

ЛИНГВИСТИЧЕСКИЕ АСПЕКТЫ ИССЛЕДОВАНИЯ АРГУМЕНТАЦИИ НА ОСНОВЕ ОНТОЛОГИИ

Кононенко Ирина Семеновна

Н.с., e-mail: irina_k@cn.ru,

Ахмадеева Ирина Равильевна

М.н.с., e-mail: i.r.akhmadeeva@iis.nsk.su,

Сидорова Елена Анатольевна

К.ф.-м.н., с.н.с., e-mail: lsidorova@iis.nsk.su,

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

630060 г. Новосибирск, пр. Лаврентьева 6.

Аннотация. Обсуждаются вопросы аннотирования корпуса русскоязычных научно-популярных текстов. Особое внимание уделяется схеме разметки, в которой предметом аннотирования являются не только прагматические и семантические аспекты аргументации, но и лингвистические индикаторы аргументации. Предлагается представление индикаторов в виде лексических и лексико-грамматических шаблонов, которые автоматически формируются по размеченному фрагменту текста, а затем при необходимости уточняются экспертом вручную по результатам поиска в корпусе текстов и анализа всех соответствующих фрагментов. Полученный словарь индикаторов используется для поиска и автоматического выделения индикаторов аргументации в неразмеченном тексте.

Ключевые слова: онтология аргументации, аргументативная разметка, схема аргументации, индикатор аргументации, онтологическая модель разметки текста, словарь индикаторов.

Цитирование: Кононенко И.С., Ахмадеева И.Р., Сидорова Е. А. Лингвистические аспекты исследования аргументации на основе онтологии // Информационные и математические технологии в науке и управлении. 2020. № 4 (20). С. 44-55. DOI: 10.38028/ESI.2020.20.4.004

Введение. В последнее десятилетие появилась потребность в решении ряда прикладных задач автоматической обработки текста, требующих обращения к глубинным уровням представления дискурса для анализа приведенных в тексте аргументов - мнений и их обоснований. Основой развития теории аргументации и предпосылкой успеха в области автоматического распознавания аргументов является создание корпусов текстов значительного объема, снабженных аргументативной разметкой. Такого рода ресурсы развиваются для разных языков на базе текстов различных жанров и функциональных стилей, однако для русского языка аргументативно размеченные корпуса текстов до сих пор не созданы. Ниже представлены некоторые результаты работы в рамках исследовательского проекта, направленного на создание корпуса русскоязычных научно-популярных текстов с разметкой аргументации.

В процессе предварительного анализа научно-популярных текстов замечено, что нередко аргументы вводятся в текст с помощью явных текстовых индикаторов. О присутствии

в тексте аргумента могут сигнализировать дискурсивные коннекторы (*потому что, именно, следовательно, и т.п.*) или свободные индикаторные конструкции. В качестве примера таких конструкций для рассуждений, ссылающихся на мнение экспертов (авторитетных в данной предметной области), можно привести конструкцию с ментальным предикатом и авторитетом в позиции субъекта:

*Недавно **ученые** из Университета Йорка **показали, что шимпанзе все-таки способны выучивать новые звуки.***¹

Одним из важнейших аспектов аннотирования аргументации является регистрация индикаторов аргументации, которые играют ключевую роль в процессе понимания дискурса, облегчая идентификацию и реконструкцию шагов аргументации, осуществляемых в аргументативном диалоге или тексте [12].

Основы индикаторного подхода к извлечению информации из текста были заложены еще в 70-е годы прошлого столетия (см. достаточно детальное описание в [2]). Извлечение онтологической информации из научных текстов в [4] основано на индикаторном методе, в основе которого лежит обнаружение в тексте подсказок в виде словесных клише (образцов, маркеров), которые являются показателями аспектов содержания текста. В последнее время появились исследования, направленные на выявление шаблонных индикаторов структуры дискурса: так, в [11] описываются основные модели образования свободных конструкций, сигнализирующих о наличии причинно-следственного риторического отношения в дискурсе.

В работе [8] предлагается пилотный вариант инструкции по аннотированию схем аргументации в корпусе аргументированных эссе, уже снабженных разметкой аргументативной структуры. Для решения новой, существенно более сложной, задачи предлагается использование лингвистических индикаторов, которые приписаны отдельным схемам в неформальной инструкции для аннотаторов и могут служить им подсказками при разметке в текстах схем рассуждения (в особенности речь идет о сильных, т.е. однозначных индикаторах). Однако нам неизвестны проекты создания корпусов текстов, в которых предметом аргументативного аннотирования являлись бы не только структуры и схемы аргументации, но и лингвистические индикаторы и ставилась бы задача автоматического выделения индикаторов в неразмеченных текстах.

Целью данного исследования является обеспечение возможности использования индикаторов аргументации при решении задачи автоматического распознавания аргументов. Для этого при разметке текстов корпуса производится первичное ручное аннотирование индикаторов аргументации и создание словаря индикаторов, который далее используется для автоматического обнаружения индикаторов в еще не размеченных текстах по лингвистическим свойствам текстовых фрагментов (наличие лексического маркера, грамматического или пунктуационного признака) и/или свойствам, связанным с онтологией аргументации (т.е. по связи с типовой схемой рассуждения), что, в свою очередь, позволяет привлечь внимание аннотатора к потенциально аргументативным фрагментам размечаемого текста.

1. Модель аргументативной разметки текста. Любая разметка текста опирается на схему, которая задает перечень сущностей, используемых при аннотировании, их типизацию,

¹ В примере индикаторы выделены жирным шрифтом, а заключительное утверждение подчеркнуто.

формат и интерпретацию. В нашей работе в качестве схемы для разметки аргументации используется формат AIF (Argument Interchange Format) [6] и онтология, построенная на его основе [3, 9].

Отличительной особенностью этой онтологии является ее ориентированность на графовое представление аргументации, а также наличие разветвленной системы классов для метаописания сущностей предметной области. Таким образом, в онтологии AIF можно выделить три базовые концепции:

$$O = \langle O^G, O^C, O^F \rangle, \text{ где}$$

O^G – графовое представление аргументации, содержащее вершины Node и бинарные отношения между ними. В классе Node выделяются два типа вершин, S-node, предназначенные для визуализации аргументов, и I-node – для визуализации утверждений, входящих в состав аргументов;

O^C – предметная область аргументации, содержащая классы типовых схем рассуждения (аргументов), классы утверждений и отношения между ними, а также экземпляры этих классов. Эта онтология реализует Уолтоновский подход [13] к представлению аргументации на основе аргументативных схем, которые формально описывают внутреннюю структуру аргументов;

O^F – онтология метаописаний, содержащая классы дескрипторов (форм) для схем аргументации и их атрибутов. Эта онтология позволяет задавать языковое описание схем аргументов.

На рис. 1 приведен пример описания типовой схемы рассуждения, построенного по онтологическому представлению класса *CauseToEffect_Inference* и его формы.

Аргумент «CauseToEffect_Inference»		
Роль	Тип утверждения	Описание утверждения
Causal_Premise	Causal_Statement	A causes B
Conclusion	OccurrenceB_Statement	In this case, B will occur
OccurrenceA_Premise	OccurrenceA_Statement	In this case, A occurs

Рис. 1. Описание типовой схемы аргумента *CauseToEffect*.

Модель аргументативной разметки корпуса текста, в соответствии с заданной онтологией, можно представить как следующую систему:

$$M^C = \langle T, O^{CA}, A \rangle, \text{ где}$$

T – класс, предназначенный для представления текстов, включая информацию о его источнике, жанровой принадлежности и набор мета-атрибутов (*автор*, *дата_публикации* и т.п.);

$O^{CA} = CS \cup CA \cup C^R \cup P, O^{CA} \in O^C$ – часть онтологии аргументации, содержащая набор классов размечаемых сущностей:

$$CS = UCS_i \text{ – конечное непустое множество классов утверждений,}$$

$CA = UCA_i$ – конечное непустое множество схем (классов) аргументов,

$C^R = UC^R_i$ – конечное непустое множество классов конфликтных отношений;

$P = DT \cup R \cup RC$ – множество свойств классов онтологии, где

DT – конечное множество атрибутов простых типов, данных V , и

$F^{DT}: CSUCAUC^R \rightarrow 2^{DT \times V}$ – функция, которая определяет имена и типы атрибутов для классов онтологии,

$R = Ua_i$ – конечное множество ролей (атрибутов) в структуре схем аргументов, и

$F^R: CA \rightarrow 2^{R \times (CSUCA)}$ – функция, которая определяет имена и типы ролей для классов аргументов (отметим, что значением атрибута (роли) в аргументе могут выступать как утверждения, так и другие аргументы);

$RC = Urc_i$ – конечное множество связей конфликтных отношений, и

$F^{RC}: C^R \rightarrow 2^{R \times (CSUCA)}$ – функция, которая определяет имена и типы ролей для конфликтных отношений (конфликт может быть, как между утверждениями, так и между утверждением и аргументом);

$A = UA_t$ – множество аннотаций и $F^A: T \rightarrow 2^A$ – функция, сопоставляющая тексту множество его аннотаций $F^A(t) \subseteq A$.

Отличительной особенностью рассматриваемой модели аннотирования текста в сравнении с другими схемами представления аргументации (например, AIF) является то, что в предлагаемую схему явным образом вводятся компоненты для разметки индикаторов. Таким образом, аннотация текста представляется системой вида:

$A_t = \langle Fr_t, S_t, Arg_t, Ind_t \rangle$, где

$Fr_t = Ufr_i$ – множество размеченных текстовых фрагментов (в общем случае фрагмент может быть разрывным и представляться мультиинтервалом);

$S_t = Us_i$, $s_i \in CS_i$ – множество размеченных утверждений:

$s_i = \langle desc, ws \rangle$, где

$desc$ – выраженная на естественном языке пропозиция, или пропозициональное содержание текстового фрагмента, размеченного в тексте, и

$F^S: S \rightarrow 2^{Fr}$ – функция, сопоставляющая утверждению множество фрагментов текста (в общем случае утверждение может несколько раз встречаться в тексте),

ws – числовая характеристика, сопоставленная утверждению и характеризующая степень уверенности автора в его истинности или достоверности.

$Arg_t = Ua_i$ – множество аргументов, найденных в тексте, где a_i класса $CA_i \in CA$ имеет набор ролей r_i , связывающих аргумент с утверждениями и другими аргументами:

$a_i = U_j(r_j, s_j)U_k(r_k, a_k)$, так, что $(r_j, s_j) \in F^{SA}(CA_i)$, $(r_k, a_k) \in F^{SA}(CA_i)$.

$Ind_t = U(ind_i, tm_i)$, $ind_i \in M$, $tm_i \in T_m$ – множество индикаторов ind_i различных аспектов аргументации tm_i , и

$R^{AM}: ArgUS \rightarrow 2^{M \times T_m}$ – функция, сопоставляющая аргументу или утверждению, входящему в состав аргумента, набор индикаторов и аспектов аргументации.

Множество T_m определяет различные типы соотношения индикатора и сигнализируемых им аспектов аргументации:

- 1) тип аргументативного отношения (поддержка vs. конфликт);
- 2) структура аргументации (множественная vs. последовательная аргументация);

- 3) сила или убедительность аргумента;
- 4) отношение вывода между двумя утверждениями (наличие аргументации);
- 5) роль утверждения в отношении вывода (посылка vs. заключение);
- 6) семантико-онтологическое отношение, на котором основана применяемая в данном случае типовая схема рассуждения;
- 7) степень уверенности автора в утверждении.

Аспекты 1-4 сигнализируют о наличии аргументации, что в случае ручной разметки фокусирует внимание эксперта на соответствующих фрагментах текста, а при автоматической – требует наличия других показателей для более точного выявления класса аргумента. Аспекты 4-6 соотносят индикатор не столько с аргументом, сколько с утверждением (или несколькими утверждениями) в структуре аргумента. Аспекты 3, 7 влияют на оценку убедительности аргументации.

Таким образом, аннотация состоит из множества фрагментов текста, каждому из которых поставлено в соответствие утверждение или индикатор. Аргументы определяются как n-местные отношения над размеченными утверждениями. Роль индикаторов заключается в фиксации свойств и границ аргументов и/или утверждений, входящих в их состав.

2. Индикаторы. Индикаторы аргументации – это используемые в дискурсе языковые средства (слова и конструкции), которые служат подсказками при определении структуры аргументации: они помогают установить наличие аргументов и их составляющих в данном сегменте текста, реконструировать связи между утверждениями, соотносят аргумент с определенной схемой рассуждения (формой вывода, выражающей взаимосвязь посылок и заключений).

Индикаторы различаются не только по типу семантической и прагматической информации (что отмечено выше), но и по степени грамматикализации (первичные и вторичные индикаторы, см. [11]), по семантике опорного знаменательного слова (например, в них используются такие лексико-семантические классы, как речевые и ментальные предикаты, предикаты вывода и ментального воздействия, модальные слова, выражающие различные степени убежденности в высказываемом мнении) и по типу конструкции (см. классификацию и примеры в [7]).

2.1. Язык описания индикаторов. Множество индикаторов, образующих словарь индикаторов, можно создать с помощью специального синтаксиса, формальное описание которого в расширенной форме Бэкуса-Наура представлено ниже:

```
<словарь индикаторов> ::= {<шаблон>}
<шаблон> ::= <имя индикатора> '=' (<альтернатива> ['->' <атрибуты>])+ ['=>' <атрибуты>]
<альтернатива> ::= <цепочка> | <разрыв>
<цепочка> ::= '[' <индикатор> (';' <индикатор>) ']'
<разрыв> ::= '[' <ограничение> ':' <индикатор> {';' <ограничение> ':' <индикатор>} ']'
<индикатор> ::= (<имя индикатора> | <простой индикатор>) [<ограничения на атрибуты>]
<простой индикатор> ::= <тип простого индикатора> '/' <описание простого индикатора>
<тип простого индикатора> ::= 'w' | 'ph' | 't' | 's'
<ограничение> ::= 'begin' | 'end' | 'include' | 'not_include' | 'begin!' | 'end!' | 'main'
<ограничения на атрибуты> ::= <атрибуты>
<атрибуты> ::= '<' <атрибут> {';' | '><'} <атрибут> '>'
```

<атрибут> ::= <имя атрибута> [=] <значение> {, <значение>}

Можно выделить два основных способа задания индикаторов: индикаторы с простой структурой *<простой индикатор>* и составные конструкции, описываемые с помощью шаблонов *<шаблон>*. Словарь индикаторов состоит из множества шаблонов, которые в своем описании используют индикаторы с простой структурой, а также ссылаются на другие составные индикаторы.

К индикаторам с простой структурой относятся однословные и многословные дискурсивные коннекторы, и полнозначная лексика (предикаты вывода и причины, речевые и ментальные предикаты и т.п.).

Описание составных индикаторов включает несколько альтернатив, каждая из которых может быть цепочкой индикаторов *<цепочка>* или разрывной конструкцией *<разрыв>*. Любой из альтернатив (или сразу всем) при желании можно приписать грамматические или семантические значения *<атрибуты>*. Для простых индикаторов значения атрибутов извлекаются автоматически при анализе текста. Полученные значения атрибутов можно в дальнейшем использовать для фильтрации вхождений *<ограничения на атрибуты>*.

3. Индикаторы как инструмент исследования аргументации. Для поддержки исследования аргументации была разработана платформа, которая включает набор веб-инструментов, обеспечивающих создание корпусов текстов, визуализацию используемой онтологии аргументации (т.е. схем аргументации и их метаописаний), построение графа аргументации, поиск различных сущностей в корпусах текстов в терминах онтологии и т.п.

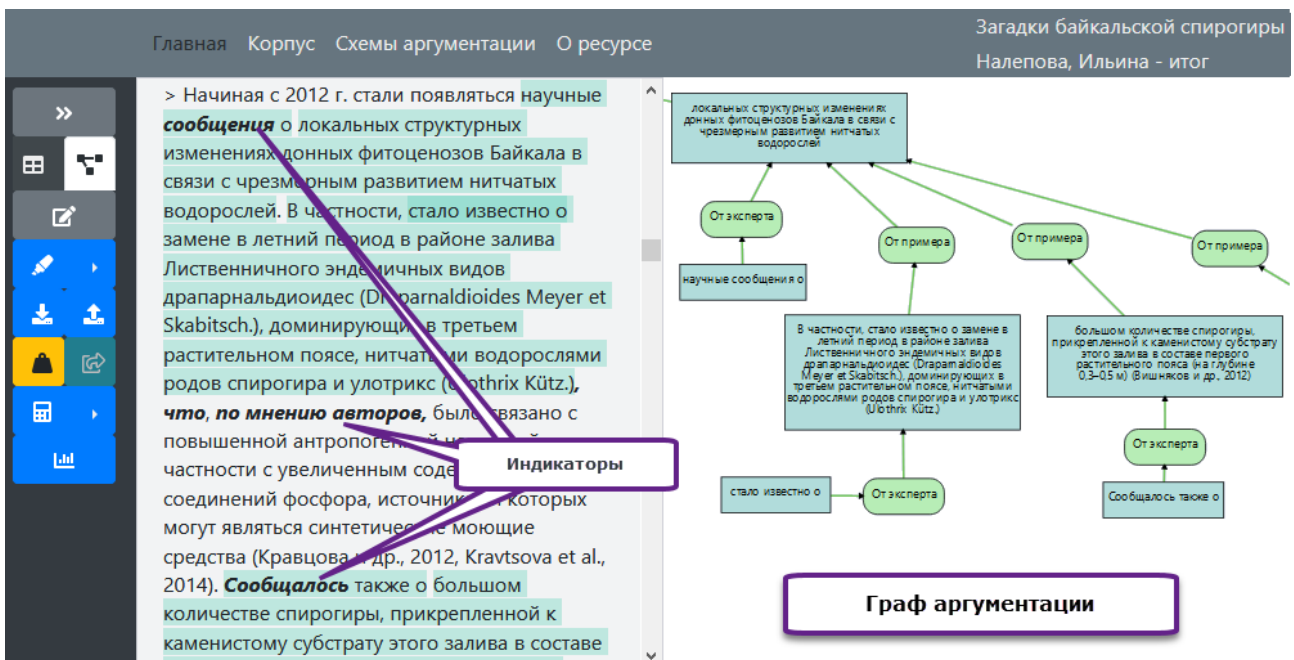


Рис. 2. Выделение индикаторов в аргументативной разметке текста.

В рамках данного исследования были созданы дополнительные инструменты для поиска и исследования аргументации в текстах на основе индикаторов. Созданные инструменты обеспечивают следующие функциональные возможности:

- подключение словаря индикаторов аргументации, созданного на основе риторических маркеров, и автоматический поиск и выделение в текстах индикаторов аргументации в режиме аргументативной разметки текста и построения графа аргументации (рис. 2),
- исследование адекватности выявления особенностей аргументации с помощью индикаторов на основе корпусного исследования и построения конкордансов.

3.1. Формирование словаря индикаторов. Формирование словаря индикаторов происходит на базе ручной разметки индикаторов. На рис.3 приведены основные этапы процесса создания и исследования индикаторов (блоки со светлым фоном сопоставлены с полностью автоматическими процедурами, блоки с темным фоном - с процедурами, осуществляемыми экспертом).

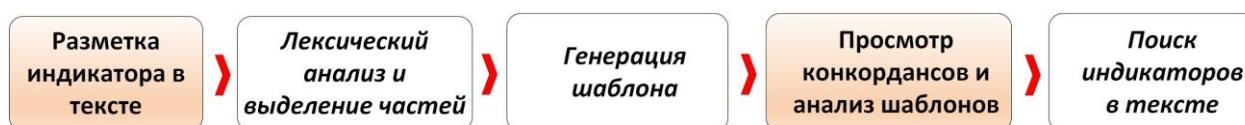


Рис. 3. Основные этапы формирования словаря индикаторов.

Автоматически можно сформировать шаблоны для первичных индикаторов (функциональных слов) и задать базовое структурное определение шаблонов для вторичных индикаторов, включающих опорное слово. Далее полученное формальное описание индикаторов исследуется экспертом. С этой целью осуществляется поиск индикаторов в тексте, построение конкорданса.

Для исследования индикаторов аргументации были созданы веб-инструменты, обеспечивающие:

- визуализацию словаря индикаторов, просмотр состава индикатора, поиск индикаторов и/или их составных частей в тексте,
- создание новых индикаторов и пополнение словаря индикаторов пользователем,
- поиск индикаторов и/или их частей в выбранном(ых) корпусе, построение конкордансов и визуализация расширяемых контекстов.

3.2. Пополнение словаря. Список знаменательных слов, которые могут служить индикаторами или опорными словами индикаторных конструкций, разнороден и принципиально неполон, как далек от полноты и словарь шаблонов индикаторных конструкций. Поэтому при разметке аннотатор сталкивается с индикаторными фрагментами, не покрытыми текущим словарем индикаторов. В этом случае у аннотатора есть возможность создать пользовательский шаблон, выделив соответствующий фрагмент текста (возможно, разрывный).

На основе размеченного аннотатором фрагмента автоматически генерируется шаблон, в котором зафиксированы состав и нормализованная форма лексических единиц, знаки препинания и разрывы (если пользователь выделил разрывный фрагмент).

Так, на рис.4 в числе прочих представлен пользовательский шаблон *президент ...отметил, что*, который не был автоматически выделен идентичным по структуре шаблоном *expert_opinion*, поскольку слово *президент* не было заранее отнесено к лексико-

семантическому классу **expert**. В данном случае, не являясь авторитетным представителем лингвистической науки, президент, как представитель власти, является экспертом в области языковой политики, о которой идет речь в аннотируемом тексте, так что ввод данного пользовательского шаблона вполне правомерен и шаблон удален не будет.

Рис. 4. Исследование индикаторов в тексте и пополнение словаря.

Кроме того, рис.4 демонстрирует еще один пользовательский шаблон *проректор по науке*, который также не был автоматически выделен в качестве индикатора типа экспертного мнения, но по другой причине – ввиду отсутствия речевого глагола. Такая структура (использующая пунктуацию прямой речи без речевого глагола) характерна для жанра интервью и в таком жанре имеет полное право на существование: соответствующий шаблон является кандидатом на пополнение словаря индикаторов.

3.3. Корпусное исследование индикаторов. Для расширения и обобщения лексического состава шаблона проводится анализ конкорданса (множества всех контекстов) шаблона и его частей.

При проведении исследований могут выполняться следующие шаги:

- формирование гипотез об обобщении отдельных опорных слов и создание лексико-семантических классов (так, в результате анализа аргументов, представляющих экспертное мнение, все глаголы, лежащие в основе конструкций прямой и косвенной речи, объединены в лексико-семантический класс речевых предикатов);
- проверка по конкордансу гипотезы об обобщении опорных слов разных классов в рамках одного шаблона (так, конкорданс показывает функциональное тождество и общность конструкций, формируемых речевыми и ментальными предикатами, что означает целесообразность объединения этих классов в рамках одного шаблона **_opinion**);

- создание служебных шаблонов с альтернативами (так, рассмотрение контекстов уже введенных лексических индикаторов **_opinion**, представленных речевыми и ментальными предикатами, позволяет увидеть альтернативные способы заполнения актантных позиций, т.е. средства ввода актанта, представляющего экспертное мнение, и сформировать варианты служебных шаблонов для прямой и инверсной актантных позиций, **_opinion_act_dir**, см. рис. 5, и **_opinion_act_inv**, которые далее выступают в качестве подшаблонов альтернативных вариантов полной индикаторной конструкции **expert_opinion**);
- корректировка и проверка всех вхождений созданных индикаторов в корпус.

The screenshot shows a search interface with the following components:

- Parameters search:** A tree view on the left with checkboxes for various indicators. The indicator **_opinion_act_dir** is checked. Below the tree are search buttons: 'Искать' (Search), a pause button, and a close button. A status bar indicates '76 результатов' (76 results).
- Corpus:** A list of sources with checkboxes. 'STRF' is checked and highlighted with a blue circle and the number '76'.
- Search Results Table:** A table with columns: 'Левый контекст' (Left context), 'Вхождение' (Occurrence), 'Правый ко...' (Right context), 'Текст' (Text), 'Шаблон' (Template), and 'Кор...' (Corpus). The table contains 10 rows of search results, including text snippets and the corresponding template **_opinion_act_dir**.
- Navigation:** At the bottom, there are navigation buttons: 'Назад' (Back), 'Страница 2 из 8' (Page 2 of 8), '10 строк' (10 lines), and 'Вперед' (Forward).

Рис. 5. Поиск индикаторов в корпусе текстов.

Заключение. Данное исследование проводилось в рамках проекта по созданию программного комплекса, предназначенного для поддержки исследования аргументации в русскоязычных научно-популярных текстах.

Предложено представление индикаторов аргументации в виде лексических единиц и лексико-грамматических шаблонов. Процедура автоматизированного создания словаря индикаторов аргументации позволяет на основе размеченного текстового фрагмента автоматически сформировать предварительную формальную спецификацию индикатора и уточнить ее вручную по результатам поиска в корпусе и анализа всех соответствующих ей фрагментов. Полученные индикаторы используются для поиска и автоматического выделения индикаторов аргументации в неразмеченном тексте. Это облегчает процесс аннотирования и исследования текстов, поскольку привлекает внимание эксперта-аннотатора к определенным фрагментам текста, потенциально содержащим аргументы.

Кроме того, проведены эксперименты в области автоматического распознавания аргументации, в которых индикаторы, извлекаемые с помощью достаточно надежных лексико-грамматических шаблонов (демонстрирующих точность более 70 %, см. [1]), используются в качестве признаков для машинного обучения при решении задачи автоматического извлечения аргументов [5]. В работе [10] предложен подход к частичному восстановлению аргументативной структуры текста, который может быть реализован при

недостатке размеченных коллекций. Утверждения, содержащие аргументы, выявляются автоматически на базе индикаторов аргументации. В условиях дефицита больших размеченных коллекций, с пополнением арсенала индикаторных шаблонов и расширением словарей индикаторов данный метод показывает себя вполне работоспособным.

Статья подготовлена по итогам исследования, проведённого в рамках проекта Российского фонда фундаментальных исследований № 18-00-01376 (18-00-00889).

СПИСОК ЛИТЕРАТУРЫ

1. Ахмадеева И.Р., Кононенко И.С., Саломатина Н.В., Сидорова Е.А. Подход к построению шаблонов индикаторов для извлечения аргументов из научно-популярных текстов // Тр. Международной конференции “Знания - Онтологии - Теории” (ЗОНТ-2019). Институт математики им. С. Л. Соболева СО РАН. Новосибирский государственный университет. 2019. С. 24-32.
2. Блюменау Д.И., Гендина Н.И., Добронравов И.С., Лахути Д.Г., Леонов В.П., Федоров Е.Б. Формализованное реферирование с использованием словесных клише (маркеров) // Научно-техническая информация. Сер.2. 1981. №2. С. 16–20.
3. Загорюлько, Ю.А., Гаранина Н.О., Боровикова О.И., Доманов О.А. Моделирование аргументации в научно-популярном дискурсе с использованием онтологий // Онтология проектирования. 2019. Т. 9. №4 (34). С.496-509.
4. Саломатина Н.В., Гусев В.Д. Автоматизация формирования индикаторных словарей и возможности их использования // Труды межд. конференции Диалог-2006 «Компьютерная лингвистика и интеллектуальные технологии». Бекасово. 2006. Москва: "Наука". С. 121–125.
5. Akhmadeeva I., Kononenko I., Salomatina N., Sidorova E. Indicator Patterns as Features for Argument Mining // 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). Novosibirsk. Russia. 2019. Pp. 0886-0891.
6. Chesñevar C.I., McGinnis J., Modgil S., Rahwan I., Reed C., Simari G., South M., Vreeswijk G., Willmott S. Towards an argument interchange format. The knowledge engineering review 2006. № 21(4). Pp. 293-316.
7. Kononenko I., Sidorova E. Development of the Lexicon of Argumentation Indicators. In: Kuznetsov S., Panov A. (eds) Artificial Intelligence. RCAI 2019. Communications in Computer and Information Science. vol 1093. Springer. Cham. 2019. Pp. 154-168.
8. Musi E., Ghosh D., Muresan S. Towards Feasible Guidelines for the Annotation of Argument Schemes// Proceedings of the third workshop on argument mining (ArgMining 2016). Pp. 82-93.
9. Rahwan I, Banihashemi B, Reed C, Walton D, Abdallah S. Representing and classifying arguments on the semantic web. The Knowledge Engineering Review 2011. №26(4). Pp.487-511.
10. Salomatina N.V., Kononenko I.S., Sidorova E.A., Pimenov I.S. Identification of connected arguments based on reasoning schemes “from expert opinion” // Journal of Physics: Conference Series (в печати).
11. Toldova S., Pisarevskaya D., Vasilyeva M., Kobozeva M. The cues for rhetorical relations in Russian: "Cause-Effect" relation in Russian Rhetorical Structure Treebank // Computational

Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue”. 2018. Issue 17. Pp.747-761.

12. Van Eemeren F.H., Houtlosser P., and F. Snoeck Henkemans. Argumentative Indicators in Discourse: A Pragma-Dialectical Study // Dordrecht: Springer. 2007.

13. Walton D, Reed C, Macagno F. Argumentation schemes. Cambridge: Cambridge University Press. 2008.

UDK 004.82:004.89:519.816

LINGUISTIC ASPECTS OF ONTOLOGY-BASED ARGUMENTATION STUDY

Irina S. Kononenko

Researcher, e-mail: irina_k@cn.ru,

Irina R. Akhmadeeva

Junior researcher, e-mail: i.r.akhmadeeva@iis.nsk.su,

Elena A. Sidorova

PhD, Senior researcher, e-mail: lsidorova@iis.nsk.su,

A.P. Ershov Institute of Informatics Systems

630090, Russia, Novosibirsk Acad. Lavrentjev pr., 6.

Abstract. Annotation of the corpus of Russian-language popular science texts is discussed. Particular attention is paid to the annotation scheme, in which the objects of annotation are not only pragmatic and semantic aspects of argumentation, but also their linguistic indicators. The presentation of argumentation indicators in the form of lexical and lexical-grammatical patterns is proposed, which are automatically generated from a marked-up fragment of text, and then, if necessary, are manually specified by an expert based on the results of a search in the corpus and analysis of all relevant fragments. The resulting dictionary of indicators is used to search and automatically highlight indicators of argumentation in unannotated text.

Keywords: ontology of argumentation, argumentation annotation, argumentation scheme, argumentation indicator, ontological model of text markup, lexicon of indicators.

References

1. Akhmadeeva I.R., Kononenko I.S., Salomatina N.V., Sidorova E.A. Podhod k postroeniju shablonov indikatorov dlja izvlechenija argumentov iz nauchno-populjarnyh tekstov [Approach to building indicator patterns for extracting arguments from popular science texts] // Tr. Mezhdunarodnoj konferencii “Znanija - Ontologii - Teorii” (ZONT-2019) = International Symposium “Knowledge - Ontology - Theory” (KONT-19). Sobolev Institute of Mathematics. Novosibirsk State University. 2019. C. 24-32. (in Russian).
2. Bljumenau D.I., Gendina N.I., Dobronravov I.S., Lahuti D.G., Leonov V.P., Fedorov E.B. Formalizovannoe referirovanie s ispol'zovaniem slovesnyh klishe (markerov) [Formalized abstracting using verbal cliches (markers)] // Nauchno-tehnicheskaja informacija = Scientific and technical information. Vol.2. 1981. №2. Pp. 16–20.1. (in Russian).

3. Zagorulko, Yu.A., Garanina N.O., Borovikova O.I., Domanov O.A. Modelirovanie argumentacii v nauchno-populjarnom diskurse s ispol'zovaniem ontologij [Argumentation modeling in popular science dis-course using ontologies] // *Ontologija proektirovanija = Ontology of designing*. 2019. T. 9. №4 (34). Pp. 496-509 (in Russian).
4. Salomatina N.V., Gusev V.D. Avtomatizacija formirovanija indikatornyh slovarej i vozmozhnosti ih ispol'zovanija [Automation of cue dictionaries formation and their applications] // *Trudy mezhd. konferencii Dialog-2006 «Komp'juternaja lingvistika i intellektual'nye tehnologii»*. Bekasovo = Dialogue-2006 “Computational Linguistics and Intellectual Technologies”. Bekasovo. Moscow. "Nauka". 2006. Pp. 121–125. (in Russian).
5. Akhmadeeva I., Kononenko I., Salomatina N., Sidorova E. Indicator Patterns as Features for Argument Mining // 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). Novosibirsk, Russia. 2019. Pp. 0886-0891.
6. Chesñevar C.I., McGinnis J., Modgil S., Rahwan I., Reed C., Simari G., South M., Vreeswijk G., Willmott S. Towards an argument interchange format. *The knowledge engineering review* 2006. № 21(4). Pp. 293-316.
7. Kononenko I., Sidorova E. Development of the Lexicon of Argumentation Indicators. In: Kuznetsov S., Panov A. (eds) *Artificial Intelligence. RCAI 2019. Communications in Computer and Information Science*, vol 1093. Springer. Cham. 2019. Pp. 154-168.
8. Musi E., Ghosh D., Muresan S. Towards Feasible Guidelines for the Annotation of Argument Schemes// *Proceedings of the third workshop on argument mining (ArgMining 2016)*. Pp. 82-93.
9. Rahwan I, Banihashemi B, Reed C, Walton D, Abdallah S. Representing and classifying arguments on the seman-tic web. *The Knowledge Engineering Review* 2011. №26(4). Pp.487-511.
10. Salomatina N.V., Kononenko I.S., Sidorova E.A., Pimenov I.S. Identification of connected arguments based on reasoning schemes “from expert opinion” // *Journal of Physics: Conference Series*. (to be published).
11. Toldova S., Pisarevskaya D., Vasilyeva M., Kobozeva M. The cues for rhetorical relations in Russian: "Cause-Effect" relation in Russian Rhetorical Structure Treebank // *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference “Dialogue”*. 2018. Issue 17. Pp.747-761.
12. Van Eemeren F.H., Houtlosser P., and F. Snoeck Henkemans. *Argumentative Indicators in Discourse: A Pragma-Dialectical Study* // Dordrecht: Springer. 2007.
13. Walton D, Reed C, Macagno F. *Argumentation schemes*. Cambridge: Cambridge University Press. 2008.