

Дефиниция как компонент содержательной структуры научного текста¹

Аннотация. В работе на теоретическом уровне показано, что модель текста, в качестве элементов которой выступают имеющиеся в нем дефиниции, может отражать содержание более адекватно по сравнению с моделями, построенными на терминах и их отношениях.

Ключевые слова: научный текст, содержательная структура, дефиниция, автоматическая обработка текста.

Введение

Дефиниция является необходимым компонентом научного текста, поскольку именно благодаря ей выполняется такое важное требование к стилю научного изложения, как однозначность выражения. По формулировке Ю.Д. Апресяна, формальность, точность, однозначность – это свойство языка, на котором излагается теория [3]. В самом общем виде можно сказать, что в дефиниции приводятся существенные свойства предмета, благодаря которым возможно его выделение в качестве элемента познаваемой реальности, а также сообщение о предмете в ходе коммуникации по итогам познавательной деятельности. Используемые для дефинирования специальные языковые средства выражают логико-семантические отношения тождества, различия, противопоставления, включения, исключения, причины, следствия и др. [8, 9, 30, 33].

Дефиниция представляет собой интерес для различных наук и дисциплин: логики, лингвистики, терминологии, психологии научного творчества и т.д. [1, 12, 28, 34]. Являясь важным «строевым элементом» как семантики, так и структуры научного текста, дефиниция закономерно оказывается также одним из предме-

тов междисциплинарных исследований при разработке средств компьютерного анализа научной литературы [6, 7, 18, 21, 36-38].

1. Проблема выявления содержательной структуры научного текста

Такие задачи автоматической обработки текста, как реферирование и аннотирование, поиск по запросу, оценка сходства и т.п. подразумевают умение работать с содержательной составляющей публикации. Содержательная (смысловая) структура текста определяется смысловыми связями и отношениями, которые устанавливает автор между используемыми при изложении материала понятиями. Понятия же в тексте выражаются с помощью языковых единиц разного уровня [24]. Для научного стиля изложения характерно применение для выражения понятий терминов – специальных лексических единиц (слов или словосочетаний), из перечня важнейших характеристик которых мы выделим тенденцию к однозначности и наличие дефиниции [31].

Существует точка зрения, согласно которой именно термины выступают в качестве единиц,

¹Работа выполнена при поддержке РФФИ (грант ОФИ_М 14-29-05028).

отражающих содержание научного текста. Поэтому задача извлечения информации может быть сведена к выделению терминов, выражающих ключевые для данного текста понятия, установлению отношений между ними и формированию на этой основе некоторой схемы в виде, например, семантической сети, графа или кластерной структуры. Получаемая в результате схема рассматривается в качестве модели содержания текста. Такой подход можно видеть в ряде работ:

- «В текстовых массивах, где форма документа и набор лексики ограничены, новую информацию можно извлекать, анализируя статистику на уровне отдельных ключевых слов (терминов). ... Семантическая сеть представляет собой наиболее важные или ключевые понятия, извлеченные из текста, с выражением взаимосвязей или отношений между ними, оцененных на основе их относительной значимости» [4].

- «Автоматическое извлечение знаний из научных текстов предполагает не только выявление терминов, но и извлечение знаний о терминах. Для этого необходимо распознать в тексте семантические отношения между терминами» [20].

- «В состав формализованного смыслового описания документа должны быть включены наименования понятий, сопровождаемые коэффициентом, определяющим степень их смысловой значимости в тексте» [13].

Однако согласно другой точке зрения, единицами содержания текста выступают не знаки – слова естественного или специально создаваемого языка, а фрагменты реальной действительности, выраженные в тексте соответствующими языковыми средствами (в частности, терминами) [23]. В этом смысле, пишет О.Л. Каменская, текст представляет собой некую «инструкцию» для читателя, которая включает в себя сведения двух видов: «первый – какие понятия (о предметах или отношениях) реципиент должен отобрать из имеющегося в его распоряжении «запаса», с тем, чтобы использовать для модели, воссоздаваемой в его сознании в соответствии с инструкцией – текстом отправителя; второй – с какими именно компонентами уже ранее построенного фрагмента модели реципиент должен соотнести... вновь «отобранные» им из своей памяти понятия о предметах и отношениях» [16]. При таком

подходе содержание текста не может быть напрямую выведено из используемых в нем языковых единиц (терминов) и устанавливаемых между ними в данном же тексте отношений. Для адекватного отражения содержания необходимо учитывать значения терминов, попадающих в позицию узлов семантической сети или вершин графа, поскольку описываемая в тексте предметная реальность отражается не в именах знаков (лексических единицах), а именно в их значениях. Функцию определения значения и выполняет дефиниция.

2. Терминологическая неопределенность

Представляется очевидным, что вычленение и исследование предмета научного познания, описание и концептуализация полученных о нем знаний специфичны для различных научных дисциплин, задач и ситуаций познания. В результате один и тот же предмет может быть представлен в конкретной дисциплине, концепции или публикации по-разному. В частности, по-разному будут выделяться существенные, с точки зрения создателя конкретного текста, свойства предмета, поскольку они становятся таковыми при решении вполне конкретных познавательных задач.

Даже словари одной предметной области могут давать термину определения, акценты которых соответствуют узкой тематической специализации издания. Так, в значении термина *нейрон* «Медицинская энциклопедия» в качестве первого компонента дефиниции приводит наиболее общие характеристики: *Нейрон – это высокоспециализированная клетка нервной системы, являющаяся ее главной структурной и функциональной единицей* (Медицинская энциклопедия // http://www.erom.ru/encik/encik_n/neiron.html). В словаре «Физиология человека» выделяется функциональный аспект: *Нейроны – специализированные клетки, способные принимать, обрабатывать, кодировать, передавать и хранить информацию, организовывать реакции на раздражения, устанавливать контакты с другими нейронами, клетками органов* (Физиология человека / Под ред. В.М. Покровского, Г.Ф. Коротько. М.: Медицина, 2001). А в словаре «Гистология, цитология и эмбриология» – *нейроны состоят из тела (перикариона) и отростков, среди кото-*

рых выделяют дендриты и аксон (нейрит) (Гистология, цитология и эмбриология // <http://cytohistology.ru/>).

В текстах же первичных публикаций, ориентированных, прежде всего на изложение авторской точки зрения, значение термина зависит от специфики концептуального осмысления феномена, связанной с установками исследователя (или исследовательской школы), на которые он опирается: «Речь идет о полисемии, обусловленной расхождением в референтной области, обслуживаемой данной терминологической системой, которые целенаправленно выделяются различными исследователями» [14]. При этом «информация со стороны номинатора может быть представлена по принципу «минимального толкования», т.е. эксплицируются только основополагающие признаки, без перечисления других, поскольку они могут восполняться фоновыми знаниями человека» [15]. Следовательно, в соответствии с задачами конкретного исследования те или иные компоненты представления о предмете помещаются на первый план в качестве существенных признаков или, наоборот, за пределы текста в качестве фоновых знаний, что закономерно отражается в составе приводимых автором текста дефиниций. Так, в работах, излагающих результаты конкретных исследований, термин *нейрон* дефинируется с указанием совершенно различных признаков:

- ...нейрон, как и любая живая клетка, реализует генетическую программу, нуждаясь в метаболитах, поступающих к нему от других клеток (Александров Ю.И. Системогенез и смерть нейронов // *Нейрохимия*. 2004. Т. 21. № 1. С. 5-14).

- Нейроны обладают метаболизмом, сопоставимым с нагруженными мышечными клетками, что требует оценки происхождения нервной ткани именно с этих позиций (Александров Ю.И., Анохин К.В., Безденежных Б.Н. и др. Нейрон. Обработка сигналов. Пластичность. Моделирование / Под ред. Е.Н. Соколова, В.А. Филиппова, А.М. Черноризова. Тюмень: Издательство ТюмГУ, 2008).

- Каждый элемент среды (нейрон) является автогенератором, который в автономном режиме генерирует кратковременные импульсы (спайки) (Кащенко С.А., Майоров В.В. Модели волновой памяти. № 43. М.: URSS, 2014.).

Дополнительно усложняет ситуацию существование межпредметных заимствований, когда присвоенный из «чужой» терминологии термин в рамках принимающей системы лексических единиц приобретает принципиально отличное от исходного значение: *Легко видеть, что каждый нейрон представляет собой какое-либо конкретное понятие*, – читаем в работе по искусственному интеллекту (Пинчук А. Искусственные нейронные сети в попытке приблизиться к биологической нейронной среде // <http://www.gotai.net/documents/doc-nn-011.aspx>). Вариативно и значение терминов, обозначающих высокообобщенные понятия, которые используются представителями различных наук и входят в состав так называемой общенаучной лексики. Например, широко распространившийся в последнее время термин *фрейм*, обозначающий в самом общем случае структуру, содержащую некоторую информацию, в программировании приобретает значение «область данных», в психологии – «идеальный образ действительности, зафиксированный в памяти человека», в социологии – «определенная модель поведения», в языкознании – «парадигма», в библиографии – «форма сноски внизу страницы» и т.д. [22]. О.Д. Митрофанова показывает, как при употреблении в рамках определенной дисциплины приобретают специализированное значение общенаучные лексические единицы *масса, форма, узел, кора, сила, сопротивление* и т.п. [19].

В результате на уровне научного дискурса отмечается ситуация «дефинитивного плюрализма», когда «для одного понятия можно создать не одно, как считал Аристотель, а целый ряд отличающихся в нюансах определений, которые все же можно будет признать истинными» [29]. Невыполнение требования однозначности для термина не является экстремальным положением: «Согласно современным представлениям, одно понятие может иметь несколько различных языковых форм выражения в специальном тексте, в соответствии с теми коммуникативными задачами, которые возникают при использовании этого понятия» [5]. На уровне собственно дефиниций такой плюрализм выражается, как видно из приведенных выше примеров, в наличии у одного понятия, выражаемого термином, множества определений, которые «имеют не только различное

вербальное и структурное оформление, но и могут содержать разные наборы концептуальных признаков, объективированных на различных уровнях категоризации мира» [11].

Одновременное существование разных значений термина на уровне научного дискурса следует признать неизбежным, поскольку оно вытекает из самой природы познавательного отношения к реальности. По определению С.В. Ракитиной «Научный текст обладает мощным когнитивно-дискурсивным потенциалом: способностью углубления смысловой перспективы ранее созданной картины мира за счет приращения новых смыслов. Это становится возможным благодаря системным взаимоотношениям концептов, способных при осмыслении нового знания выстраивать в многоярусной концептосфере ученого новые схемы, новые научные концепции» [26]. Но именно эти «бесконечные творческие потенции в освоении новых знаний», отмечает Б.Ю. Городецкий, определяют принципиальную нечеткость значения как всех вообще элементов языковой системы, так и ее дефинируемых элементов, т.е. терминов [10]. Особенно ярко проявляется терминологический плюрализм в гуманитарных науках, где термин – это «явление более индивидуальное и мировоззренческое, напрямую зависящее от определенной системы взглядов, научных убеждений или собственной логики исследователя» [2]. Но даже когда речь идет о техническом знании, для терминологии которого в целях исключения разночтений законодательно закреплена необходимость стандартизации, обнаруживается пересмотр таких понятий, как «качество» или «надежность», причем изменения происходят в достаточно высоком темпе [27].

3. Выражение в текстовой дефиниции актуального значения термина

Поскольку любое научное знание, с одной стороны представляющее собой результат познавательной деятельности, с другой является лишь ее моментом [17], процесс преобразования значений единиц, служащих средством выражения научного знания, имеет перманентный характер. Однако полисемия терминов, неизбежная для уровня научного дискурса, на уровне конкретной научной публикации существенно ограничивается: «при описании конкретной предметной ситуации, конкретной

научно-технической сферы термины не допускают неверного их толкования» [32]. Наиболее эффективным средством избавления от терминологической неопределенности выступает дефиниция, сопровождающая термин в тексте. Можно сказать, что дефиниция выполняет функцию локального устранения дефинитивного плюрализма и реализации в рамках конкретной публикации требования однозначности научной информации. Для задачи формализации содержания текста это означает необходимость работы, во-первых, со значениями используемых в научной публикации терминов (а не только с самими терминами), а во-вторых, с теми дефинициями терминов, которые приводятся в конкретной публикации (а не теми, которые занесены в терминологические словари). Нужно отметить, что дефиниция является в данном случае и средством, наиболее технически удобным с точки зрения задач интеллектуального анализа текста, поскольку при всей вариативности ее текстовых форм возможно выделение определенного набора шаблонов, позволяющих решать задачу извлечения дефиниций достаточно уверенно (например, такая функция реализована в системе Exactus Expert [25]. Средства автоматического извлечения дефиниций описаны также в работах [6, 35-43].

Заключение

Проведенный анализ позволяет рассматривать дефиницию в качестве строевого компонента содержания научной публикации, а построение модели текста на основе содержащейся в нем дефинитивной информации – как шаг к более адекватному отражению его содержания по сравнению с моделями, учитывающими только имена знаков-терминов и отношения между этими именами.

Литература

1. Аллаhverдян А.Г., Мошкова Г.Ю., Юревич А.В., Яровский М.Г. Психология науки. М.: Изд-во «Флинта», 1998.
2. Анисимова А.Г. Методология перевода англоязычных терминов гуманитарных и общественно-политических наук. Автореф. дисс. докт. филол. наук. М., 2010.
3. Апресян Ю.Д. Идеи и методы современной структурной лингвистики (краткий очерк). М.: Просвещение, 1966.
4. Беленький А. Извлечение информации из неструктурированных текстов // КомпьютерПресс, 2008. № 2. С. 74-79.

5. Большакова Е.И., Васильева Н.Э. Терминологическая вариантность и ее учет при автоматической обработке текстов // Одиннадцатая Национальная конференция по искусственному интеллекту с международным участием КИИ-2008. Труды конференции в 3-х томах. М.: Физматлит, 2008. Т. 2. С. 174-182.
6. Большакова Е.И., Васильева Н.Э., Морозов С.С. Лексико-синтаксические шаблоны для автоматического анализа научно-технических текстов // Труды Десятой национ. конф. по искусствен. интел. с междунар. участ. КИИ-2006. М.: Физматлит, 2006. Т. 2. С. 506-524.
7. Васильева Н.Э. Шаблоны употреблений терминов и их использование при автоматической обработке научно-технических текстов // Компьютерная лингвистика и интеллектуальные технологии: Труды Международной конференции Диалог '2004 / Под ред. И.М. Кобозевой, А.С. Нариньяни, В.П. Селегея. М.: Наука, 2004.
8. Гак В. Г. Пространство мысли: (Опыт систематизации слов ментального поля) // Логический анализ языка: Ментальные действия. М.: Наука, 1993.
9. Гибатова Г.Ф. Мир мысли в русских предикатах // Известия РГПУ им. А.И. Герцена. 2010. № 124 // <http://cyberleninka.ru/article/n/mir-mysli-v-russkih-predikatah>
10. Городецкий Б.Ю. Компьютерная лингвистика: моделирование языкового общения // Новое в зарубежной лингвистике: Вып. XXIV. Компьютерная лингвистика: Пер. с англ. / Сост., ред. и вступ. статья Б.Ю. Городецкого. М.: Прогресс, 1989. С. 36-51.
11. Дроздова Т.В. Научный текст и проблемы его понимания (на материале англоязычных экономических текстов). Автореф. дисс. докт. филол. наук. М., 2003.
12. Зарва А.М. Дефиниция как типологическая разновидность научного текста. Автореф. дисс. канд. филол. наук. Нальчик, 2003.
13. Захаров В.Н., Хорошилов А.А. Автоматическая оценка подобия тематического содержания текстов на основе сравнения их формализованных смысловых описаний // Электронные библиотеки: перспективные методы и технологии, электронные коллекции: Труды XIV Всероссийской научной конференции RCDL'2012 (Переславль-Залесский, 15-18 октября 2012). Переславль-Залесский: Университет города Переславля, 2012. С. 189-195.
14. Зимовая М.В. Многозначность в терминологии. Автореф. дисс. канд. филол. наук. Орел, 2011.
15. Извольская И.В. Текстобразующая роль термина // Вопросы филологических наук. 2008. № 6. С. 56-58.
16. Каменская О.Л. Текст и коммуникация. М.: Высшая школа, 1990.
17. Котюрова М.П., Баженова Е.А. Культура научной речи: текст и его редактирование. М.: Флинта: Наука, 2008.
18. Кузнецова Ю.М., Осипов Г.С., Чудова Н.В. Изучение положения дел в науке с помощью методов интеллектуального анализа текстов // Управление большими системами: Сб. трудов. 2013. № 44. С. 106-138.
19. Митрофанова О.Д. Научный стиль речи: проблемы обучения. М.: Русский язык, 1985.
20. Найханова Л.В. Основные типы семантических отношений между терминами предметной области // Известия высших учебных заведений. Поволжский регион. Технические науки. Информатика, вычислительная техника и управление. 2008. № 1. С. 62-71.
21. Найханова Л.В. Технология создания методов автоматического построения онтологий с применением генетического и автоматного программирования. Улан-Удэ: Изд-во БНЦ СО РАН, 2008.
22. Некрасов С.И., Молчанова Н.С. Значение теории фреймов в современной науке // Научные ведомости. 2009. № 16 (71). С. 13-17.
23. Новиков А.И. Семантика текста и ее формализация. М.: Наука, 1983.
24. Нотина Е.А. Межъязыковая опосредованная научная коммуникация. М.: Изд-во РУДН, 2005.
25. Осипов Г.С., Смирнов И.В., Тихомиров И.А., Соченков И.В., Швец А.В. Сервисы поддержки научно-образовательной деятельности в системе Exactus Expert // <http://www.gpntb.ru/win/inter-events/crimea2014/disk/032.pdf>
26. Ракитина С.В. Когнитивно-дискурсивное пространство научного текста. Автореф. дисс. докт. филол. наук. Волгоград, 2007.
27. Стуль Я.Е., Суханов К.Н. Философские вопросы технического знания. М.: Наука, 1984.
28. Суперанская А.В., Подольская Н.В., Васильева Н.В. Общая терминология: Вопросы теории. М.: Наука. 1989.
29. Трусов В.Е. Своеобразие дефиниции в различных функциональных стилях как разновидностях типов мышления (на материале научного и художественного стилей). Автореф. дисс. канд. филол. наук. Саратов, 2008.
30. Хлодионова С.И. Культура научной речи: смысловая структура научного текста // http://journal.kfrgteu.ru/files/1/2012_6_21.pdf
31. Цымбал Н.А. Проблемы изучения научной терминологии в современной лингвистике // <http://dspace.udpu.org.ua:8080/jspui/handle/6789/1552>
32. Челнокова Ю.А., Шабунько Л.Г. Научно-технические и медицинские термины: проблемы перевода // Патенты и лицензии. 2007. № 4. С. 50-52.
33. Шатуновский И. Б. Эпистемические глаголы: коммуникативная перспектива, презумпции, прагматика // Логический анализ языка: знание и мнение. М.: Наука, 1988. С. 18-22.
34. Шилова Е.В. Терминологическая дефиниция как метатекст в русскоязычной и англоязычной научно-технической литературе. Автореф. дисс. канд. филол. наук. Екатеринбург, 2005.
35. Anke L.E. Towards Definition Extraction Using Conditional Random Fields // Proceedings of the Student Research Workshop associated with RANLP 2013, Hissar, Bulgaria, September 9-11, 2013: 63-70 // <http://www.aclweb.org/anthology/R13-2010>
36. Borg C., Rosner M., Pace G.J. Automatic Grammar Rule Extraction and Ranking for Definitions // LREC, 2010/5 // <http://staff.um.edu.mt/cbor7/publications/2010LREC.pdf>
37. Del Gaudio R., Branco A. Language Independent System for Definition Extraction: First Results Using Learning Algorithms // Sierra G. Pozzi M., Torres J.-M. (Eds.) Proceedings of International Workshop on Definition Extraction. RANLP-2009, Borovets, Bulgaria, September 14-16, 2009:33-40 // <http://anthology.aclweb.org/W/W09/W09-44.pdf>
38. Iftene A., Trandabăt D., Pistol I. Grammar-based Automatic Extraction of Definitions. Applications for Romanian // In proceedings of 10th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2008), Published by the IEEE Computer Society. Timișoara, România, September 26-29, 2008: С. 112-117 // http://scholar.google.com/citations?view_op

- =view_citation&hl=ru&user=p2ScknsAAAAJ&citation_for_view=p2ScknsAAAAJ:UeHWp8X0CEIC
39. Sierra G., Pozzi M., Torres J.-M. (Eds.) Proceedings Of International Workshop On Definition Extraction. RANLP-2009, Borovets, Bulgaria, 14-16 September, 2009 // <http://anthology.aclweb.org/W/W09/W09-44.pdf>
40. Storrer A., Wellinghoff S. Automated detection and annotation of term definitions in German text corpora // Proceedings of LREC, Genoa, 2006: 275-295 // http://www.studiger.tu-dortmund.de/images/Storrer_2006_Automated_detection_and_annotation_of_term_definitions_in_German_text_corpora.pdf
41. Walter S. Linguistic Description and Automatic Extraction of Definitions from German Court Decisions // http://www.lrec-conf.org/proceedings/lrec2008/pdf/742_paper.pdf
42. Westerhout E.N. Definition extraction for glossary creation. A study on extracting definitions for semi-automatic glossary creation in Dutch. PhD Dissertation. Utrecht University, Utrecht: LOT, 2010.
43. Yiping J. Mining Scientific Terms and their Definitions: A Study of the ACL Anthology // Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP). Seattle, USA // http://wing.comp.nus.edu.sg/publications/2013/emnlp2013_yp.pdf

Кузнецова Юлия Михайловна. Старший научный сотрудник ИСА РАН. Окончила МГУ им. М.В. Ломоносова в 1991 году. Кандидат психологических наук. Автор 78 печатных работ, в том числе двух монографий. Область научных интересов: психосемантика, психолингвистика, психология когнитивных процессов. E-mail: kuzjum@yandex.ru