждений достигается выделением сегмента текста в левом окне (с содержанием текста) мышкой и нажатием на пустое место в правом окне (где будет представлен аргументационный граф). Двойное нажатие на созданную вершину с утверждением позволяет отредактировать его, одиночное с удерживанием — переместить по экрану.

2.2) Создание аргументационных связей доступно по нажатию кнопки «Соединить вершины» под кнопкой с сохранением проекта. Для создания связи следует нажать на одну вершину и, удерживая кнопку мыши зажатой, нажать на вторую. Если обе вершины соответствуют утверждениям, то между ними появится связь в виде двух стрелок с вершиной-схемой между ними. Направление стрелок определяется порядком нажатия на вершины (от первой ко второй). Если хотя бы одна из двух вершин является вершиной-схемой, будет построена одиночная стрелка без новой вершины схемы (однако зажатие

клавиши «Shift» при проведении связи позволяет построить связь с вершиной-схемой и двумя смежными стрелками, что полезно при моделировании атак на аргументы).

2.3) Для указания аргументационной схемы в основе связи требуется двойное нажатие на вершину-схему. В окне «Категория» следует указать тип схемы из всплывающего списка («Inference» для аргументов в поддержку, «Conflict» для опровержений), а затем указать конкретную схему из всплывающего списка в окне «Схема». После этого появятся окна, соответствующие компонентам аргумента (заключению и посылкам разных типов в зависимости от схемы). В этих окнах можно указать роль каждого из утверждений, соединенных стрелкой с вершиной-схемой.

2.4) Сохранение проекта в режиме редактирования возможно после указания точных схем и ролей задействованных в них утверждений для каждого аргумента. Рекомендуется регулярно сохранять проект по мере аннотирования текста ввиду риска утраты внесённых изменений при утрате интернет-соединения или закрытии браузера.

3) Экспорт размеченных текстов в формате .json для машинной обработки доступен двумя способами. Первый способ — отдельно для каждого текста, через кнопку с дискетой в списке проектов для этого текста (список проектов открывается изнутри режима проекта кнопкой «Открыть список проектов»). Второй способ — в режиме просмотра корпусов через нажатие правой кнопкой мыши на имя текста или всего корпуса в целом и выбор функции «Экспортировать с проектами» во всплывающем меню (в этом случае будет сгенерирован и сохранён один общий .json для нескольких вариантов разметки или всех текстов корпуса).

Список литературы

- Пименов И.С.** Сочетаемость аргументов разных функциональных групп в научных текстах // Филологические науки. Вопросы теории и практики. Тамбов: Грамота, 2022. № 11. С. 3672–3680.
- Пименов И.С.** Анализ расхождений в аргументационной разметке научных статей на русском языке // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация, 2023, т. 21, № 2.
- Сидорова, Е.А.** и др. Платформа для исследования аргументации в научно-популярном дискурсе / Е.А. Сидорова, И.Р. Ахмадеева, Ю.А. Загорюлько, А.С. Серый, В.К. Шестаков // Онтология проектирования. – 2020. – Т. 10, №4(38). – С. 489-502.
- Amossy, R.** Argumentation in discourse. A socio-discursive approach to arguments // *Informal Logic*, 29(3), 2009, pp. 252–267.
- Krippendorff, Klaus** (2013). *Content analysis: An introduction to its methodology*, 3rd edition. Thousand Oaks, CA: Sage.
- Landis, J. Richard and Gary G. Koch.** 1977. The measurement of observer agreement for categorical data. *Biometrics*, 3:159–174.
- Lawrence J., Reed C.** Argument Mining: A Survey // *Computational Linguistics*, Vol. 45 (4), 2019, pp. 765–818.
- Rahwan I., Reed C.** The argument interchange format // *Argumentation in artificial intelligence*, Rahwan I. and Simari G., Eds. Springer, 2009, pp. 383–402.
- M Teruel, C Cardellino, F Cardellino.** 2018. Increasing argument annotation reproducibility by using inter-annotator agreement to improve guidelines. *Proceedings of the ...*, 2018 - aclanthology.org.
- Walton D., Reed C., Macagno F.** *Argumentation schemes* // New York: Cambridge University Press, 2008, 443 p.