

РАСПОЗНАВАНИЕ СМЫСЛА АНАЛИТИЧЕСКОГО МЕДИАТЕКСТА ФОРМАЛЬНЫМИ МЕТОДАМИ

Современный этап автоматической обработки текстов характеризуется бурным развитием корпусных технологий, дальнейшим совершенствованием традиционных систем машинного перевода, аннотирования и реферирования, а также постоянным вниманием к таким новым направлениям, как распознавание мнений, эмоциональных оценок и состояний коммуникантов, извлечение из текстов знаний различных типов. Эти новые направления фактически направлены на выявление прагматики текста – такого семиотического измерения, которое вплоть до настоящего времени считалось невозможным формализовать для алгоритмической обработки.

Разработанные нами ранее алгоритмы поиска мнений участников медийного общения, включая автора текста [1], базируются, в первую очередь, на анализе таких поверхностных сигналов, как лексико-синтаксические структуры прямой и косвенной речи и далее – на лексических и грамматических средствах, маркирующих высказывания мнения. В качестве следующего шага в формализации прагматики медиатекста представляется целесообразным рассмотреть возможности распознавания основной идеи, или смысла целого текста (в качестве соответствующего понятия в современном научном и разговорном обиходе часто используется английское слово *message*). Рассмотрим вначале основные подходы к пониманию термина *смысл текста* в современной лингвистике.

Прежде всего, следует заметить, что необходимо разграничивать термины *смысл текста* и *семантика текста*. С точки зрения решения алгоритмических задач, представляется возможным рассматривать понятие семантики текста как синоним его содержания. Сравните в этом отношении следующее определение: «семантику текста составляет структура его содержания как целостное и системное образование, возникающее в интеллекте человека, в его отношении к линейной структуре текста» [2, с. 11]. Сходную точку зрения высказывает А. И. Новиков: с позиций разработанного им денотатного подхода, семантическую структуру текста (его внутреннюю форму)

составляет «совокупность денотатов, связанных предметными отношениями в целостный семантический комплекс, представляющий собой модель ситуации, задаваемую языковыми средствами данного языка» [3, с. 117]. В его понимании содержание – это ментальное образование, базирующееся на денотативных (референтных) структурах, в то время как смысл представляет собой мысль о действительности, интерпретацию того, что сообщается в тексте [4]. Тем не менее в ряде работ понятия семантики текста и его содержания разграничиваются. Так, семантическая структура связывается со знаковым характером текста, содержательная – с кодированием заключенной в нем информации, а смысловая – с декодированием, т.е. с восприятием произведения как целого [5, с. 69].

Важным для настоящего исследования является дифференциация таких прагматических видов информации, как содержательно-фактуальная (СФИ) и содержательно-концептуальная информации (СКИ), данная И. Р. Гальпериным. Второе из вышеуказанных понятий, имеющее непосредственное отношение к смыслу всего текста, определяется как «индивидуально-авторское понимание отношений между явлениями, описанными средствами СФИ, понимание их причинно-следственных отношений, их значимости в социальной, экономической, политической, культурной жизни народа. <...> Такая информация извлекается из всего произведения и представляет собой творческое переосмысление указанных отношений, фактов, событий, процессов» [6, с. 28].

Соотношение содержания и смысла зависит от типа текста. Можно согласиться с А. И. Новиковым в том, что «минимальное расхождение содержания и смысла характерно для текстов научной, технической тематики, для деловой прозы, а максимальное – для различных жанров художественных текстов» [4, с. 112]. Безусловно, наиболее сложным представляется процесс определения смысла художественного текста, который выявляется не только на основе лингвистических параметров, но и комплекса экстралингвистических факторов. Этой проблеме посвящен целый ряд работ в функциональной стилистике и лингвистике текста (А. Н. Васильева, Г. И. Богин, Н. А. Купина, Н. С. Болотнова и др.). Что касается таких типов текстов, как публицистические и разговорные, то их смысловые структуры и глобальные смыслы остаются наименее изученными.

Как следует из вышеприведенных определений смысла целого текста, он представляет собой авторскую интерпретацию текста, в большинстве случаев его словесная формулировка не дается в готовом виде. Вместе с тем можно предположить, что в каждом тексте есть некоторые поверхностные (т.е. формальные) признаки, сигнализирующие о его глубинных структурах, в том числе и об общем (глубинном) смысле. К таким признакам относятся конкретные слова, лексико-семантические группы слов, словосочетания или лексико-синтаксические структуры. Именно на таких признаках и могут базироваться алгоритмы распознавания главного смысла того или иного типа текста, на основе которого создаются компьютерные программы.

В нашей концепции коммуникативно-когнитивного моделирования текстовых жанров понятия семантики и прагматики текста рассматриваются

в неразрывном единстве [1], поскольку семантика текста без прагматики является неполной, а прагматика без семантики не существует. Семантика текста в нашей модели анализируется на основе понятия суперструктуры, предложенной Т. ван Дейком [7]. Хотя суперструктура постулируется этим автором как формальная схема дискурсивного жанра, в конкретном анализе текстов газетных новостей, наряду с формальными категориями, т.е. категориями композиции (*заголовок, резюме, лид* и др.), он использует семантические категории (*предшествующие события, фон, вербальные реакции* и др.). Используя понятие суперструктуры, мы наделяем его семантическими свойствами и определяем как общую структуру содержания текста, т.е. как семантическую структуру.

Конкретный анализ текстов показал, что каждый жанр обладает типовой суперструктурой, включающей семантические компоненты, характерные для подавляющего большинства текстов данного жанра [8; 9; 10; 11; 12]. При этом установлено, что семантические составляющие репрезентируются в текстах с помощью определенных языковых маркеров, что дает возможность их алгоритмического распознавания [8; 12] и построения сжатых аналогов исходных текстов – аннотаций и рефератов.

В прагматически-ориентированных алгоритмах анализа текста, например, поиска мнений или определения тональности текста, нами используется понятие *семантико-прагматическая структура*. Его основу составляет рассмотренное выше понятие семантической структуры, модифицированное с точки зрения прагматической значимости того или иного семантического компонента. Прагматическая значимость, или прагматическая «заряженность» семантического компонента суперструктуры определяется на основе совокупности критериев: количества высказываний, выражающих мнения; количества оценочных высказываний; степени эмоциональной насыщенности; наличия непосредственной связи с основной идеей текста. Так, анализ такого вида аналитической статьи, как проблемная статья, показал, что в ее суперструктуре наиболее прагматически нагруженными являются компоненты «описание проблемной ситуации» и «пути решения проблемы». При этом второй компонент содержит мнения не только автора текста (журналиста), но, главным образом, мнения других участников медийной коммуникации – представителей государственных, частных, общественных организаций. В связи с этим, смысл такой разновидности аналитической статьи можно рассматривать как коллективное мнение о путях разрешения проблемы.

В современной прессе аналитические статьи представлены и многими другими субжанрами. Так, распространенным видом являются статьи, которые публикуются под рубриками «Opinion», «Мнение», «Меркаванне», «Погляд» и др., часто подразделяемые на ряд подрубрик. Интересно отметить, что в текстах подрубрики «В стране / Страна» в белорусской прессе мнение автора (журналиста), как правило, выражено эксплицитно, хотя привлекаются точки зрения и других участников медийной коммуникации. В отличие от проблемных статей, в таких статьях рубрики «Мнение» авторская точка зрения превалирует, и она сигнализируется эксплицитными средствами: личные и притяжательные местоимения (*я, мой, мы, наши*), эллиптические струк-

туры с пропущенным подлежащим 1-го лица (*уверен, не верю, понимаю, думаю, рискую, смею, начал, обнаружила* и др.). Например: ... *И, кстати, я не права – интрига остается* (СБ Беларусь сегодня, 6.05.2016); *И здесь именно в музыке я вижу огромный потенциал* (СБ Беларусь сегодня, 18.05.2016).

Суперструктура таких текстов может быть как сходной, так и отличной от проблемной статьи. Например, в статье «Капитальные расчеты», опубликованной в подразделе «В стране» рубрики «Мнение» (СБ Беларусь сегодня, 5.05.2016) рассматривается проблемная ситуация, связанная с отчислениями граждан на капитальный ремонт жилья. Суперструктура этой статьи, как и большинство статей данной рубрики, посвященных внутренним проблемам страны, включает такие компоненты, как *описание проблемной ситуации, формулирование проблемы, выявление ее причин, пути решения проблемы*. Как было отмечено выше, в соответствии с установленной нами ранее семантико-прагматической структурой проблемной статьи одним из наиболее важных и прагматически нагруженных ее составляющих является компонент «пути решения проблемы». Именно в этой структурной части, как показывает анализ, содержится главный авторский смысл рассматриваемого текста, который в концентрированной форме обобщен в последнем предложении следующего текстового фрагмента, передающего мнение журналиста:

*Но если мы собираемся выйти на 100-процентное возмещение затрат на ЖКХ, то уже не отдельные наши читатели, а все начнут задавать вопросы: сколько моих денег собрали и куда они ушли? К полному отчету надо бы готовиться. Кроме того, оценив накопленную сумму, жильцы, возможно, захотят сами выбрать подрядчика. <...> А если к жилью бережно относиться, можно и 50 лет не думать о комплексном капремонте. **Должен быть выбор – вот главный посыл*** (СБ Беларусь сегодня, 5.05.2016).

Проведенный анализ статей раздела «В стране» рубрики «Мнение» показал, что алгоритм распознавания их авторского смысла должен быть инкорпорирован в детальный алгоритм выявления семантико-прагматического компонента «пути решения проблемы», который базируется на комплексе языковых маркеров этого компонента [1]. После установления текстового фрагмента, содержащего точку зрения автора на способы решения обсуждаемой проблемы, в качестве смысла всего текста выбирается его финальный отрезок, который, как правило, представляет обобщенное мнение автора и может дополнительно маркироваться такими языковыми средствами, как *главное; главный вопрос, главный посыл* и др.:

Главное, показать себя на мировом рынке. А лучшие – по-хорошему удивить (СБ Беларусь сегодня, 14.04.2015);

*<...> самый **главный** вопрос – как убедить человека быть в ответе за собственную единственную и бесценную жизнь?* (СБ Беларусь сегодня, 4.05.2015);

В отличие от раздела «В стране» рубрики «Мнение», статьи раздела «В мире» часто имеют иную прототипическую структуру: это последовательность аргументативных блоков, каждый из которых состоит из тезиса, аргументов различных типов и следствия. Например:

<Тезис> С приходом к власти президента Роухани Иран сделал ряд существенных шагов в этом направлении. <Аргумент> В первую очередь это касается снятия санкций в связи с иранской ядерной программой, <Следствие> что позволило Ирану существенно нарастить торгово-экономическое и технологическое сотрудничество с другими государствами (Рэспубліка, 3.05.2016).

Как показал анализ текстов, посвященных ситуации в мире, в большинстве случаев их структура характеризуется четкостью и логичностью развертывания, что выражается в частой оформленности аргументативных блоков в рамках отдельных абзацев. Подобное структурирование придает сходство данного субжанра аналитической статьи с научной статьей и облегчает его алгоритмический анализ. Как любой научный текст, рассматриваемые статьи часто завершаются четким выводом, который можно интерпретировать, на наш взгляд, как выражение главного смысла. При этом такой заключительный вывод может маркироваться не только финальной позицией в тексте, но и маркерами обобщения: *в целом, таким образом, ясно, очень важно; вынікі, у выніку* и др. Например:

*Но **ясно** одно: движение иракских исламистов переживает бурный расцвет, и ожидать его политической кончины пока не приходится (СБ Беларусь сегодня, 8.07.2014);*

Тем не менее может возникнуть эффект домино: кризис распространится на все компании, оказывающие услуги в нефтяной промышленности (СБ Беларусь сегодня, 13.01.2015);

*Можна чакаць, што ў **выніку** мінулых выбараў Еўрапарламент стане менш эфектыўным інструментам еўраінтэграцыі. А ўрады краін, дзе еўраскептыкі дасягнулі важных вынікаў, будуць вымушаны ўлічваць гэты факт у сваёй рабоце ў рамках Еўрасаюза (Звязда, 27.05. 2014);*

***Таким образом,** противостояние «либерального» и консервативного крыла иранского общества входит в новую стадию, и новый состав иранского меджлиса, по всей видимости, станет отражением этой непростой внутренней ситуации (Рэспубліка, 3.05.2016).*

Как показывают приведенные примеры, вывод, или общий смысл статей данного субжанра, наряду с подведением итогов рассмотрения конкретной ситуации, часто включает и прогноз развития описываемого положения дел в будущем.

Несмотря на сходства с научными текстами, рассматриваемые газетные аналитические статьи сохраняют все характеристики публицистического стиля, в частности, его стилистические свойства. Этим объясняется возможность вербального оформления вывода в виде вопроса к читателям, вопросно-ответного блока или риторического вопроса, незавершенного высказывания, перифразы, парцелляции и др.:

Как мне кажется, для нас очень важно вовремя вписаться в концепцию нового Шелкового пути, которую так активно сейчас продвигает Китай. <...> Для нас это может означать инвестиции в новые рабочие места. Настало время собирать камни. Великие в том числе (СБ Беларусь сегодня, 5.05.2015);

Можжа, не трэба абагульненняў, калі іх атрымліваецца зрабіць толькі шляхам замоўчвання і скажэння фактаў? Можжа, трэба імкнуцца крытычна і аб'ектыўна ставіцца не толькі да асобных старонак гісторыі, а выкарыстоўваць такі падыход увогуле? (Звязда, 26.02.2016).

Таким образом, проведенный анализ показал, что формальное распознавание общего смысла медийных статей рубрик «Мнение», «Погляд» и др. в белорусской прессе целесообразно проводить на основе семантико-прагматических структур таких текстов и их языковых маркеров. Установленная в результате первого этапа работы алгоритма конкретная структура будет определять дальнейшее ветвление принципиального алгоритма и выявление наиболее вероятного вербального выражения глобального авторского смысла в структурном компоненте, имеющем наивысший прагматический потенциал.

ЛИТЕРАТУРА

1. Карпилович, Т. П. Семантико-прагматический интерфейс в алгоритмах анализа медиадискурса / Т. П. Карпилович // Вестник МГЛУ. Сер. 1, Филология. – № 4 (71). – 2014. – С. 28–35.
2. Слюсарева, Н. А. Проблемы функционального синтаксиса современного английского языка / Н. А. Слюсарева. – М. : Наука, 1981. – 207 с.
3. Новиков, А. И. Семантика текста и ее формализация / А. И. Новиков. – М. : Наука, 1983. – 215 с.
4. Новиков, А. И. Текст, его содержание и смысл / А. И. Новиков // Текст: Структура и функционирование. – Барнаул, 1997. – С. 108–113.
5. Каменская, О. Л. Текст и коммуникация / О. Л. Каменская. – М. : Высш. шк., 1990. – 152 с.
6. Гальперин, И. Р. Текст как объект лингвистического исследования / отв. ред. Г. В. Степанов. – изд. 8-е. – М. : Книжный дом «Либроком», 2014. – 144 с.
7. Дейк, Т. А. ван. Язык, познание, коммуникация / Т. А. ван Дейк. – М. : Прогресс, 1989. – 312 с.
8. Карпилович, Т. П. Моделирование процесса смысловой компрессии текста: когнитивно-дискурсивный подход / Т. П. Карпилович. – Минск : Изд-во МГЛУ, 2003. – 226 с.
9. Сысоева, Т. А. Прагматическая структура малоформатных медийных текстов / Т. А. Сысоева // Вестник МГЛУ. Сер. 1, Филология. – № 6 (49). – 2010. – С. 130–137.
10. Лешкевич, И. В. Семантическая структура газетных рецензий в британской и американской прессе / И. В. Лешкевич // Медiateкст и культура / Минский гос. лингвист. ун-т; под ред. Т. П. Карпилович. – Минск, 2015. – С. 26–34.
11. Шевцова, А. К. Дискурсивные маркеры компонентов суперструктуры английских и белорусских радиодискуссий / А. К. Шевцова // Актуальные проблемы лингвистики и гуманитарных наук : сб. науч. ст. по материалам междунар. науч.-практ. конф., Москва, 23 марта 2012 г. / Рос. ун-т дружбы народов ; редкол. : Е. В. Сафонова (отв. ред.) [и др.]. – М. : РУДН, 2012. – С. 288–299.
12. Глазко, П. П. Суперструктура публицистического текста как основа его автоматического реферирования / П. П. Глазко // Сб. тр. VI Междунар. науч. конф. «Прикладная лингвистика в науке и образовании», Санкт-Петербург, 5–7 апр. 2012 г. – СПб, 2012. – С. 60–63.