

УДК 81'42

К ПРОБЛЕМЕ ИНТЕРПРЕТАЦИИ СТРУКТУРЫ ТЕКСТА

© 2010 г.

С.Н. Виноградов

Нижегородский госуниверситет им. Н.И. Лобачевского

vinogradov54@mail.ru

Поступила в редакцию 06.07.2010

Рассматривается текст как подобная слову семиотическая единица, структура которой может быть различным образом интерпретирована. Множество наиболее частотных лексических единиц, выделенных из научного текста, рассматривается как его структурная часть, выражающая его основное значение. Обсуждаются некоторые аспекты интерпретации текста с помощью ключевых слов.

Ключевые слова: слово, морфемная структура, текст, интерпретация текста, структура текста, теоретико-множественная модель, ключевые слова.

Под структурой обычно понимается внутреннее устройство чего-либо, а внутреннее устройство связано с категориями целого и его частей. В лингвистике принято говорить о структуре слова, морфемы, основы, синтагмы, предложения, текста и т. п., но о системе гласных, форм одного слова, падежей, фонем, значений многозначного слова и т. п. [1, с. 453]. Структурой обладают различные единицы языка и других семиотических систем. На такое понимание структуры мы и будем опираться. В данной статье в основном будет анализироваться структура текста, который можно рассматривать как семиотическую единицу и который, как и другие языковые единицы, имеет внутреннее устройство, внутреннее строение.

Со структурой текста мы имеем дело тогда, когда текст рассматривается как единица некоторой семиотической системы и требуется выявить специфику данной единицы. Ранее мы проводили семиотические и структурно-семантические параллели между словом и текстом [2], которые во многом сходны благодаря своей знаковой природе, вовлечённости в коммуникативную деятельность как её системообразующие и системно организованные фрагменты. Методики лингвистического изучения слова и текста также до известной степени аналогичны: «В любом случае мы ищем структурные элементы, выражающие определённое содержание, и выявляем закономерности системного взаимодействия этих элементов, правила их употребления и интерпретации» [2, с. 269–270].

Будучи знаковыми единицами, и слово, и текст интерпретируются [3, 4], причём интерпретация приводит к созданию материального

результата [4, с. 19]. Например, результатом интерпретации слова *объёмистый* в процессе морфемного анализа будет последовательность *объ-ём-ист-ый* (побуквенное представление морфем, разделённых дефисом; в этой записи важен также порядок морфем, так как их место в слове связано с их классификацией).

В зависимости от характера и целей деятельности результат интерпретации может быть другим, например последовательностью терминов «приставка – корень – суффикс – окончание» или последовательностью надстрочных обозначений, принятых в практике школьного обучения при выделении значимых частей слова. Из этих примеров видно, что результаты интерпретации при осуществлении даже более или менее одинаковой деятельности могут быть различными. Это обусловлено частными целями и задачами деятельности. Тем более будут отличаться результаты интерпретации в иных видах деятельности, например при толковании значения слова или при написании творческой работы, трактующей выражаемое словом понятие.

Те же соображения применимы и к тексту. Результаты его интерпретации могут быть различными, и это хорошо известно всем, кто работает с текстами в разных сферах человеческой деятельности. По-разному могут интерпретироваться не только художественные, но и научные, и юридические тексты. Однозначность истолкования – это идеал, который если и достигим, то очень редко.

Здесь может возникнуть вопрос: достигим ли этот идеал при интерпретации слова? На первый взгляд, кажется, что при выделении морфемы в слове можно добиться практически

абсолютной объективности и однозначности. В результате многолетнего изучения производного слова в русском языке сложилась (с рядом упрощений и огрублений) следующая модель материальной структуры слова: оно содержит ряд морфем, одни из которых выражают его основное лексическое значение (корневые морфемы), другие модифицируют это значение, образуя новые слова (словообразовательные морфемы), третьи образуют форму данного производного слова (грамматические морфемы). При этом учитывается линейная последовательность этих морфем: те, которые находятся перед корнем, называются префиксами, после корня – суффиксами, после флексий – постфиксами.

Стройность этой модели разрушается под влиянием ряда факторов, которые в конечном итоге приводят к множественности результатов интерпретаций. Во-первых, возможно разное понимание сущности морфемы и её разновидностей и в связи с этим различные операционные определения соответствующих понятий. Так, определение корня как морфемы, выражающей основное лексическое значение слова [5, с. 58], не позволяет выделять отрезок *бужен* как корневую морфему в слове *буженина*, поскольку этот отрезок не выражает лексического значения (в отличие от *кон'* в *конина*, *баран* в *баранина* и т.д.). С этой точки зрения морфемой (значимой частью) данного слова является *буженин-*. Если же рассматривать корень как морфему, выражающую не обязательно лексическое значение, но также и значение структурное, то отрезок *бужен* будет выделен как морфема на том основании, что он отличает значение слова *буженина* от значений слов *конина*, *баранина* и т.д. С.И. Богданов считает морфему по преимуществу средством структурной организации единицы более высокого порядка [6, с. 82] и по этой причине при выделении морфем считает существенным не их значение, а выражаемую ими общесистемную структурную аналогию (например, членение слова *гон-и-тель-о* по аналогии с *победитель*, *родитель* и др.) [6, с. 48–55]. Подобные различия в определениях и других исходных посылках анализа приводят к тому, что границы морфем в слове могут быть проведены по-разному и, следовательно, результаты интерпретации морфемной структуры слова могут быть различными. Таким образом, объективные основы множественности интерпретаций коренятся в множественности видов деятельности, в тех предпочтениях, которые в ней проявляются.

Однако наличие морфем как элементарных значимых единиц в русском слове не вызывает

сомнения. Нельзя ли выделить аналогичные единицы и в тексте? Насколько далеко простираются аналогии между словом и текстом в области структуры?

Чисто внешние структурные отличия текста от слова проявляются в том, что в тексте больше структурных элементов, к которым относятся употребления слов, предложений и других единиц языка, и линейная организация знака играет меньшую роль, чем в слове (хотя отчасти эта роль сохраняется). Более глубокие отличия текста от слова заключаются в том, что структурой текста управляют иные внешние причины. Согласно определению Р. Барта, связный текст (дискурс) – любой конечный отрезок речи, представляющий собой некоторое единство с точки зрения содержания, передаваемый с вторичными коммуникативными целями и имеющий соответствующую этим целям внутреннюю организацию, причём связанный с иными культурными факторами, нежели те, которые относятся собственно к языку [7, с. 443–444]. Обратим внимание на специфику текста как знаковой единицы. Вторичные коммуникативные цели отличают текст от языка, на котором он написан: языковая единица используется для кодирования информации в любом сообщении, а текст предназначен для выражения информации научной, бытовой, эстетической, управленческой и др. в зависимости от сферы его создания и употребления. Эта сфера и есть совокупность тех факторов, которые управляют внутренней организацией текста.

Сложность человеческой деятельности приводит к разнообразию функций языка, реализованных в текстах различных стилей и жанров. Так, в научном стиле большое значение приобретает когнитивная функция (это не означает, что другие функции языка здесь полностью исключены), в художественном – эстетическая функция и т.д. По-видимому, семиотическая «система систем», образуемая текстами, гораздо сложнее, чем собственно язык. Очевидно, при всём структурно-языковом сходстве, например, художественных и научных текстов структурные элементы, выделяемые в них как в знаках особых семиотических систем, будут существенно различаться.

В силу вышесказанного ограничимся рассмотрением текстов определённой сферы, главным образом научной. Научный текст связан с когнитивной функцией языка и с логическими аспектами деятельности, так как трактует научные понятия, выражает их признаки, содержит суждения и умозаключения. Будучи знаковым образованием, научный текст подвергается раз-

личным интерпретациям. Одной из важных разновидностей интерпретаций научного или технического текста является его компрессия (сжатие). Название текста (заголовок), формулировка темы, реферат, аннотация, набор ключевых слов – всё это разновидности результатов такой компрессии.

Если мы поставили целью научное исследование структуры научного текста, надо найти объективные основания выделения элементов этой структуры. Таким объективным основанием является деятельность интерпретатора (в случае компрессии текста – деятельность библиографа, индексатора, автора, формулирующего название своей работы). В этой деятельности сплетены психофизиологические, лингвистические, логические аспекты. Как, например, индексатор (составитель поискового образа документа в информационно-поисковой системе) отбирает ключевые слова из индексируемого документа? Есть данные, что на процесс отбора индексатором ключевых слов из текста наибольшее влияние оказывают следующие факторы: повторяемость слов в тексте, характер распределения повторяющихся слов по различным предложениям и фрагментам текста, структура и лексическое наполнение начальных и конечных фраз текста, морфологические и синтаксические показатели связи слов и предложений в тексте [8, с. 23]. По-видимому, выделенные здесь свойства (повторяемость и комбинирование текстовых элементов, линейная организация, языковая структура словосочетаний и предложений) являются релевантными свойствами текстовой структуры, которые учитываются в процессе интерпретации текста.

Повторы текстовых элементов использовались при создании формализованных моделей смыслового анализа текста для прикладных целей [4, с. 146–147]. Различные исследователи обратили внимание на следующую закономерность: чем чаще встречается в тексте некоторое слово, тем вероятнее, что оно передаёт основное содержание данного текста. Эта закономерность позволяет выработать некоторую теоретико-множественную модель структуры текста с точки зрения разделения лексических единиц, существенных и несущественных для выражения его основного содержания. Текст можно представить как множество всех словоупотреблений знаменательных частей речи, на котором заданы отношения эквивалентности. Эквивалентными следует считать словоупотребления, в которых основы слов передаются одинаковыми буквенными последовательностями (например, *морфема*, *морфемы*, *морфеме* и т.д.). Чем

больше элементов связаны таким отношением (то есть чем больше мощность множества, содержащего эквивалентные единицы), тем вероятнее, что соответствующая словоформа существенна для выражения основного содержания текста. Если проводить аналогию между анализом структуры слова и анализом структуры текста, то данная теоретико-множественная модель аналогична отделению корня – морфемы, несущей основное лексическое значение слова, – от служебных морфем, выполняющих вспомогательную функцию, модифицирующих основное лексическое значение.

Описанная модель структуры текста в виде системы множеств словоупотреблений отличается от описания структуры слова в виде последовательности морфем. Данная теоретико-множественная модель не отражает многих аспектов структуры текста. Из всей сложной структуры текста для создания модели взята лишь частота повторов текстовых единиц. Очевидно, есть и другие структурные показатели, которые могут быть проанализированы, объективно описаны и оказаться полезными для изучения тех или иных содержательных аспектов текста. В частности, можно рассматривать не только повторы отдельных слов, но и словосочетаний, учитывая синтаксические структуры, реализованные в научном тексте. В этом случае нужен более сложный алгоритм генерации элементов множества: в состав этих элементов следует включать не только отдельные словоформы (последовательности букв «от пробела до пробела»), но и словосочетания. Повторяющиеся сочетания слов важнее для передачи основного содержания, чем повторяющиеся отдельные слова, так как более детально выражают признаки понятий. Нами была разработана и опробована методика обнаружения лексических единиц, предположительно передающих основное содержание текста, на основе выделения повторяющихся слов и словосочетаний [4, с. 137–184; 9; 10]. В теоретико-множественной модели в явном виде не использована линейная организация текста, но в некоторых исследованиях по автоматизированному аннотированию текста предлагается использовать элементы его линейной организации. Например, «чем больше расстояние по тексту... между первым и последним употреблением термина, тем больше вероятность того, что термин обозначает тему текста» [11, с. 17].

Использование различных структурных критериев для выделения лексики, существенной для данного текста, позволяет выделить её точнее и с меньшим информационным шумом. Но не следует стремиться к точности любой ценой.

Дело в том, что психофизиологическая реальность предполагает нечёткость выбора, которая должна отражаться и в выборе лексических единиц. Множество таких единиц носит нечёткий, размытый характер, что проявляется в известной нечёткости границ искомым множеств эквивалентности. Например, в отрывке из академической грамматики, посвящённом понятию морфемы и содержащем определение морфемы, с использованием формальных критериев (на основе частоты лексической единицы и количества знаменательных слов, входящих в неё) были выделены следующие «самые важные» лексические единицы: *морф*, *морф* *корневой*, *морфема*, *морфема* *корневая*, *морфы* *одноименные*, *словоформа*, *совокупность* [4, с. 180]. Эти лексические единицы отражают содержание данного отрывка, в котором речь идёт о том, что морфема представляет собой совокупность морфов, выделяемых в различных словоформах, и что эти морфы представлены различными разновидностями (корнями, суффиксами и т. д.). Но лексическая единица *корень* оказалась «хуже» перечисленных, попала в более низкий ранг с точки зрения её формальных признаков, хотя имеет практически то же значение, что и словосочетание *морфема* *корневая*. С другой стороны, в число выделенных единиц попало слово *совокупность*, являющееся не узкоотраслевым термином, а общенаучным словом (хотя его появление здесь можно объяснить его терминологизацией, уподоблением по значению математическому термину *множество*). Вероятно, при изменении формальных критериев (или при выполнении этой работы «вручную» разными людьми) состав выделенного множества будет меняться в некоторых пределах. Выше было показано, что даже при анализе морфемной структуры слова состав и границы морфем могут быть разными в зависимости от исходных посылок. Тем более это справедливо при выделении единиц такого многоэлементного образования, как научный текст.

Задача выделения лексики, выражающей основное содержание, может быть поставлена именно для текста, реализующего когнитивную функцию языка, связанную с логическими аспектами деятельности. Данная функция реализуется не только в научном стиле. Например, стихотворение Н. Гумилёва «Сахара», будучи отрывком из своеобразного поэтического путевого дневника, точно передаёт понятия о некоторых описываемых реалиях [9, с. 264], что отражено в составе многих повторяющихся в этом стихотворении слов: *верблюды*, *песок*, *Сахара*, *пустыня*, *солнце* (всё это либо наименования

пустыни, либо её атрибуты – номинации «пустынной» темы). Но, разумеется, эта роль словесных повторов характерна не для всех поэтических текстов. Правомерно ли задавать вопрос об основном содержании поэтического текста? Скорее всего, содержание этого вопроса и ответа на него будет различным для текстов, реализующих разные языковые функции – эстетическую и когнитивную. Исходя из представлений о поэтическом тексте в работах Р.О. Якобсона и Ю.М. Лотмана, в поэзии важны не столько повторяющиеся лексические единицы, сколько лексические единицы, образующие парадигматические отношения, возникшие в результате осуществления поэтической функции и реализуемые в пределах поэтического произведения. Нахождению таких единиц помогает структура поэтического произведения, в частности, формальный параллелизм стихотворных строк, их рифмованных окончаний, различных композиционных элементов. Подобная организация текстов совершенно нехарактерна для других сфер употребления языка.

Таким образом, тексты, имеющие различные вторичные коммуникативные цели, имеют различную внутреннюю организацию и связаны с различными задачами интерпретационной деятельности. Это различие имеет отдалённый аналог в разнообразии интерпретаций слов, состоящих более чем из одной морфемы. Результатами этих интерпретаций могут быть деление слова на морфемы, выделение производящей базы и словообразующего форманта, формулировка толкования лексического значения и другое. Благодаря сложности текстовой структуры можно ожидать, что в тексте можно выделить и другие структурные элементы, релевантные различным элементам его содержания.

Список литературы

1. Лингвистический энциклопедический словарь / Гл. ред. В.Н. Ярцева. М.: Сов. энциклопедия, 1990. 685 с.: ил.
2. Виноградов С.Н. Семиотические и структурно-семантические параллели между словом и текстом // Вестник ННГУ. 2009. № 4. С. 267–270.
3. Моррис Ч. Основания теории знаков // Семиотика. М.: Радуга, 1983. С. 37–89.
4. Виноградов С.Н. Термин как средство и объект описания (на материале русской лингвистической терминологии): Монография. Нижний Новгород: Изд-во ННГУ, 2005. 229 с.
5. Немченко В.Н. Современный русский язык: Морфемика и словообразование: Учебник. Нижний Новгород: Изд-во ННГУ, 1994. 296 с.
6. Богданов С.И. Форма слова и морфологическая форма. СПб.: Изд-во СПбУ, 1997. 248 с.

7. Барт Р. Лингвистика текста // Новое в зарубежной лингвистике. Вып. 8. Лингвистика текста. М.: Прогресс, 1978. С. 442–449.

8. Бондаренко Г.В., Яровенко О.И. Использование структурных закономерностей текста при автоматической обработке информации // Научно-техническая информация. Сер. 2. 1984. № 3. С. 23–29.

9. Виноградов С.Н. Лексические повторы и понятийное содержание поэтического текста // Семантика. Функционирование. Текст: Межвузовский сборник научных трудов. Киров: Изд-во ВятГУ, 2006. С. 260–265.

10. Виноградов С.Н. Некоторые особенности лексических повторов в текстах сходной тематики // Активные процессы в современном русском языке: Сборник научных трудов, посвящённый 80-летию со дня рождения профессора В.Н. Немченко. Нижний Новгород: Издатель Ю.А. Николаев, 2008. С. 39–44.

11. Новиков А.И., Якушин Б.В. Алгоритм индексирования текстов взвешенными ключевыми словами по методу семантической фильтрации // Научно-техническая информация. Сер. 2. 1972. № 6. С. 15–20.

ON THE PROBLEM OF INTERPRETATION OF THE TEXT STRUCTURE

S.N. Vinogradov

The author considers the text as a word-like semiotic unit, the structure of which may be interpreted in different ways. The set of most frequent lexical units picked out from scientific text is considered as its structural part expressing its main meaning. Some aspects of text interpretation by means of keywords are discussed.

Keywords: word, morphemic structure, text, interpretation of text, text structure, set theory model, keywords.