

ЛЕНИНГРАДСКИЙ ОРДЕНА ЛЕНИНА  
И ОРДЕНА ТРУДОВОГО КРАСНОГО ЗНАМЕНИ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ имени А. А. ЖДАНОВА

ЛИНГВИСТИЧЕСКИЕ ПРОБЛЕМЫ  
ФУНКЦИОНАЛЬНОГО  
МОДЕЛИРОВАНИЯ  
РЕЧЕВОЙ ДЕЯТЕЛЬНОСТИ

ВЫПУСК I

**756712**

ИЗДАТЕЛЬСТВО  
ЛЕНИНГРАДСКОГО УНИВЕРСИТЕТА  
1973

*Печатается по постановлению  
Редакционно-издательского совета  
Ленинградского университета*

Статьи сборника представляют самостоятельные исследования либо конкретных лингвистических проблем точными методами, либо самих методов описания языка. Значительная часть работ посвящена разработке отдельных вопросов, связанных с созданием информационной системы «запрос—ответ», принципиально отличающейся от существующих информационных систем. В ряде исследований развивается проблематика, связанная с автоматическим переводом, автоматическим синтаксическим анализом, переводом с естественного языка на искусственные языки, описываются и анализируются результаты экспериментальной проверки разработанных в лаборатории систем автоматической обработки текста.

Сборник рассчитан на читателей, интересующихся проблемами математической лингвистики.

Ответственный редактор Г. С. Цейтин

## ПРЕДИСЛОВИЕ

Работы, входящие в данный сборник, посвящены функциональному моделированию различных аспектов речевой деятельности. Наиболее интересным здесь представляются те аспекты, которые затрагивают не только форму, но и смысл текстов: имеется в виду моделирование способности человека выполнять эквивалентные смысловые преобразования над текстом, получать из него следствия, отвечать на вопросы по этому тексту. Наиболее общей постановкой задачи для такого моделирования является построение информационной системы типа «запрос—ответ».<sup>1</sup> Значительная часть работ в этом сборнике касается частных проблем, связанных с этой общей задачей.

Системы типа «запрос—ответ» в отличие от информационно-поисковых систем должны на основе имеющейся информации обеспечивать выдачу в ответ на запрос не только информации, которая содержится в памяти системы в явном виде, но и информации, которая логически следует из имеющейся. Предполагается, что информация на входе и выходе системы представлена в виде текстов на естественном языке. В основе действия системы лежат два типа правил преобразования текстов: правила эквивалентного преобразования и правила вывода.

Для построения таких правил необходимо предварительно решить целый ряд лингвистических проблем, связанных с переходом от поверхностной структуры текста к представлению смысла текста на некотором достаточно строгом семантиче-

<sup>1</sup> Общую постановку задачи см.: Г. М. Ильин, Б. М. Лейкина, Т. Н. Никитина, М. И. Откупщикова, С. Я. Фитиалов. Лингвистический подход к задаче построения информационной системы. В сб.: Информационные вопросы семиотики, лингвистики и автоматического перевода, вып. 2. М., 1971.

ском языке (к «внутренней» структуре текста), так как содержательные преобразования значительно проще формулировать на языке формализованном, лишенном неоднозначностей, «неполноты» и «избыточности» естественного языка, короче, всяких расхождений между формой и содержанием в языке. Предполагается, что любая единица смысла, вычлененная из текста и допускающая определенные преобразования, может быть представлена во внутренней структуре как некоторая совокупность объектов и функторов, определенным образом связанных друг с другом. Объекты и функторы — это элементарные единицы внутренней структуры, из которых с помощью операции заполнения «мест» функторов строятся более сложные единицы — сложные объекты и ситуации. Для описываемой задачи класс функторов представляется целесообразным разбить на несколько подклассов в зависимости от характера единицы, заполняющей места функтора и от характера результирующей единицы внутренней структуры.<sup>2</sup>

Переход от внешней структуры текста к внутренней и обратно состоит в установлении такого соответствия между элементами внешней и внутренней структуры, при котором правила преобразования, сформулированные в терминах внутренней структуры, задавали бы необходимые содержательные преобразования над текстами.

Одной из основных проблем этого рода является разграничение информативных элементов текста, которые должны быть отражены в его внутренней структуре, и неинформативных, служащих лишь для организации текста, которые при переходе к внутренней структуре являются лишними. Требуется также описание элементов, в явном виде в поверхностной структуре не представленных, но восстанавливаемых при восприятии речи на основе учета общей структуры высказывания, элементов, учет которых приведет к потере информации.

Чрезвычайно существенным для формулирования правил действия системы является понятие функтора. Практическая задача здесь заключается в составлении списка функторов и в их классификации. Особое внимание уделяется функторам, значение которых задается явно (полностью или частично) в виде правил преобразования, приложимых к ним. Такие функторы задают логические отношения между объектами

<sup>2</sup> Подробно о классификации функторов см. там же, стр. 7.

(например, предикаты тождества и включения) или между ситуациями (в частности, коннекторы причины, условия и т. п.). Определение этих функторов и описание способов их выражения в тексте — одна из важных лингвистических задач при решении вопроса о построении информационной системы. Выделение такого рода «грамматических» функторов означает проведение некоторой условной границы между «внутренней грамматикой» и «внутренней лексикой».

На основе выделения и описания логических функторов формулируются правила преобразования текстов. С этим связаны такие проблемы, как разбиение сложных предложений на простые, т. е. правила вывода простых предложений из сложных, правила эквивалентности частей сложного предложения определенным элементам простого предложения (причастным, деепричастным и инфинитивным оборотам, предложным конструкциям) и, наоборот, правила построения сложных структур из более простых.

Существенной операцией при переходе от внешней структуры к внутренней является учет различных межфразовых связей в тексте, в частности, необходимо отождествление различных упоминаний одного и того же объекта в связном тексте. В лингвистике эта проблема интерпретируется как проблема анафорических связей и разнообразных категорий слов-заместителей; логически она может быть интерпретирована как проблема тождества. С анализом текста связана и задача выявления кванторных характеристик объектов в зависимости от способов их обозначения в поверхностной структуре.

Кроме перечисленных выше лингвистических задач, при построении действующей информационной системы рассматриваемого типа возникают и чисто математические проблемы, в первую очередь алгоритмические. Процесс поиска ответа на поступивший запрос состоит в применении к имеющимся в системе текстам разрешенных преобразований, причем преобразования должны применяться в последовательности, которая приведет к искомому результату. Требование возможности алгоритмизации процесса поиска неизбежно вызывает определенные ограничения, накладываемые на способ внутреннего представления текстов.

Затронутые выше лингвистические задачи следует рассматривать как определенную программу работ, выполнение которой даст возможность строить информационные системы различной сложности.

Разработкой некоторых из перечисленных проблем занимаются сотрудники лаборатории математической лингвистики ВЦ ЛГУ. Часть результатов исследований в этой области представлена в данном сборнике, некоторые работы по рассматриваемым вопросам были опубликованы в сборнике «Информационные вопросы семиотики, лингвистики и автоматического перевода» (вып. 2, М., 1971).

Ряд статей этого сборника рассматривает моделирование более низких уровней речевой деятельности. Имеется в виду проблематика, порожденная задачами автоматического анализа и синтеза. В настоящее время в области такого моделирования уже накоплен значительный опыт, и поэтому модели, описывающие только синтаксический (или морфологический) уровень языка, имеют гораздо более конкретный и разработанный вид, а также более широкое экспериментальное подтверждение, чем семантические модели. Хотя опыт исследований в этой области показал, что удовлетворительное решение практической задачи автоматического анализа (и синтеза) текстов, даже только на синтаксическом уровне, немислимо в отрыве от семантики, представляется, что возможности развития чисто синтаксических моделей еще не исчерпаны. Построение моделей, хорошо описывающих функционирование языка на синтаксическом уровне, в частности выбор подходящей формы для таких моделей, является актуальной задачей как в теоретическом плане (разработка новых типов лингвистических описаний), так и для практического построения информационных систем, поскольку синтаксический анализ и синтез являются этапами обработки текста в такой системе.

В некоторых статьях синтаксический и семантический аспекты моделирования рассматриваются совместно, например, строится синтаксическая модель, использующая признаки семантического характера. В одной из статей исследование синтаксических характеристик речи употребляется как средство для выявления расстройства мышления. Последнюю работу можно рассматривать также как попытку моделирования совсем другого аспекта речевой деятельности, а именно различения речевых стилей.

В одной из статей при описании определенных этапов автоматической обработки текста приведены некоторые результаты комплексных экспериментов по системе автоматического перевода, разработанной в лаборатории математической лингвистики ВЦ ЛГУ.

## СКОБОЧНЫЕ КОНСТРУКЦИИ В СИСТЕМЕ АВТОМАТИЧЕСКОГО АНАЛИЗА

Проблема скобок является малоизученной вообще и еще менее изученной для прикладных задач автоматического анализа.

В системе АП, разработанной в ВЦ ЛГУ, скобочные конструкции пока не учитывались, хотя представляют собой интересную и сложную проблему, так как ими выделяются единицы разных уровней: слово, словосочетание, предложение<sup>1</sup>. Обычные грамматические описания рассматривают конструкции со скобками только с точки зрения их функции и поэтому не являются достаточными для целей автоматического анализа.

Данная работа представляет собой попытку описать некоторые формальные признаки выявления различных функций скобочных конструкций, встречающихся в технических текстах. Материалом исследования явились английские и американские тексты по радиолокации (2000 предложений).

В литературе отмечаются следующие функции употребления скобочных конструкций. Круглые скобки: 1) выделение вводных членов предложения и вводных вставных предложений; 2) выделение обособленных членов предложения, носящих характер второстепенного замечания; 3) выделение ссылок в научно-технической, учебной или справочной литературе; 4) выделение цифр или букв в рубриках при пе-

---

<sup>1</sup> Г. М. Ильин, Б. М. Лейкина, Т. Н. Никитина, М. И. Откупщикова, Г. С. Цейтин, С. Я. Фитиалов. Система автоматического перевода, разрабатываемая в группе математической лингвистики ВЦ ЛГУ. НТИ, 1966, № 1.

речислении. Квадратные скобки: 1) выделение пояснений, вставляемых автором в цитируемый текст; 2) выделение части текста, уже заключенного в круглые скобки<sup>2</sup>. Фигурные скобки обычно не учитываются.

В исследованных текстах встретились все разновидности скобок: круглые, квадратные и фигурные. Квадратные и фигурные скобки — только в формулах:

1. Inserting V into the screening range equation

$$R = \left[ \frac{(1 \ 10^6) (316) (1) (2.6 \ 10^{10})}{(12.6) (2) (2) (3.16 \ 10) (0.6) (1.15)} \right];$$

$$P_{im} = \left\{ \begin{array}{l} \text{the probability density of the sample } F... \text{ under the} \\ \text{hypothesis that one or more targets are present} \end{array} \right.$$

Такие случаи немногочисленны и к тому же связаны с особой проблемой анализа формул, поэтому в данной работе рассматриваться не будут.

Круглые скобки могут быть односторонними и двусторонними. Односторонние скобки однофункциональны: они используются только для выделения цифр или букв в рубриках при перечислении.

3. Solutions to these problems were obtained before the results mentioned here could be realized:

1) work was done on Supermandur, to insure that its superior magnetic properties were obtained in actual cores of the size needed;

2) heat runs were made to ascertain the need for external cooling.

Двусторонние скобки тоже могут выступать в этой же функции:

4. The following signals may be switch selected:

(1) Moving or stationary target pulse with range variable from 7 to 200 nautical miles, pulse width variable from 0,2 to 10 sec. and amplitude variable from 0 to 3,5 volts (peak).

(2) ...

<sup>2</sup> Н. А. Кобрин, Л. В. Малаховский. Английская пунктуация. М., 1961.



5. Manual controls are provided for

(a) Rate of turn: Calibrated Fort and Starboard 0 to 90°/minute or any other desired rate.

(b) ...

Основным признаком скобок, выполняющих функцию рубрикации, является то, что они выделяют не содержащиеся в словаре словоформы (цифры или буквы) и выделяются сами из основного текста либо графически (после скобок слово начинается с заглавной буквы, хотя точка отсутствует), либо типографским способом (скобки стоят в начале предложения левее основного текста предложения или каждого из однородных членов). Дополнительным формальным признаком в большинстве случаев является двоеточие в конце предшествующего предложения (см. примеры 3 и 4).

Двусторонние скобки, выделяющие цифру или букву, встречаются и в конце предложения (чаще всего после формулы). В этом случае они будут выполнять функцию идентифицирующего приложения:

6. The equation for the self-screening range,  $R \dots$  is

$$R = (P_T J_T / 4 P_I J_I V). \quad (1)$$

В этом случае скобки также несколько выделяются из текста предложения, занимая крайнюю правую позицию, отделенную от предшествующего текста большим интервалом, чем обычный пробел между словами.

Особую трудность представляют собой конструкции, заключенные в круглые двусторонние скобки, так как они могут быть единицами разных синтаксических уровней: словом, словосочетанием, предложением. В тех случаях, когда скобочная конструкция представляет собой предложение, его целесообразно вычленить в качестве самостоятельного сегмента.

В единичных случаях такая сегментация задана в самом тексте; скобочная конструкция с обеих сторон выделяется точкой и пробелом; слово, следующее за левой скобкой, имеет заглавную букву.

7. The new system also has a moving target indicator (MTI) with only one speed blind spot. (Previous MTI<sub>3</sub> for ARSR radar had several speed blind spots between zero and 600 knots.)

Иногда скобки выделяют даже два предложения.

8. (If greater Doppler resolution is desired, the beam can be held as long as necessary to provide the desired resolution in only those beam positions in which the presence of a target has been detected.

This would not increase the average dwell time significantly in a low target density environment).

Эти случаи не представляют особых трудностей. В грамматике синтаксического анализа следует лишь учесть такую возможность. В остальных случаях скобочную конструкцию следует подвергнуть анализу для выявления того, не будет ли она предложением.

Для того чтобы определить, является ли конструкция, заключенная в скобки, предложением, необходимо найти его показатели. Достаточными признаками можно считать однозначно финитную глагольную форму и беспредложную именную группу слева от нее:

9. Figure 3C shows how the four horn feed of the sequentially-lobed radar is located with its aperture in the focal plane (the phase center of the feed horn is placed in the focal plane)...

Количество таких случаев невелико (всего 8).

Ссылочные конструкции к иллюстративному материалу, также имеют статус предложения, устанавливаемый на основании иного формального признака — повелительной формы глагола to see непосредственно после левой скобки (в сочетании с позиционным признаком форма see является достаточным показателем предложения, хотя вне контекста она представляет собой неоднозначно финитную форму):

10. ...two transistorized video stages make up the *i-f* and video amplifiers (see Fig. 3).

Другие глаголы в аналогичной функции в рассмотренных предложениях не встречаются.

Вместо figure и Fig. в подобных конструкциях наблюдаются также существительные: photo, discussion, map.

Не все скобочные конструкции, как уже отмечалось, обладают признаками предложения: в скобки может быть заключено словосочетание и слово, которые являются единицами более низкого синтаксического уровня. В этих случаях скобки стоят непосредственно за словом или целым высказыванием, которому подчинено заключенное в скобки выражение, свя-

занное со словом или высказыванием одним типом связи— приложением. В скобках даются некоторые дополнительные сведения, иногда раскрывающие значение слова, которому они подчинены:

11. Acknowledgment is also made of numerous beneficial discussions of radar beacon problem areas with F. H. Dyke, RCA Service Co., L. F. O'Kelly, Radiation, Inc. (formerly RCA Service Co.) and R. C. Wilson, AFHTC.

12... analog deflection data inputs to the PPDD system are  $x$  and  $y$  coordinate data (resolved ppi sweeps, computer coordinate, etc.).

Встречаются причастные обороты, выделенные скобками:

13. Furthermore, development of instantaneous electronic-plotting boards (operating in rectangular coordinates) had already been carried out for use in close-support bombing and in mortar location.

Сокращения, заключенные в скобки, принимаем за слово и будем обрабатывать в «Блоке слов, не предусмотренных в словаре».<sup>3</sup>

14. Basic line-type radar modulators consist of a switch tube, load and pulse-forming network (PFN).

Таким образом, как показали предварительные наблюдения, односторонние скобки встретились только в одной функции — функции рубрикации. Особую же трудность представляют конструкции, заключенные в круглые двусторонние скобки.

При дальнейшем исследовании функции скобочных конструкций случаи их употребления могут быть расширены.

**В. Д. Буторов**

### **ОТОЖДЕСТВЛЕНИЕ И СВЯЗЫВАНИЕ ПЕРЕМЕННЫХ ПРИ СЕМАНТИЧЕСКОМ АНАЛИЗЕ МАТЕМАТИЧЕСКИХ ТЕКСТОВ**

Под семантическим анализом предложения (текста) будем понимать «извлечение» и запись в какой-либо форме,

<sup>3</sup> Б. М. Лейкина. Независимый морфологический анализ английского научно-технического текста. Труды ЦНИИПИ, сер. 3. Автоматический анализ патентных текстов. М., 1971, стр. 12—25.

т. е. в некотором языке, содержания предложения. Саму запись содержания будем называть интерпретацией предложения.

Семантический анализ предложения естественно проводить путем анализа смысла его составляющих, начиная с отдельных слов и переходя постепенно к более крупным единицам, вплоть до получения анализа целого предложения.

1. Представление синтаксической структуры предложения. Структура составляющих предложения может быть задана некоторой порождающей синтаксической моделью, в которой процесс построения предложения описывается последовательным применением ряда операций, образующих, начиная с отдельных слов, все более крупные составляющие.

В данной работе для представления синтаксической структуры используется дерево зависимостей с некоторой дополнительной информацией. Все связи в дереве помечаются знаком типа, например в терминах отношений непосредственной доминации,<sup>1</sup> которые специально обозначаются: S — предикативное отношение, O<sub>1</sub> — первое объектное, A — определительное и т. д. Могут вводиться также новые типы связей.

Порождающей операцией в принятой синтаксической модели считается присоединение некоторого поддерева к узлу, которому оно должно быть подчинено.<sup>2</sup>

Дополнительная информация в дереве зависимостей, состоящая во введении дополнительных узлов и связей, требуется, в частности, для отражения в структуре так называемых анафорических связей — связей между словами, указывающими на один и тот же объект, упоминаемый в тексте несколько раз. Мы предполагаем, что такие связи в тексте «восстановлены».

Существенно различать первичное упоминание объекта, который обозначается при этом каким-либо существительным, и все последующие, вторичные, которые могут вводиться с помощью различных «слов-заместителей» — повторений самих существительных, называющих объект при его первичном

<sup>1</sup> И. А. Мельчук. Автоматический синтаксический анализ, т. 1. Новосибирск, 1964.

<sup>2</sup> Г. С. Цейтин, Язык математической логики как средство исследования семантики естественного языка. Проблемы прикладной лингвистики. Тезисы межвузовской конференции, 16—19 декабря 1969. М., 1969.

упоминании, местоимений, специальных обозначений и др.<sup>3</sup> Иногда вторичные упоминания объектов подразумеваются без явного выражения.

Назовем первичным вхождением существительного то его вхождение в текст, которое соответствует первичному упоминанию объекта, обозначенного данным существительным.

Анафорическая связь считается направленной от первичного вхождения существительного к словам-заместителям.

Понятие первичного вхождения существительного распространяем и на существительные, не имеющие анафорических связей. Такие существительные выражают первичное (и единственное) упоминание объекта в тексте.

**Пример 1. Во всякой конечной группе порядок любой подгруппы является делителем порядка самой группы.**

Дерево зависимостей этого предложения с учетом анафорических связей в нем (см. пунктирные стрелки) можно изобразить следующей схемой:



В скобках указан заместитель, пропущенный в данном предложении и восстановленный при построении дерева.

Строго говоря, с включением анафорических связей структура зависимостей перестает быть деревом. Можно «устранить» из структуры зависимостей пунктирные стрелки, изображающие анафорические связи между словами, перенести информацию об «отождествлении объектов», которую эти связи содержат, с самих связей в узлы, соединяемые ими. Для этого отметим оба конца анафорической связи одной и той же переменной, которая будет обозначаться малой латинской буквой в квадратных скобках. Существительным, не имеющим анафорических связей, также припишем свою переменную. Тогда в дереве зависимостей информация об

<sup>3</sup> Ограничимся лишь видом анафорических связей, состоящих в «отождествлении объектов». Возможны более сложные типы таких связей.

«отождествлении объектов» будет представляться с помощью приписанных некоторым узлам переменных и указанием их первичных вхождений таким образом, что всем упоминаниям различных объектов будут соответствовать различные переменные и все упоминания в тексте одного объекта будут обозначены одной и той же переменной.

Первичным вхождением переменной в текст<sup>4</sup> назовем то ее вхождение, которое соответствует первичному вхождению существительного; остальные вхождения, соответствующие противоположным концам анафорических связей данного существительного, будем называть вторичными (см. пример 1).

2. Порождающая семантическая модель. Для представления как промежуточных, так и конечных результатов семантического анализа предложения используется порождающая семантическая модель,<sup>5</sup> которая определяет интерпретацию предложения «рекурсивно» — через анализ значения его составляющих.

Язык порождающей семантической модели, принятый в данной работе (язык СМ), представляет собой синтаксическое расширение языка узкого исчисления предикатов (ИП). Необходимость такого расширения обусловлена отсутствием в ИП синтаксических возможностей для записи содержания отдельных составляющих (меньше целого) предложения, т. е. для представления промежуточных результатов семантического анализа. Это расширение логического языка исходит из некоторых свойств структуры предложения, и в этом смысле язык СМ ближе к естественному языку, чем язык ИП.

В словарном представлении отдельных слов естественного языка СМ учитывает их синтаксические валентности. Например, интерпретацией прилагательного «смежный» в словаре языка СМ будет формула вида  $AxO_1y$  **Смежный** ( $x, y$ ), отражающая тот факт, что в предложении данное прилагательное будет связано с существительным «по месту  $x$ » определенной связью  $A$ , а «по месту  $y$ » связью  $O_1$  — первой объектной (**смежный с чем**).

Существительному «порядок (группы)» сопоставляется формула  $JxGy$  **Порядок** ( $x, y$ ), в которой символом  $G$  обозна-

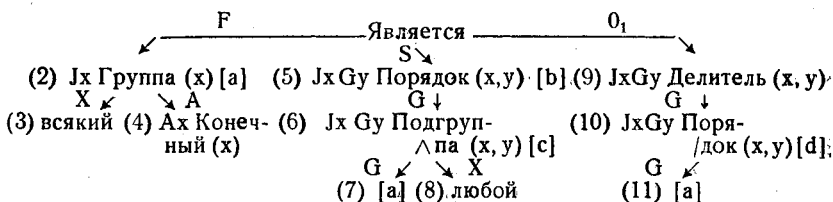
<sup>4</sup> Представление текста в виде дерева зависимостей также будем для краткости называть текстом.

<sup>5</sup> В тезисах Г. С. Цейтина «Язык математической логики как средство исследования семантики естественного языка» эта модель названа «переводящей семантической моделью» (стр. 329).

цена обязательная синтаксическая валентность (генитивного типа) данного слова и символом J помечена переменная x, значение которой определяется объектом, называемым самим существительным.

Синтаксические валентности слова в языке СМ записываются с помощью валентных кванторов, указывающих типы этих валентностей по каждому «месту» предиката, соответствующего данному слову.

Заменив в узлах дерева зависимостей отдельные слова их интерпретациями на языке СМ и разметив узлы переменными по описанному выше правилу, получим структуру, являющуюся представлением входного текста для семантического анализа (в данной семантической модели). Для примера 1 эта структура будет следующей (узлы произвольно пронумерованы числами в скобках).



Интерпретация слов, оставшихся незамененными (так называемые служебные слова), задается особыми правилами их обработки при анализе предложения, совокупность которых для каждого такого слова и является описанием их семантики.

Вхождения переменных a, b, c, d в узлы 2, 5, 6, 10 соответственно будут первичными вхождениями данных переменных. Вхождения переменной a в узлы 7 и 11 являются вторичными.

Для интерпретации более крупных (чем отдельное слово) составляющих порождающая семантическая модель существенно использует информацию о синтаксической структуре предложения. Семантическая модель сопоставляет порождающей операции в синтаксической модели (присоединение поддерева к узлу-хозяину по некоторому типу связи) порождающую операцию, дающую по уже известным интерпретациям присоединяемой и присоединяющей подструктур интерпрета-

цию<sup>6</sup> полученной подструктуры. Назовем эту операцию замыканием связи.

Например, замыкание связи типа G в поддереве (10)  $\xrightarrow{G}$  (11) дает интерпретацию всего поддерева с вершиной (10) в виде формулы: Jx Порядок (x, a). Учитывая полученную интерпретацию поддерева (10), замкнем связь G узла (9):

$$\frac{Jx Gy \text{ Делитель } (x, y) \rightarrow Jx \text{ Порядок } (x, a) [d]}{Kd \text{ Порядок } (d, a): Jx \text{ Делитель } (x, d)}$$

Формула под чертой является интерпретацией всего поддерева (9) и имеет следующий смысл: «x, являющийся делителем d, где d является порядком a». Из примера видно, в чем состоит в данном случае выполнение операции замыкания связи.

Кроме валентных кванторов, в язык СМ включены ограниченные первичные кванторы. Они предназначены для записи «кванторной информации», извлекаемой из текста в ходе семантического анализа и используемой в дальнейшем для выработки типа кванторов и их расстановки в логической формуле. Ограниченный первичный квантор представляет собой выражение вида KxΦ(x);, которое можно интерпретировать так: «...где x удовлетворяет условию Φ(x)». Здесь K — тип первичной кванторной информации по переменной x; Φ(x) — условие, или оговорка, ограничивающая область допустимых значений выделенной переменной x.

Произвольную последовательность ограниченных первичных кванторов и валентных кванторов (с переменными) будем называть приставкой. В предыдущем примере выражение Kd Порядок (d, a): Jx является приставкой.

Формулой языка СМ является любая формула узкого исчисления предикатов, перед которой стоит (или не стоит) приставка.

3. Первичная кванторная информация. Наблюдения над текстами из различных областей математики показывают, что естественный язык придерживается достаточно регулярных правил «обращения» с кванторами, зависящих не столько от конкретного содержания математических

<sup>6</sup> В дальнейшем как синоним термина «интерпретация» употребляется также термин «перевод».



утверждений, сколько от формы этих утверждений. Эта регулярность проявляется как в выборе в пределах однотипных синтаксических конструкций некоторых стандартных средств для выражения типа квантора — существования или общности, так и в использовании некоторых характеристик структуры предложения (текста) для осмысления области действия соответствующих кванторов.

Задача, следовательно, состоит в выработке лингвистических признаков в терминах структуры предложения, необходимых для правильного анализа «кванторных значений», т. е. для определения типа квантора и его области действия в интерпретации предложения на логическом языке.

Указание на тип квантора вырабатывается в процессе анализа для данной переменной один раз при первичном вхождении ее в текст, т. е. при первичном вхождении существительного, называющего некоторый объект. Вторичные вхождения переменной (вхождения слов-заместителей) обычно в тексте никакой кванторной информации не имеют. Однако окончательный тип квантора, каким переменная будет связана в логической формуле, не всегда удается определить по первичному вхождению переменной в текст. Для этого часто требуется рассматривать более широкий контекст, учитывая, например, тип конструкции, содержащей данное первичное вхождение существительного, и некоторые другие признаки.

Кроме того, первичное вхождение существительного при переводе на логический язык определяет лишь начало области действия возникающего в этот момент квантора по переменной, о правой же границе области действия пока ничего не известно, поскольку в нее должны попасть также все вторичные вхождения данной переменной, если таковые имеются, но они могут быть еще не учтены из-за отсутствия интерпретаций других составляющих предложения.

Для учета перечисленных особенностей естественного языка и представления промежуточных результатов перевода в логический язык и введены ограниченные первичные кванторы.

Кванторная информация при обработке первичного вхождения существительного фиксируется при помощи первичного квантора по переменной, приписанной этому существительному. Само существительное, переводимое  $n$ -местным предикатом, дает при этом условие, или оговорку, ограничивающую область допустимых значений выделенной переменной.

В зависимости от способа выражения кванторной информации при существительных выделяется несколько типов первичных кванторов. Разделение этих типов связано с тем, что для каждого из них формулируются свои правила замены в какой-то момент семантического анализа обычными кванторами существования или общности с необходимым преобразованием оговорок (правила реализации).

Выделяются два типа первичных кванторов, которые не изменяются при реализации. Тип этих кванторов удастся точно установить уже при первичном вхождении переменной.

I. Первичный квантор типа квантора общности. Выражается словами «любой», «каждый», «всякий», «произвольный», «любой наперед заданный» и др. Обозначается знаком квантора общности —  $\forall$ .

II. Первичный квантор типа квантора существования. Средства выражения в языке этого типа квантора весьма разнообразны. Кроме тех, с помощью которых значение «существования» выражается как бы в чистом виде, например, «существует», «имеется», «найдутся», «имеет», «определенный», «некоторые» (только во множественном числе!), встречается много слов с различными «сопутствующими» значениями — «проведем», «построим», «можно построить», «можно представить в виде», «фиксированный».

Поскольку конечный результат семантического анализа предложения представляется формулой узкого исчисления предикатов, многие из этих сопутствующих значений при анализе просто теряются. Извлекается лишь информация о кванторе существования  $\exists$ .

Известно, что в естественном языке кванторная информация весьма часто не выражается явным образом, лексически. В этом случае говорят о «нулевом» выражении ее. Тем не менее, предложения с таким нулевым выражением кванторов понимаются однозначно, причем в одних случаях эти нулевые способы соответствуют квантору  $\forall$ , в других —  $\exists$ . В примере 1 в интерпретации всего предложения переменная  $b$  должна быть связана квантором общности, а переменная  $d$  — квантором существования.

Аналогичные замечания можно сделать и по поводу слов «некоторый» (только в единственном числе!), «какой-нибудь», «какой-либо», которым также соответствуют различные кванторы.

Наличие нулевого выражения кванторной информации или слова «некоторый» при первичном вхождении существительного требует дополнительных признаков для уточнения типа квантора, а именно учета относительного расположения слов в предложении (в некотором смысле «порядка слов» в предложении).

В указанном случае и вырабатывается третий тип первичной кванторной информации.

III. Первичный квантор типа неопределенного квантора — Н.

Следует отметить один частный случай нулевого выражения кванторной информации при существительных, когда такой способ вполне «закономерен» и объясним. Речь идет о существительных, которым в логическом языке можно сопоставить функтор, например «сумма (чисел), произведение (чисел)». Если вводить переменную, соответствующую слову «сумма», то ее значение однозначно определяется значением упоминаемых чисел. Такая переменная (область допустимых значений которой ограничена единственным объектом) не требует квантора или по ней можно ставить любой квантор без искажения смысла. Существительные, определяющие единичные объекты, можно специально пометить в словаре и, учитывая эту информацию, «снять» для них проблему кванторов. Если же такому существительному в ходе анализа предложения и припишется неопределенный первичный квантор Н, на правильность семантического анализа это не повлияет.

Возможны выделения и других типов первичных кванторов, в частности, в связи с обработкой множественного числа существительных, количественных числительных и др. Выделенные выше три типа являются наиболее типичными, наиболее частыми.

В соответствии с выделенными первичными кванторами допускаются следующие виды выражений — ограниченных первичных кванторов:

$\forall x\Phi(x)$ : — «для всякого  $x$ , такого, что  $\Phi(x)$ ...»;

$\exists x\Phi(x)$ : — «существует  $x$ , такой что  $\Phi(x)$ ...»;

$\text{H}x\Phi(x)$ : — «...где  $x$  удовлетворяет условию  $\Phi(x)$ ».

Справа дано возможное чтение этих выражений.

4. Предикативные и полупредикативные узлы в структуре предложения. Для анализа

семантики предложения, в частности для анализа кванторных значений, особую роль играют предикативные и полупредикативные узлы. Предикативными узлами являются вершины главного и придаточных предложений, в полупредикативных узлах могут стоять, например, постпозитивные прилагательные, имеющие обязательные активные валентности, причастия.

По отношению к этим узлам выделяют части предложения, называемые темой и ремой высказывания.

Гипотеза о выработке первичных кванторов при нулевом их выражении состоит в том, что для переменных с первичным вхождением в тему высказывания наблюдается интерпретация всеобщности, а переменные из ремы связываются квантором существования.

Сравним следующие предложения (отвлекаясь от оценки их истинности).

(1) «Всякая скалярная матрица перестановочна с любой матрицей  $A$ ». Смысл предложения не изменяется, если устранить кванторное слово «всякая».

(2) «Скалярная матрица перестановочна с любой матрицей  $A$ ». «Трансформируем» это предложение.

(3) «\*Любая матрица  $A$  перестановочна со скалярной матрицей». (Предикат «перестановочна» симметричен.)

Предложение (3) не эквивалентно по смыслу предложению (2). В интерпретации предложения (3) по «скалярной матрице» должен стоять квантор существования, тогда как в предложении (2) — квантор общности.

Указанная гипотеза подтверждается также и тем, что противоположные кванторы (по переменным темы — квантор  $\exists$ , в реме — квантор  $\forall$ ) в теме и реме всегда имеют явное лексическое выражение<sup>7</sup>, так что для них удается сразу выработать точный тип квантора.

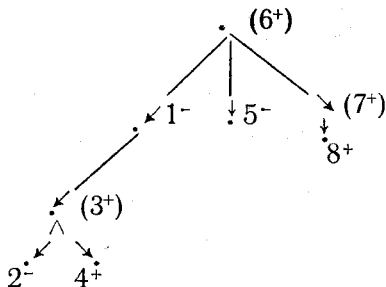
Ввиду трудности формальной выработки признаков темы и ремы можно заменить их, видимо, эквивалентным для письменного математического текста признаком порядка слов следующим образом.

Все узлы поддеревьев, которые в линейном порядке предложения расположены левее данного предикативного узла, составляют тему этого узла, узлы правых поддеревьев — рему. Сам предикативный узел включается в рему.

<sup>7</sup> Например, **Всякая скалярная матрица перестановочна с любой матрицей  $A$** , где кванторное слово **любой** в реме высказывания обязательно.

Если в одном из поддеревьев, в свою очередь, встречается предикативный (полупредикативный) узел, для него признаки темы и ремы вырабатываются независимо по тому же правилу. Для полупредикативных узлов указывается лишь рема. В следующем примере предикативные и полупредикативные узлы дерева заключены в скобки, признак темы в узлах помечен знаком  $-$ , ремы — знаком  $+$ .

Пример 2. Если лупа  $Q_0$  изотопна группе  $Q_1$ , то она является группой, изоморфной группе  $Q_1$ .



Таким образом, в структуре предложения признаки темы и ремы могут встретиться на нескольких уровнях<sup>8</sup>.

Каждое первичное вхождение существительного получит признак темы или ремы. Будем приписывать этот признак также и возникающему при переводе данного существительного первичному неопределенному квантору  $H$  (для других типов первичных кванторов этот признак несуществен) в виде  $H^-$  — для признака темы и  $H^+$  — для признака ремы. В приведенном примере 2 первичный квантор при существительном 2 вырабатывается в виде  $H^-$ , а для существительного 4 — в виде  $H^+$ .

Если возникают условия для реализации первичного квантора (см. п. 5), то  $H^-$  (из темы) трактуется как квантор общности, а  $H^+$  (из ремы) — как квантор существования.

Пока не возникнут условия, разрешающие реализацию, первичные кванторы в процессе перевода предложения на логический язык могут оставаться какое-то время нереализованными. При этом знак нереализованного квантора типа  $H$ , вы-

<sup>8</sup> См. также: М. И. Откупщикова. Грамматика порядка слов для синтеза при автоматическом переводе. Автореф. канд. дисс. Л., 1967.

работанный первоначально, может быть изменен в зависимости от дальнейшей «судьбы» составляющей, в которой он возник. Если в переводе какого-либо поддерева содержатся нереализованные, положительные первичные кванторы  $H^+$  и это поддерево присоединяется к отрицательному узлу, то в переводе подструктуры, получившемся после замыкания связи, все вхождения первичных кванторов  $H^+$  заменяются на  $H^-$ . Замена  $H^-$  на первичный квантор  $H^+$  никогда не производится.

Например, интерпретация поддерева с вершиной  $3^+$  будет содержать два нереализованных квантора, возникших при обработке первичных вхождений существительных 2 и 4.

$H^+$  x **Группа** (x) :  $H^-$  y **Луна** (y) : **Изотопны** (x, y).

При замыкании связи  $1^- \rightarrow 3^+$  положительный первичный квантор  $H^+$  заменяется на  $H^-$  и будет получен перевод:

$H^-$  x **Группа** (x) :  $H^-$  y **Луна** (y) : **Изотопны** (x, y).

Содержательное пояснение этого преобразования может быть следующим. Тема некоторого предикативного узла остается темой и для всего последующего высказывания, для дальнейшего текста, в то время как рема одного предикативного (полупредикативного) узла может входить в тему другого, лежащего выше в структуре высказывания, узла. Оставшиеся нереализованными первичные кванторы с оговорками «переносят» некоторые объекты высказывания в тему следующего предикативного узла, рема которого будет содержать об этих объектах какую-то новую информацию (рема этого предикативного узла будет содержать вторичные вхождения переменных, кванторы по которым еще не реализованы).

Особая роль предикативных (полупредикативных) узлов состоит, как будет показано ниже, еще и в том, что только при замыкании их синтаксических связей могут возникать условия для реализации первичных кванторов с оговорками и, следовательно, они являются ограничителями области действия кванторов, выполняя для них как бы роль скобок.

5. Условия реализации первичных кванторов с оговорками. Формула расширенного логического языка  $SM$  представляет собой формулу ИП, которой может предшествовать последовательность ограниченных первичных кванторов и валентных кванторов — приставка.

Правило замены (реализации) первичного квантора окончательным определяется типом первичного квантора, но момент применения этого правила, момент реализации первичного квантора зависит от вида формулы СМ, стоящей справа от данной оговорки (после двоеточия).

Если на некотором шаге перевода предложения на логический язык непосредственно справа от какого-либо ограниченного первичного квантора (с оговоркой) получилась формула СМ, которая не имеет приставки, т. е. является «чистой» формулой ИП и переводом такой части текста, в которой содержатся все вторичные вхождения переменной, выделенной в данной оговорке (т. е. учтены все анафорические связи первичного вхождения существительного), то данный первичный квантор следует реализовать.

Для первичных кванторов типа I и II правила реализации будут следующими:

$$\forall x\Phi(x) : \Psi \implies \forall x(\Phi(x) \supset \Psi);$$

$$\exists x\Phi(x) : \Psi \implies \exists x(\Phi(x) \& \Psi).$$

Здесь  $\Psi$  — формула СМ, являющаяся «чистой» формулой ИП, в которой учтены все вхождения переменной  $x$  в переводимый текст<sup>9</sup>.

Реализация первичного квантора, как это видно из приведенных правил, окончательно устанавливает область действия данного квантора и его вид. Дальнейшая обработка текста (если она имеет место) ничего в полученной после реализации формуле (подформуле) изменить не может.

Поэтому для обеспечения правильного порядка кванторов в переводе необходимо, в соответствии с первым условием реализации, определить порядок первичных и валентных кванторов в приставке таким образом, чтобы в конце стояли те первичные кванторы, которым в интерпретации должна соответствовать меньшая область действия.

6. Определение порядка кванторов. Порядок кванторов в логической формуле не всегда совпадает с «порядком слов» в предложении.

(1) Любое четное число делится на некоторое простое число.

$\forall x$  (Четное число( $x$ )  $\supset$   $\exists y$  (Простое число( $y$ )  $\&$  Делится( $x, y$ )))

<sup>9</sup> Из правил видно, что в принятой семантической модели роль существительных (именных групп) вместе с кванторной информацией при них состоит во введении ограниченных кванторов.

(2) Любое четное число делится на одно и то же простое число.

Эу(Простое число (у) & Vx (Четное число (х) ) Делится (х, у))

В предложении (1) «порядок слов» (имеется в виду порядок именных групп с кванторными словами при них) совпадает с порядком кванторов в логической формуле, а в предложении (2) — не совпадает.

Кванторному слову «один и тот же» соответствует более «сильный» по сравнению со словом «некоторый» квантор существования, что выражается в расширении области действия этого квантора. Подобные «сильные» кванторные слова употребляются в предложениях в тех случаях, когда при заданном порядке слов требуется выразить несовпадающий с ним порядок кванторов.<sup>10</sup> Таким образом, роль многих кванторных слов состоит не только в указании типа квантора (который выражается и другими средствами), но и в «усилении» квантора. Так, «сильный» квантор существования дают слова: «один и тот же», «определенный», «фиксированный», «некоторые» (множ. число); «сильный» квантор общности: «любой наперед заданный».

Замыкание синтаксических связей в дереве зависимостей производится таким образом, чтобы последовательность первичных кванторов в приставке соответствовала необходимому порядку кванторов в интерпретации предложения на логическом языке.

При замыкании связей важно учитывать следующие формальные ограничения.

1) Квантор с переменной, свободно входящей в оговорку другого квантора, должен предшествовать этому квантору. Область действия второго квантора должна быть уже, так как множество допустимых значений его переменной зависит от конкретного значения переменной первого квантора.<sup>11</sup>

2) В остальных случаях квантор с переменной, не имеющей вторичных вхождений вне рассматриваемой составляющей, должен стоять после квантора с переменной, имеющей

<sup>10</sup> Ср.: Е. В. Падучева. Логический анализ русских кванторных прилагательных. НТИ, 2 сер., Информационные процессы и системы, 1972, № 2.

<sup>11</sup> Г. С. Цейтин. О промежуточном этапе при переводе с естественного языка на язык исчисления предикатов. Тезисы докладов на конференции по обработке информации, машинному переводу и автоматическому чтению текста. М., 1961, стр. 107—111.



такие вхождения (чтобы первый мог реализоваться и до присоединения упомянутых вторичных вхождений).

3) «Сильный» первичный квантор предшествует «обычному» первичному квантору (при любом порядке соответствующих кванторных слов в предложении).

Необходимый порядок первичных кванторов в приставке, удовлетворяющий этим ограничениям, достигается определенным порядком замыкания связей в узлах дерева и определенным способом расстановки первичных и валентных кванторов при замыкании связи, зависящим от «силы» квантора.

Нормально присоединение поддеревьев к узлу дерева производится в порядке линейного расположения этих поддеревьев в предложении. При этом «порядок слов» (относительно рассматриваемого узла) и порядок кванторов будут совпадать. Однако если присоединяемое поддерево имеет признак «сильного» первичного квантора, то это поддерево присоединяется раньше других поддеревьев, не имеющих такого признака, т. е. замыкание связей в узле дерева упорядочено с учетом «силы» первичной кванторной информации в поддеревьях.

В заключение приведем общий вид правил замыкания «стандартной» синтаксической связи а) предикативного и б) полупредикативного узла. Правила учитывают изложенные выше ограничения на порядок кванторов. В записи правил над чертой указывается условие применения правила, под чертой — результат применения. Результатом правила является перевод на язык СМ некоторой подструктуры предложения, построенной по переводам ее составляющих:<sup>12</sup>

$$(a) \quad \frac{\Delta R^1 R_x \Phi(x) \xrightarrow{R} \Sigma^1 \Sigma J_x \Psi(x)[a]}{\Delta \Sigma^1 \Sigma K_a \Psi(a) : R^1 \Phi(a)}$$

$$(б) \quad \frac{Az R^1 R_x \Phi(x, z) \xrightarrow{R} \Sigma^1 \Sigma J_x \Psi(x)[a]}{\Sigma^1 Az \Sigma K_a \Psi(a) : R^1 \Phi(a, z)}$$

Здесь  $\Delta R^1 R_x \Phi(x)$  — интерпретация предикативного узла-хозяина;  $Az R^1 R_x \Phi(x, z)$  — интерпретация полупредикатив-

<sup>12</sup> См. также: В. Д. Буторов. Интерпретация составляющих предложения в одной семантической модели. В сб.: Автоматическая переработка текста методами прикладной лингвистики. Материалы Всесоюзной конференции 6—8 октября 1971 г. Кишинев, 1971, стр. 173—175.

ного узла-хозяина;  $\Sigma^1 \Sigma J x \Psi(x)[a]$  — интерпретация зависимого узла с приписанной переменной  $a$ ;  $\Delta, \Sigma$  — некоторые последовательности ограниченных первичных кванторов;  $\Sigma^1$  — последовательность первичных кванторов с переменными, имеющими вторичные вхождения в текст;  $R^1$  — последовательность валентных кванторов, незамкнутых связей узла-хозяина;  $R$  — валентный квантор;  $A$  — пассивная валентность полу-предикативного узла;  $\Psi(x), \Phi(x), \Phi(x, z)$  — формулы  $SM$ , не содержащие приставок;  $K$  — первичный квантор по переменной  $a$ .

Если квантор  $K$  является «сильным» первичным квантором, то результатом правила (б) будет формула:

$$\Sigma^1 \Sigma K a \Psi(a) : A z R^1 \Phi(a, z).$$

А. Г. Давтян

### ОПЫТ СЕГМЕНТАЦИИ СЛОЖНЫХ ПРЕДЛОЖЕНИЙ НА ОСНОВЕ ОГРАНИЧИТЕЛЬНЫХ ЗНАКОВ ПРЕПИНАНИЯ

Автоматический анализ текстов выдвигает в качестве одной из задач проблему исследования знаков препинания в научно-технической литературе. При переводе с английского языка мы сталкиваемся с многозначностью некоторых знаков препинания, присущей и другим языкам. Специфическую трудность представляет то обстоятельство, что система пунктуации английского языка опирается больше на семантический и связанный с ним ритмико-интонационный аспект членения предложения, чем на строгие формальные правила.

Задача настоящей работы состоит в том, чтобы определить функции некоторых ограничительных знаков препинания (точки с запятой и двоеточия) и выделить формальные признаки для сегментации внутри предложения. Для удобства разделим все знаки препинания на два класса: точечные (точка, многоточие, вопросительный знак и восклицательный знак) и неточечные (все остальные), при этом точечные знаки препинания можно считать однозначными в научно-техническом языке, в то время как неточечные знаки препина-

ния однозначными не являются.<sup>1</sup> Синтаксический анализ в системе автоматического перевода, разработанный в лаборатории математической лингвистики ВЦ ЛГУ, является множественным и построен не на прямом алгоритме, а на алгоритме, выявляющем все возможные варианты анализа, допускаемые условиями, предусмотренными в грамматике.<sup>2</sup> С увеличением длины предложения число вариантов анализа, подлежащих проверке на соответствие условиям грамматики, резко возрастает.<sup>3</sup> Поэтому предварительная сегментация предложения на части, приравниваемые к предложению, может существенно ускорить анализ. Отсюда вытекает целесообразность создания такого блока, в котором посредством не очень сложного набора правил производилось бы членение предложения на сегменты, эквивалентные предложению как отрезку текста, структура которого анализируется на этапе синтаксического анализа. Такой блок должен, видимо, следовать за блоком предварительного решения грамматической омонимии, устраняющим в большинстве случаев неоднозначность глагольных форм.

Нами были исследованы тексты по радиолокации из журнала «Electronics» за 1961—1963 гг. (9000 предложений), а также статьи из журнала «Science Journal» за 1970 г.

Рассмотрим сначала точку с запятой. В данном объеме текста встретилось 130 предложений с этим знаком препинания. Как известно, точка с запятой служит для обозначения сравнительно крупного членения предложения. Употребление точки с запятой наиболее типично для сложносочиненного предложения, части которого обладают смысловой авто-

---

<sup>1</sup> См., например: Т. М. Николаева. Анализ знаков препинания при машинном переводе с русского языка. Труды ИТМ и ВТ АН СССР, вып. 1. Сб. статей по машинному переводу, М., 1958, стр. 33—46; Д. Л. Розенберг, Л. Т. Борисова. О проблеме ограничительных знаков при машинном переводе с английского языка. Уч. зап. Горьковск. ун-та, вып. 66, 1963, стр. 65—72; А. Р. Белопольская, А. Б. Ковригин. Членение сложного предложения (немецкий язык). Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста, вып. 2. М., 1961.

<sup>2</sup> Б. М. Лейкина, Т. Н. Никитина, М. И. Откупщикова, С. Я. Фитиалов, Г. С. Цейтин. Система автоматического перевода, разрабатываемая в группе математической лингвистики ВЦ ЛГУ. НТИ, 1966, № 1.

<sup>3</sup> Б. М. Лейкина. Принципы построения английской грамматики для синтаксического анализа при автоматическом переводе. НТИ, 1964, № 11.

номностью и относятся к одному и тому же синтаксическому уровню (parts of equal rank). В этом случае одно из сочиненных предложений или оба сочиненных имеют осложненную структуру и содержат обособленные (или однородные) члены или придаточные предложения. В 98 предложениях из 130 (т. е. примерно в 75% случаев) функцию точки с запятой можно приравнять к функции точки с полным сохранением смысла предложения. Например:

The second fence effectively reduces the intensity of the out-of-phase elements caused by the main fence, leaving a virtually uniform, uncluttering wave front; the slight *distortion* remaining *poses no problem* for the acquisition radar.

Правила предварительной сегментации для предложений, содержащих точку с запятой, содержат ряд проверок на наличие отдельных признаков. В число этих признаков входят глагол с однозначным признаком — «личная форма» справа от точки с запятой, беспредложное существительное слева от этого глагола и в случае переходности глагола наличие беспредложной формы существительного справа от него. В результате ручной проверки блока предварительной сегментации 88 предложений сегментировались правильно, например:

At the power levels anticipated, minority carrier injection into the heavily doped regions is negligible; hence, recombination of holes and electrons takes place almost entirely in the first layer.

Во всех исследованных предложениях встретились следующие неоднозначные глагольные формы, неоднозначность которых была разрешена в блоке грамматической омонимии: can cause, will be controlled, can be overridden, can be tolerated, will be applied, will result, must be, perfect, yield, shows, shifts, means, passes и т. д.

Сочиненные предложения не выделились в сегменты-предложения лишь в 8 предложениях, так как, во-первых, в блоке сегментации нет правил для случаев инверсии, что не является типичным для научно-технических текстов, во-вторых, отсутствуют правила и для эмфатических конструкций с глаголом to do. Например:

The maximum rate of digital output is one set of five variables per second per tape punch; should an output rate of not more than one function value per second be adequate, the typewriter may be connected directly to the computer.

Now consider the time interval when pump II is on and pumps I and III are off; signals do flow from  $P_2$  to both  $P_1$  and  $P_3$  (both idle).

The aeriels of the interrogator-respondors of these systems have usually been designed for mounting on top of a primary radar aerial and have consequently been of small vertical aperture; there has thus been little chance of modifying the vertical coverage pattern to compensate for the depredations of swept gain and considerable loss of high cover has had to be accepted.

Отрицательный результат первой проверки, т. е. отсутствие однозначной финитной формы глагола, устраняет необходимость дальнейших проверок. При невыполнении хотя бы одного из перечисленных условий сегментация не производится.

Вторую группу составляют случаи, когда точка с запятой служит для связи однородных беспредложных или предложных именных групп, инфинитивных оборотов (обособленных и пояснительных членов предложения), особенно при наличии внутри них других знаков препинания. Например:

A short-duration-pulse radar can be used to advantage in the laboratory to conduct remote thickness measurement, evaluate recovery times, breakdown phenomena, relaxation times and other time-dependent microwave effects; to determine locations of imperfections in wave guides and microwave tubes, and to explore propagation phenomena such as multipath reflections and atmospheric indirect time delays due to weather, clouds, temperature etc.

В случаях этого типа сегментация недопустима. Таким образом, задача сводится к определению того, являются ли части предложения, разделенные точкой с запятой, сочиненными предложениями или однородными членами более низкого синтаксического уровня.

При исследовании двоеточия выделились четыре основных типа его употребления в тексте.

**Тип 1.** Двоеточие после бокового заголовка (side-heading) для отделения его от следующего за ним текста. Четким формальным показателем для выделения бокового заголовка в самостоятельный сегмент служит типографский шрифт (курсив либо жирный шрифт) и/или графические особенности (заглавные буквы у всех знаменательных слов именной группы, составляющей боковой заголовок). Например:

*Operational Flight Simulator Tie-In:* When this radar simulation device is to be used as a portion of a flight...

Первый тип самый простой. После выделения бокового заголовка в сегмент приписываем ему информацию «боковой сегмент».

**Тип 2.** Двоеточие перед перечислением однородных членов, которому предшествует обобщающее слово. Здесь выделяются два подтипа.

**Подтип 1.** Перечисление, вводимое словами *as follows*, «the following + существительное», количественными числительными, квантитативными словами (*several, many* и т. п.), а также словами типа *these, such*, например:

*Briefly, these units interoperate as follows:*

1) When the power to the printer is switched on, a pair of contacts on the function translator starts to transmit 50 sec control signals to the buffer stage at intervals of 10 msec.

2) When the tape reader is switched on it reads the first character in the tape from the model 3000 tape punch and transmits the combination on the five wires to the buffer stage.

Developers of Research Institutes radar-acoustic remote air velocity measuring system are predicting it will find *many uses*: assisting jet landings on carriers, as a wind tunnel calibration standard, for air pollution studies and micrometeorology.

In using modern long-range radars, such techniques are no longer feasible for *these reasons*: although radar range has been greatly extended, the video of interest is normally confined to a small segment of overall range; the extended range usually requires a round-trip propagation time that extends over several interpulse periods of the radar prf.

This has *two disadvantages*: first, such replies even though rejected by the interrogator-responder which caused them, may yet be received by other responders within range and, being on a radio frequency common to all responses, be displayed as a form of interference known as «fruit»; secondly, to avoid over-leading the transponder transmitter, it is necessary to limit the transponder reply capability by desensitizing its receiver when the rate of reply reaches a certain figure.

**Подтип 2.** Перечисление, вводимое глаголом *to include*, конструкцией *there are*, а также финитными формами глагола *to be*. Например:

New factors used in the FPQ-6, according to RCA, *include*: a revised console to provide the most useful acquisition featur-

res for the operators as well as automatic, semiautomatic and manual equipment to supplement the operators; improved built-in acquisition features in the ranging equipment; an auxiliary nonreferenced range system which permits...

*There are, however, new rating requirements that must be included.*

**Peak Fault Current** — Under certain conditions of fault, the magnitude of peak current may be as high as four times the rated continuous peak current.

**RMS Fault Current** — As with peak current, rms fault current can reach approximately four times the value of continuous rms current.

**Averaging Time** — To prevent damage to the clipper tube, because of too long or too frequent application of fault current, a fault — current averaging time must be included in the rating.

The equipments that comprise the system *are*: the PPDD console containing the direct view storage tube and operator controls; the unit processor rack required for each display containing the power supplies, video, analog and logic circuits for the display console; and the group processor rack required for each display system and containing functions common to all the display consoles.

Наиболее удобно задать список слов, характеризующих подтипы 1 и 2 из второго типа, и проверять, не содержится ли перед двоеточием слово из заданного списка. Затем производится ряд проверок справа от двоеточия, аналогичных тем, которые были описаны для случая точки с запятой. Если после проверок справа окажется, что там вычленяются сегменты-предложения, то тогда предшествующую двоеточию часть предложения также выделяем в самостоятельный сегмент. Например:

An approximation to the desired result may be obtained in several ways amongst them being *the following*: (a) *It is* sometimes possible to calculate some of the moments of the required distribution; (b) *It is* usually possible to measure the probability distribution experimentally; (c) a more versatile *method* which has become very attractive with the advent of high-speed digital computers *involves* specification of the parameter whose probability distribution is required in the form of a mathematical model.

**Тип 3.** Двоеточие между сочиненными предложениями, не соединенными союзом; при этом во втором предложении либо излагается причина того, о чем говорится в первом, либо содержатся характеристики объекта, о котором говорится в первом предложении. Например:

On the model 1000 printer the converse method is employed: the letter feed operation takes place only when a «letter feed» signal is received by the function translator from the code convertor.

Of these storms, which were not covered by bulletins, five occurred, when WLW meteorologists were not on duty: two were freak disturbances that developed and disappeared in a few moments; two occurred during a period of mechanical break down in the radar system.

В случаях типа 3 смысловая автономность сочиненных предложений очевидна, и можно считать, что здесь мы сталкиваемся с синонимией двоеточия и точки с запятой. Поэтому для типа 3 предлагается тот же набор проверок, что и в рассмотренном выше случае для точки с запятой. В результате ручной проверки 38 предложений из типа 3 все предложения выделились в сегменты-предложения. К данному типу относится также крайне редкое употребление в одном предложении двух двоеточий, несущих разъяснительную функцию. Например:

The problem is not confined to primary radar interrogation: in no practical radar aerial system can the transmitted radiation be confined to a single beam: some of the energy is radiated in other directions and can, under certain circumstances, give rise to false echoes (in a primary radar system) or false interrogation.

**Тип 4.** Двоеточие после слов, вводящих цитату или прямую речь. Для типа 4 характерным является наличие перед двоеточием глаголов речи, мысли, чувства (state, think, write и т. п.). Этот тип представлен бедно: всего 3 предложения на 9000 предложений рассмотренного текста. Например:

Yet a Government spokesman, according to Science News Service, states: «There is every reason to believe the Russians are working on such paints and they just might even be ahead of us; if they perfect anti-radar paints, what will happen to our Distant Early Warning system in Canada, and Texas Tower



radar posts in the Atlantic?». Для типа 4 производим проверку на наличие одного из глаголов, заданных списком, перед двоеточием; в случае его нахождения приписываем ему признак валентности и выделяем часть предложения перед двоеточием в сегмент-предложение.

Г. М. Ильин

## СЕМАНТИКА ИНФОРМАЦИОННЫХ ЯЗЫКОВ

Рассматривая семантику информационных языков (ИЯ), следует иметь в виду те задачи, для которых эти языки создаются.

Во-первых, информационный язык должен допускать возможность достаточно полного представления на этом языке содержания документа (в случае документографических систем) или какого-либо факта (в случае фактографических систем). В связи с этим требованием возникают различные вопросы о выборе элементов ИЯ (например, вопросы о количестве дескрипторов, необходимых для функционирования информационной системы в какой-либо конкретной предметной области), выборе грамматики ИЯ, такой, чтобы конструкции языка, составленные из одинаковых элементов, но с помощью разных правил грамматики, выражали разное содержание.

Во-вторых, информационный язык должен допускать возможность формального определения правил преобразования языковых выражений для обеспечения работы информационной системы в целом. С помощью правил преобразования формулируется основное соотношение — соответствие запроса и ответа. Хотя обычно соответствие запроса и ответа определяется просто как соответствие двух выражений языка, в любой системе еще имеются и правила преобразований, которые могут применяться к выражениям ИЯ перед проверкой указанного соответствия. Поэтому можно считать, что правила преобразований и определяют соответствие запроса и ответа. Какое-либо выражение ИЯ есть ответ на запрос, если это выражение можно преобразовать по строго опреде-

ленным правилам в выражение, соответствующее запросу. Правила преобразования, определяющие работу системы, являются содержательными, т. е. связанными с содержанием выражений ИЯ. Это либо правила синонимических преобразований, либо правила следования.

Необходимо отметить, что вторая задача — формулировка правил преобразований выражений ИЯ — основная при создании информационной системы. Первая из указанных задач — задача возможно более полного отражения содержания — всегда решается на основе второй. Бессмысленно требовать от языка отражения таких семантических различий, которые не могут быть учтены правилами преобразований.

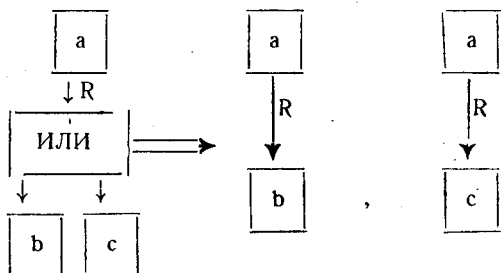
В достаточно богатых информационных языках должны существовать элементы, которые специально вводятся для того, чтобы в рамках системы можно было сформулировать правила преобразования, учитывающие более тонкие семантические различия в предложениях ИЯ. Естественно поэтому считать, что их семантика полностью определяется теми правилами, которые сформулированы на основе введенных элементов. Эти элементы не имеют никакого «терминологического» содержания. Они служат лишь средствами связи между различными содержательными конструкциями ИЯ.

Проиллюстрируем сказанное выше на примере языка одной информационной системы.<sup>1</sup> В качестве элементов в этот язык входят набор дескрипторов, набор отношений и так называемые служебные узлы. Предложением ИЯ является дерево, в узлах которого могут стоять либо дескрипторы, либо служебные узлы, а стрелки помечены какими-либо отношениями из имеющегося набора. Основное правило преобразования — последовательное отсечение веток дерева — интерпретируется как смысловое преобразование, соответствующее расширению описываемого деревом понятия. Считается, что документ соответствует запросу, если из дерева — поискового образа документа можно получить указанным способом дерево — поисковый образ запроса. Это преобразование можно интерпретировать как правило следования, если дерево ИЯ рассматривать как запись некоторой ситуации: описание более конкретной ситуации отвечает на запрос о более

---

<sup>1</sup> Подробное описание модели языка и критерия смыслового соответствия см.: Г. М. Ильин, Н. Н. Ильина, И. И. Матвеева. Модель системы поиска технической литературы по рефератам. НТИ, сер. 2. Информационные процессы и системы. 1970. № 8 стр. 24—29.

общей ситуации. Кроме этого в правила преобразования вводятся также некоторые синонимические преобразования. Например, к числу последних относится правило замены составного дескриптора поддеревом, его определяющим.<sup>2</sup> Особую роль в формулировании правил преобразования играют служебные узлы. Синтаксически служебные узлы употребляются так же, как и обычные дескрипторы. Их роль состоит в том, чтобы обеспечить более тонкое различие содержания документов. Этой ролью и определяется семантика служебных узлов. Их семантика не зависит от того, для какой предметной области разрабатывается язык по данной модели. Например, семантику узлов V и Z можно сформулировать так: в процессе преобразования дерева реферата к дереву запроса любой узел может быть заменен на Z и узел V может быть заменен на любой другой узел. Семантику узла ИЛИ можно выразить в виде такого преобразования:



Исходя уже из точно определенной семантики этих узлов, можно сделать вывод о том, какое содержание они выражают и как эти узлы можно использовать при записи поисковых образов запросов и рефератов.

Сформулированная точка зрения на принципы описания семантики информационных языков интересна и в применении к другим информационным системам. Рассмотрим с этих позиций системы типа «запрос—ответ». В этих системах правила содержательных преобразований над выражениями ИЯ также можно рассматривать как средство описания семантики тех элементов языка, которые влияют на формулировку таких правил.

<sup>2</sup> Ср. аналогичные правила преобразований в книге «Информационно-поисковая система БИТ». Киев, 1963.

Вначале обратимся к системе, построенной американским ученым К. К. Грином на базе языка исчисления предикатов.<sup>3</sup> В системе набор элементарных формул составляет терминологическую часть, т. е. часть, зависящую от конкретной предметной области. Логические же знаки — логические связки и кванторы — это те элементы, семантика которых определяется при помощи правил преобразования системы, т. е. при помощи правил логического вывода. Например, конъюнкция определяется правилом:  $A \& B \implies A, B$ . Иначе говоря, если в системе имеется предложение  $A \& B$ , то в качестве ответа может быть выдано и  $A$  и  $B$ . Знак импликации определяется правилом:  $A \supset B, A \implies B$ . Иначе говоря, если в системе имеются предложения  $A \supset B$  и  $A$ , то в качестве ответа может быть выдано  $B$ . Аналогичным образом в терминах действия системы могут быть интерпретированы и другие правила логического вывода.

В терминах правил работы информационной системы типа «запрос—ответ» может быть описана семантика и ряда слов естественного языка. В естественном языке существует большое количество элементов, служащих для выражения содержательных соотношений между различными текстами. Правила работы информационной системы на базе естественного языка<sup>4</sup> должны быть сформулированы с учетом этих элементов, и, следовательно, эти правила и будут описывать семантику таких слов.

Примеры описания семантики некоторых элементов естественного языка через правила работы системы уже приводились в различных работах.<sup>5</sup> В частности, оказалось полезным выделить средства выражения предиката «включение», так как с помощью этого типа связи можно описывать семантику других слов. Рассмотрим в качестве примера способ описания семантики слова «только». Для простоты объяснения схемы рассмотрим элементарнейшую конструкцию со

<sup>3</sup> См.: С. С. Green. The Application of Theorem Proving to Question-Answering Systems. «Technical Report», 1969, № CS-138.

<sup>4</sup> Описание постановки задачи см. в предисловии наст. сб.

<sup>5</sup> См.: Г. М. Ильин, Б. М. Лейкина, Т. Н. Никитина, М. И. Откупщикова, С. И. Фиталов. Лингвистический подход к задаче построения информационных систем. В сб.: Информационные вопросы семантики, лингвистики и автоматического перевода, вып. 2. М., 1971; Н. Н. Ильина. Об одном способе классификации связей между частями сложного предложения (там же); Н. С. Яковлева. Функция слова наоборот в тексте (там же).

словом «только»:  $P(a, \text{только } b)$ , где  $P$  — какой-нибудь предикат, « $a$ » и «только  $b$ » — места этого предиката.

Такую конструкцию, например, имеет предложение: «Я купил только посуду».

Семантику слова **только** можно определить правилами:

$$P(a, \text{только } b) \implies P(a, b);$$

$$P(a, \text{только } b), \neg \text{ВКЛ}(c, b) \implies \neg P(a, c),$$

где ВКЛ — предикат «включение»,  $\neg$  — знак отрицания. Так, из предложения «Я купил только посуду» следует: «Я купил посуду», а из двух предложений «Я купил только посуду» и «Хлеб не является посудой» следует «Я не купил хлеб».

При формулировании правил с более сложным случаем употребления слова «только» нужно учитывать, что слово **с** должно быть взято из определенного класса.

Интересно сравнить полученное здесь описание семантики слова «только» с имеющимся описанием того же слова в статье И. А. Мельчука.<sup>6</sup>

По существу в статье И. А. Мельчука формула, описывающая семантику этого слова, представляет собой запись правила синонимического преобразования предложения со словом «только». При нашем подходе рассматриваются не только синонимические преобразования, но и правила следования, поэтому формула получилась несколько иная. Кроме того, более полное описание семантики слова «только» дает привлечение дополнительных текстов для получения выводов из конструкций со словом «только» (привлечение предложений, выражающих предикат «включение»).

Таким образом, задача построения информационной системы «запрос—ответ» дает возможность успешно ставить и решать проблемы, связанные с семантикой естественного языка.

Г. М. Кириллова

#### АНАЛИЗ ФОРМУЛ ПРИ АВТОМАТИЧЕСКОМ ПЕРЕВОДЕ НАУЧНО-ТЕХНИЧЕСКИХ ТЕКСТОВ

При автоматическом переводе научно-технических текстов встает задача распознавания и анализа формул. Целесооб-

<sup>6</sup> См.: И. А. Мельчук. К принципам описания означаемых (о лингвистической семантике). В сб.: Язык и человек. Изд. МГУ, 1970, стр. 201—217.

разность семантического анализа математических формул с помощью широкого контекста убедительно показана в статье «Математические формулы в широком контексте»<sup>1</sup>. Изложенный в статье подход к проблеме анализа формул представляется чрезвычайно интересным. Однако автоматическая обработка текста в существующих в настоящее время системах АП проводится только в пределах одного предложения (пофразный перевод). Поэтому становится неизбежным вопрос автоматического распознавания, анализа и перевода математических формул без учета межфразовых связей.

В данной статье делается попытка исследования возможностей такого анализа применительно к системе АП, разработанной в ВЦ ЛГУ.<sup>2</sup> Работа проведена на материале английских и американских текстов по радиолокации. Исследовано 2000 предложений.

Под формулой будем понимать совокупность буквенных или цифровых символов (или их комбинаций), соединенных одним или несколькими математическими знаками и/или имеющих дополнительный верхний или нижний символ (индекс, сноску, показатель степени).

При этом мы исходим из предположения, что единичный символ рассматривается как отдельная словоформа и обрабатывается на этапе морфологического анализа, включающего «Бл. ан. сл.» (Блок анализа слов, не предусмотренных в словаре).<sup>3</sup>

Таким образом, показателем формулы можно считать наличие хотя бы одного из признаков: 1) математического знака, 2) линейной неоднородности.

Одна из задач синтаксического анализа сводится к явной или неявной сегментации текста, членению его на единицы определенных уровней. Трудность, которую вызывают при автоматическом анализе конструкции с формулами, заклю-

<sup>1</sup> См.: С. А. Гюльмисарян, Ф. А. Дрейзин, З. Т. Тер-Мисакянц. Математические формулы в широком контексте. НТИ, сер. 2. Информационные процессы и системы, 1971, № 3.

<sup>2</sup> Б. М. Лейкина, Т. Н. Никитина, М. И. Откупщикова, С. Я. Фитиалов, Г. С. Цейтин. Система автоматического перевода, разрабатываемая в группе математической лингвистики ВЦ ЛГУ. НТИ, 1966. № 1, стр. 40—50.

<sup>3</sup> Б. М. Лейкина. Независимый морфологический анализ английского научно-технического текста. Труды ЦНИИПИ, сер. 3. Автоматический анализ патентных текстов. М., 1971, стр. 12—25.

чается в том, что они функционируют как единицы разных уровней: слово, словосочетание, предложение.

Граматику синтаксического анализа, используемую в указанной выше системе АП, можно охарактеризовать как валентностную грамматику зависимостей. На выходе синтаксического анализа структура предложения представлена в виде дерева зависимостей и информация о каждом слове анализируемого предложения включает номер слова, управляющего этим словом (номер «хозяина»), и характеристику типа связи данного слова с «хозяином» (номер валентного «места»<sup>4</sup>. В связи с этим при анализе формул возникают две проблемы: внутренняя — определение вершины формулы и внешняя — определение «хозяина» данной формулы и типа связи ее с «хозяином».

Для выработки правил обработки формул необходимо провести их предварительную классификацию. Будем различать по синтаксическому статусу формулы «фразообразующие» и «нефразообразующие» и по форме — «чистые» (имеются чисто формальные знаки) и «смешанные» (кроме формальных математических знаков, есть словоформы, основы которых задаются в словаре).

К фразообразующим относятся формулы, содержащие математические знаки «предикативного» типа, которые при анализе можно принимать за вершину сегмента:

$\rangle, =, \equiv, >, \gg, <, \ll.$

В анализирующей грамматике следует учитывать, что все предикативные знаки могут быть аналогами как финитной, так и нефинитной формы глагола в зависимости от того, выступает ли данная формула как самостоятельное предложение или как член предложения.

К нефразообразующим относятся формулы со знаками арифметических действий<sup>5</sup>:

$(+, -, \times, :, /).$

Сюда же будем относить формулы, не содержащие математических знаков, но характеризующиеся линейной неоднородностью.

<sup>4</sup> Б. М. Лейкина, Т. Н. Никитина, М. И. Откупщикова, С. Я. Фитналов, Г. С. Цейтин, ук. ст. НТИ, 1966, № 1, стр. 40—50.

<sup>5</sup> В рассмотренных предложениях встретились только перечисленные знаки. При дальнейшем исследовании указанный ряд может быть расширен.

Предикативные знаки считаем «старшими», и наличие в формулах знаков 2-го класса (знаков арифметических действий) не мешает отнести эти формулы к фразообразующим.

Отдельного рассмотрения требуют неоднозначные формульные знаки: знак деления ( $:$ ), омонимичный знаку препинания (двоеточию), знак вычитания, омонимичный дефису, и верхний индекс, омонимичный показателю степени или сноске. Просмотренного материала недостаточно для выработки правил разрешения омонимии данных знаков, и этот вопрос здесь рассматриваться не будет. Омонимия знака / (знак дизъюнкции, знак деления или скобка) несущественна, так как все члены данного омонимического ряда относятся к нефразообразующим знакам.

Вторая проблема — подведение нефразообразующих формул под какой-то класс словоформ, так как анализ оперирует с определенными классами словоформ.

Формулы, не содержащие предикативных и омонимичных знаков, можно подвести под класс существительных, так как они выступают во всех основных функциях, присущих существительным: подлежащего, дополнения, предикативного члена, препозитивного определения, приложения. Например:

1. If  $F(Q)$  is less than unity...

2. By eliminating the part of this due to the «no target» configuration, we obtain  $F(p)$ .

3. The frequency components emitted by the transmitting antenna are  $(f)$  and  $(f + P)$ .

4. ... three vectors,  $f_1, f_2, f_3$ , lie in one plane...

Формулы, содержащие суффикс  $-th$  порядкового числительного, легко подводятся под класс числительных, так как совпадают с ними по значению и по своим синтаксическим функциям. Например,  $(n-1)th$  можно рассматривать так же, как  $n-th$  или, скажем  $4th$ .

Нижние индексы встретились только при буквенных обозначениях и во всех случаях трудностей не представляют, ибо эти конструкции полностью совпадают с существительными по своим функциям. Они не требуют дополнительных правил анализа и могут быть включены в грамматику как особый подтип существительных. Например:

1. Before  $P_{lm}$  can be calculated it is necessary...

2. The same cannot be said of the airborne aerial gains  $g_r$  and  $g_t$ ...



Подведение формул, не содержащих предикативных знаков, под класс существительных приводит к необходимости рассмотрения сочетаний «формульного» существительного ( $S_\phi$ ) с обычным.<sup>6</sup> Дистрибутивная модель SS, как известно, неоднозначна: ей может соответствовать как структура  $S \overline{S}$  (определение—определяемое), так и структура  $S \overline{\overline{S}}$  (существительное—приложение).

Исследование показало, что неоднозначность дистрибутивной модели SS разрешается в тех случаях, когда одним из S является  $S_\phi$  при учете позиции  $S_\phi$  относительно S. Иначе говоря, характеризуя дистрибутивную модель в терминах подклассов, мы вместо одной неоднозначной модели SS получаем две однозначные модели  $S_\phi S$  и  $SS_\phi$ . В первой из них реализуется атрибутивное отношение (структура  $S_\phi \overline{S}$ ), во второй — аппозитивное (структура  $\overline{\overline{S}} S_\phi$ ).

Например:

1. ...used by the Germans in the V-2 program.
2. ...target at range  $R_{max}$  with velocity  $V_{max}$ .

В случае фразообразующих формул проблема состоит в том, можно ли их вычленить как самостоятельное предложение или же они выступают как единицы более низкого синтаксического уровня. Перечислим случаи, когда сегментация предложения с формулой, содержащей предикативный знак, на два предложения, анализируемые отдельно, невозможна.

Одним из таких случаев является употребление формулы после глагола to let. В этом случае формула подводится под класс нефразообразующих, т. е. выступает в составе предложения в качестве отдельного члена. После формулы в этом случае может употребляться инфинитив. В первом случае формула подводится под класс существительных, так как выступает в функции дополнения. Во втором случае приравнивание этой формулы к существительному приводит к неправильному анализу, так как не реализуется обязательная валентность глагола to let на инфинитив. Например:

1. Let  $f_1 = f_p$  and  $f_2 = f_q$  be obtained from...
2. If we let  $Q = q/(1-q)$  this reduces to  $F = a/p_1$ .

<sup>6</sup> Выделение особого подтипа «формульного» существительного необходимо для того, чтобы на этапе синтезирования не искать переводных эквивалентов в словаре.

Видимо, в грамматике следует по возможности учитывать такие условия. Например, глаголу *to let* можно задать в словаре валентность на фразообразующую формулу.

Аналогичны случаи употребления формул с глаголом *to set*<sup>7</sup>: Now, set  $k = a(1-e)^2 \dots$

Чистые фразообразующие формулы нельзя вычленивать как самостоятельное предложение, если они стоят перед или после финитной формы глагола (причем в этом случае они могут быть отделены от глагола только союзным словом или наречием). Приведем несколько примеров:

1. Here,  $f = T/2$  is the range... 2. The result is (then)  $f = T/2$ . 3. ... we have  $F = a/p_1$ . 4. ... where  $f = S$  has poles... 5. The radar requirement becomes  $F = a/p_1$ .

Фразообразующая формула не вычленяется как самостоятельное предложение и в случае, если она употребляется после предлога:

1. ... this leads to  $F = a/p_1$ . 2. ... (for  $f = T/2$ ). 3. ... is given by  $F = a/p_1$ .

Во всех этих случаях предикативный знак рассматриваем как нефинитно-формальное определение к существительному (левой части формулы).

Особую сложность представляет употребление после глагола *to be* смешанной формулы. Хотя смешанная формула и имеет статус предложения, выделить ее в качестве отдельного сегмента, анализируемого как самостоятельное предложение, в данном случае невозможно, так как глагол *to be* остается без насыщенной валентности. Приведем примеры:

1. ...number of targets... is...  
(Number of targets) =  $N_t$ .
2. The probability of ...  $j$  targets ... is then...  
Prob. [ $j$  targets] =  $P_1$ .

В этом случае поступаем так же, как и в случае употребления формулы с глаголом *to let*.

Смешанная формула может выступать как самостоятельное предложение, и тогда предикативный формульный знак

---

<sup>7</sup> Класс таких глаголов ограничен, в рассмотренных примерах их встретилось только два, но при дальнейшем исследовании этот класс глаголов может быть расширен.

приравнивается финитной форме глагола, левая часть формулы принимается за подлежащее, правая — за предикативный член (согласно стандартному правилу порядка слов в предложении). Примеры:

1.  $f$  = local microwave reference frequency.
2.  $P$  = crystal controlled modulation frequency.

Предварительные исследования показали, что для предложений с формулами в большинстве случаев достаточно подвести эту формулу под класс существительных. Но есть сложные случаи, такие, как употребление фразообразующих формул с глаголами типа *to let*, *to set*. Не всегда поддаются анализу смешанные формулы (когда они выступают как отдельный член предложения). В этих случаях предикативный знак может быть употреблен в качестве финитной и нефинитной форм глагола, эквивалентной либо инфинитиву (в случае с глаголом *to let*), либо причастию.

Особую проблему представляют собой случаи стыка двух формул<sup>8</sup>. Однако в нашем материале таких примеров не оказалось.

Н. Л. Кириченко

### ПРЕДИКАТ ТОЖДЕСТВА И ЯЗЫКОВЫЕ СПОСОБЫ ЕГО ВЫРАЖЕНИЯ

Одной из основных задач изучения семантики для построения информационной системы «запрос—ответ» является составление различных правил преобразования текста. Под семантикой текста понимается его содержательная сторона, т. е. та информация, для сообщения которой текст был написан и которая необходима для того, чтобы информационная система могла выдать ответ на любой вопрос по тексту. Поэтому представляется особенно интересным исследование смысловых связей, существующих между отдельными элементами текста, а именно: тождества, включения и др.

<sup>8</sup> См.: С. А. Гюльмисарян, Ф. А. Дрейзин, З. Т. Тер-Микакянц, ук. ст. НТИ, сер. 2. Информационные процессы и системы, 1971, № 3, стр. 37.

Данная статья посвящена рассмотрению отношений тождества. Задача ее — показать имеющиеся в языке формальные способы выражения отношения тождества между двумя элементами в одном предложении.

Два элемента текста, связанные отношением тождества, являются синонимичными. Это значит, что они обозначают одно и то же понятие и способны заменить друг друга (или их можно подставить один вместо другого) в одном тексте.<sup>1</sup> Например: «Старинное село Никольское на острове Беринга переживает вторую молодость. На живописной террасе у подножия сопки отчетливо вырисовываются контуры нового двухэтажного поселка. Местные жители алеуты называют его островным городом. Больница, школа-интернат, гостиница, десятки благоустроенных домов, сооруженных за последние годы, придали новый облик центру Алеутского района».<sup>2</sup>

Как видно из приведенного отрывка, «старинное село Никольское» (элемент текста) получает в последующих предложениях другие синонимичные выражения — именуется новым двухэтажным «поселком», «островным городом», «центром Алеутского района», т. е. из данного отрывка можно сделать следующие выводы:

- 1) старинное село Никольское — новый двухэтажный поселок;
- 2) старинное село Никольское называют островным городом;
- 3) старинное село Никольское — центр Алеутского района.

Три словосочетания, стоящие в правой части, связаны отношением тождества с левой частью и являются синонимами словосочетания «старинное село Никольское», так как обозначают один и тот же объект и могут заменять друг друга.

<sup>1</sup> Существует целый ряд определений синонимичности. См., например: А. И. Уемов. Проблема синонимов и современная логика. В сб.: Логико-грамматические очерки. М., 1961, стр. 26—48; Ю. Д. Апресян. Проблема синонимии. ВЯ, 1957, № 6, стр. 88; М. Ф. Палевская. Проблема синонимического ряда, его границ и возможности выделения доминанты. В сб.: Лексическая синонимия. М., 1967, стр. 97; В. Г. Адмони и Г. И. Сильман. Отбор языковых средств и вопросы стиля. ВЯ, 1954, № 4, стр. 93 и др.

<sup>2</sup> Примеры, приведенные в статье, взяты из журналов: «Наука и жизнь», «Знание—сила» за 1966—1970 гг., из книг: Генетика микробов, Радиоастрономия. М., 1955; Психология. М., 1945; Курс высшей математики. М., 1956, а также из газеты «Ленинградская правда», 1970, август.

Отметим, что в двух первых предложениях отношения тождества — чисто контекстуальные.

Таким образом, для нас существенно то, что какие-то два элемента текста (обозначим их условно А и В) используются автором как эквивалентные выражения. Одним из способов, устанавливающих эквивалентность отдельных элементов в тексте, является предикат тождества, связывающий два элемента в одном предложении. Предикат тождества — это предикат, формирующий класс синонимов.

Способы выражения отношения тождества могут быть самыми различными.

1. Наиболее ярким и частым показателем отношения тождества в языке выступает глагол «называться». Например:

1. Ветровую эрозию почвы **называют** еще дефляцией, что в переводе с латинского означает «выдувание».

2. ...неосознанный компонент интеллектуального творчества **называют** часто интуицией.

Наряду с «называться» употребителен синонимичный ему глагол «именоваться». Например:

3. Заметим, что процессы биосинтеза нуклеиновых кислот **именуют** дубликацией, или редубликацией, и транскрипцией.

4. ...все разновидности появившихся сейчас так называемых «подводных аппаратов»: мезоскафы, «ныряющие блюдца» и т. д. — **обладают** перечисленными выше качествами и поэтому во избежание путаницы в терминологии должны **именоваться** традиционно — подводные лодки.

Глагол «называться» может стоять в форме причастия, либо в словосочетаниях «известный под названием», «получить (носить) название (наименование)» и т. д.<sup>3</sup> Например:

5. Мутацию вызывают изменения в химической структуре хотя бы одной пары оснований. Изменение в одной паре — это минимальная единица мутации, **названная** учеными мутоном.

<sup>3</sup> Сочетание «так называемый» обычно употребляется как определитель и стоит перед объектом в предложениях типа определяемое—определяющее. Например: «...перфузию лучше проводить так называемой „искусственной кровью“ — жидкостью, переносящей кислород... Поток радиоизлучения от Солнца часто характеризуется так называемой эффективной температурой  $T_{эфф}$ . Это — температура такого воображаемого раскаленного тела, имеющего угловые размеры энергии, как и реальное Солнце».

6. Единица рекомбинации, размер которой также не может быть меньше одной пары оснований, **получила наименование «рекон».**

7. В переходной полосе от леса к степи, в так **называемой** лесостепи, встречаются отдельные массивы леса.

II. Помимо глагола «называться» показателями отношения тождества могут выступать следующие глаголы: «отождествлять», «обозначать», «считаться», «понимать», «представлять собой». Например:

8. ...пульсары по-прежнему **отождествляют** с нейтронными звездами определенного типа...

9. ...молекулы жиров **представляют собой** укрупненные образования, состоящие из трех молекул жирных кислот и одной молекулы глицерина.

III. Языковыми способами выражения тождества, показателями связи двух эквивалентных форм выражения в отношении одного и того же предмета объективной действительности, могут служить также союзы, союзные слова, уточняющие обороты («или», «то есть», «иначе говоря», «как говорят», «а именно», «точнее», «проще», «по-научному», «означать по сути дела» и т. п.). Например:

10. Бруцеллез, **или** мальтийская лихорадка, передается человеку через травоядных.

11. ...космический корабль должен иметь свою экологическую систему, **то есть** целый миниатюрный замкнутый мир живых существ, имитирующий земные условия внутри корабля.

12. Так появились деньги — всеобщий измеритель, или **по-научному** — всеобщий эквивалент.

IV. Отношение тождества в предложении может быть выражено через пояснение в скобках или приложения, если в предложении есть и другие предикаты. Эквивалентное выражение понятия дается автором кратко и как бы мимоходом, в виде одного слова — термина или краткого определения. Например:

13. Конрад Лоренц — медик, зоолог и философ по образованию — по праву считается одним из основоположников **этологии, науки о поведении животных.**

14. Ценным дополнительным методом исследования пищевода и желудка обещает стать **эндоскопия — осмотр изнутри.**

15. Живой организм обладает чудесным и могущественным механизмом иммунитета (невосприимчивостью к болезням).

V. Отношения тождества между объектами внутри одного предложения имеют свои особенности в предложениях с однородными членами. В таких предложениях, которые могут, в частности, выражать и родо-видовые отношения, в одной части стоит общее родовое понятие, а в другой — дается перечисление объектов, составляющих это родовое понятие. В отличие от предложений с отношением включения в предложениях с отношением тождества родовое понятие раскрывается **полностью** перечислением **всех тех объектов**, которые указаны в другой части предложения. О предложениях же с включением этого сказать нельзя. Характер перечисления в этом случае таков, что мы не можем считать перечисление законченным, исчерпывающим и потому делаем два вывода, свойственных отношению включения: 1) А есть один из В; 2) существуют и другие В. Например:

16. Необходимость в строительных материалах — цементе, глине, песке, мраморе, кирпиче, гипсе и т. д. — особенно выросла после Великой Отечественной войны...

17. В курсе химии для средней школы большое внимание уделяется гидролизу органических соединений — сложных эфиров, жиров, углеводов и белков.

Сравним с точки зрения выводов предложения 16 и 17. Из предложения 16 будет следовать:

а) цемент, глина, песок, мрамор, кирпич, гипс и т. д. — виды строительных материалов (или цемент — один из строительных материалов, и т. п. в отношении каждого из перечисленных строительных материалов), т. е. А есть один из В;

б) существуют и другие виды строительных материалов (перечисление явно не закончено), т. е. есть и другие В.

С точки зрения выводов предложение 17 можно представить в следующем виде:

В курсе химии для средней школы большое внимание уделяется гидролизу органических соединений (А) — сложных эфиров, жиров, углеводов и белков (В), т. е. органические соединения (А) — это сложные эфиры, жиры, углеводы и белки (В), т. е. А=В. Таким образом, в предложении 16 налицо отношение включения между объектами (что подтверждается и выводами), а в предложении 17 между словосочетанием «органические соединения» и перечисленными

объектами правой части предложения существует отношение тождества.

Ряд слов в предложениях с перечислением объектов указывает на законченность перечисления и таким образом как бы служит показателем тождества. К таким словам относятся: союзы «и», «а», «также», «а затем» (в значении «и»), «или», заканчивающие перечисление объектов; слова: «все», «следующие»; указательные частицы: «вот», «то», «те», «таковы» и числительные: «два», «три», «пять» и пр., стоящие перед общими родовыми понятиями (отрасли, науки, продукция, районы, методы, зоны, качества, проблемы, аспекты, черты, достоинства, свойства, фазы и т. д.). Например:

18. ...мастики приготавливаются на основе термореактивных смол — фурановых, эпоксидных **или** полиэфирных...

19. До сих пор техника знала только **три** вида абразивных материалов, из которых получают круги высокого качества: алмаз, корунд и карбид кремния.

20. К числу линий второго порядка принадлежат окружность и **следующие кривые**: эллипс, гипербола и парабола.

21. Набухание ядра, проникновение в него белков из цитоплазмы, прекращение синтеза РНК и начало синтеза ДНК — **таковы** твердо установленные фазы поведения ядра после пересадки его в яйцеклетку.

VI. Как в сочетании с глаголами «определять», «рассматривать», «представлять», «использовать» и в сочетании «есть не что иное, как» служит показателем отношения тождества между объектами внутри одного предложения. Например:

22. Словарь современного русского литературного языка **определяет** акрибию **как** тщательность, точность в научных исследованиях.

23. Яркостная температура **определяется как** температура воображаемого раскаленного тела...

24. ...опыт **есть не что иное, как** восприятие окружающей действительности с помощью органов чувств.

Здесь интересно сравнить слово **как**, выступающее показателем отношения тождества, с **как** — показателем отношения включения. В предложениях с отношением включения слово **как** можно встретить в том случае, если есть слова, обозначающие множество объектов, а также дается перечисление единичных предметов, составляющих это множество.



25. ...широколиственные породы деревьев, как дуб, клен, вяз, часто встречаются в лесах южного Урала и в Предуралье,

В предложениях с отношением тождества между объектами внутри одного предложения слово **как** служит показателем определения одного объекта через другой объект, т. е. слово **как** с вышеперечисленными глаголами соединяет (сравнивает) две эквивалентные формы выражения. С точки зрения выводов предложения 24 и 25 можно представить следующим образом:

а) ...опыт (А) есть не что иное, как **восприятие окружающей действительности с помощью органов чувств (В) → (А = В)**;

б) **восприятие окружающей действительности с помощью органов чувств (В) есть не что иное, как опыт (А) → (В = А)**, но:

а1) ...дуб, клен, вяз (А), часто встречаемые в лесах южного Урала и Предуралья, — **широколиственные породы деревьев (В) → (А есть один из В)**;

б1) существуют и другие виды широколиственных пород деревьев (В) → (существуют и другие В).

Наиболее часто отношения тождества между двумя элементами текста встречаются в связочных конструкциях. Показателем отношения тождества выступают связочные глаголы «являться», «есть», «суть», а также нулевая связка. Например:

26. Н. Е. Жуковский **является** основоположником учения об авиации, его часто называют «отцом русской авиации».

27. Очевидно, постоянные и бесконечно малые величины **суть** вместе с тем величины ограниченные.

28. ...окружность **есть** частный случай эллипса... это **есть** эллипс, оси которого равны друг другу.

В приведенных примерах функция глагола «быть» сводится к тому, что «быть» является уже показателем равенства. Как по своим структурным особенностям, так и по своему значению функция глагола «быть» — показателя равенства близка функции его — показателя тождественности. В математических текстах, где чаще всего можно встретить глагол «быть» в функции показателя равенства,

нередко «есть» дополняется (и заменяется) выражением «равно»<sup>4</sup>. Например:

29. Два и три **есть** пять.

30. ...произведение любого числа на нуль **равно** нулю...

В текстах научно-популярной и научно-технической литературы довольно часто (60%) встречаются предложения, в которых дается переобозначение, описание терминов, употребляемых в данной предметной области. Для этого типа предложений с термином и его описанием характерна связочная конструкция с нулевой связкой. Иногда термин берется в кавычки. Например:

31. Но если поместить в рану **«индуктор»** — **размельченную костную ткань**, — то из клеток соединительной ткани на этом месте в конце концов образуется кость.

32. Один траулер поздравил по радио друзей с Новым годом — и больше никто его не видал и не слышал: попал в **«Красный Ллойд»** — книгу **кораблей, пропавших без вести**.

Чаще же термин употребляется без кавычек. Например:

33. **Мутации** — **скачкообразные изменения гена**, которые, естественно, вызывают изменения определяемого этим геном признака и передаются по наследству.

34. **Притча** — **рассказ-наравоучение в иносказательной форме**.

В ряде случаев в правой части предложения, в той, где дается пояснение-описание, часто встречается местоименная частица «это», которая трактуется в русской академической грамматике как заменитель связки<sup>5</sup>.

35. Палеонтология доказывает, что история органического мира — **это** непрерывная смена различного вида организмов.

36. Листопадность — **это** приспособление растений в холодном климате к морозной зиме, а в жарком, например в саваннах, — к периоду зноя.

В предложениях с нулевой связкой отношения тождества могут сопровождаться повторением одного и того же слова и в правой и в левой частях; характеризуемое понятие здесь подчеркнуто выделяется каким-то своим признаком, отличающим его от других понятий данной предметной области.

<sup>4</sup> См.: Карел Берка. Функции глагола **быть** с точки зрения современной формальной логики. В сб.: Логико-грамматические очерки. М., 1961.

<sup>5</sup> Слова «это», «вот» в Академической грамматике трактуются как указательные, часто выполняющие роль предикативной связки. См.: Грамматика русского языка, т. II, ч. I, М.—Л., Изд. АН СССР, 1960, стр. 637—641.

Например:

37. Дальнейшие работы И. Г. Пидопличко (и некоторых других ученых) привели к созданию теории, отрицающей Великие четвертичные оледенения, — теории антигляциализма.

38. Систолический объем — объем крови, выбрасываемый сердцем за одно сокращение.

Передки в текстах научно-популярной литературы и так называемые окказиональные тождества. Под этим подразумевается такой тип тождественности двух различных выражений, когда определение того или иного понятия, предмета, объекта и т. д. дается через образное, метафорическое описание, вводимое автором для большей яркости и ясности понятия. Например:

39. ...все началось в мантии Земли — огромной физико-динамической лаборатории нашей планеты.

40. Факты, — говорил И. П. Павлов, — это воздух науки.

41. Пожалуй, не преувеличим, если скажем, что ньютоновская трактовка тяготения — это подлинная цитадель классической физики...

42. Особенно выделяются расположенные на берегах Томи (правый приток Оби) Томск — «Сибирская кузница научных кадров»<sup>6</sup>, Кемерово — центр добычи угля и химической промышленности в Кузбассе...

Отождествление двух элементов текста осуществляется путем выявления формальных способов выражения предиката тождества.

Следует отметить, однако, что на данном этапе нам не удалось представить однозначные способы выражения отношений тождества. Ряд приведенных способов (союзы, союзные слова, обороты, приложение, предлог как, связочные глаголы и нулевая связка) является омонимичным: он выражает и отношение включения (родо-видовые отношения). Составление правил снятия омонимии — задача дальнейшего исследования.

---

<sup>6</sup> Иногда такое образное, яркое описание-переобозначение, появившись в каком-либо контексте, становится как бы вторым наименованием предмета, объекта, понятия и т. д., и это новое рожденное выражение получает право употребления наравне со старым, внося дополнительный оттенок яркости, образности, и читающий или слушающий легко восстанавливает первоначальную тождественность образов, объектов, понятий и т. п., давших это переименование. Так, пушнину часто называют «мягким золотом», рыбу — «живым серебром», воду — «голубым углем» и т. д. и т. п.

## МОТИВАЦИЯ СОСТАВНЫХ МАТЕМАТИЧЕСКИХ ТЕРМИНОВ

В ситуации определения составного термина будем различать сам термин  $T$  и его определяющее описание  $D$ . Это определяющее описание вводит новый термин, делает его понятным на основании известных знаний. Мотивацией составного термина  $T$  мы будем называть:

1) возможность более или менее точного предсказания  $D$  на основании известных сведений о составляющих  $T$  и о типе синтаксической связи между этими составляющими;

2) возможность для данного  $D$  подобрать соответствующее  $T$ .

В данной статье будут рассматриваться некоторые случаи, связанные в основном с первым аспектом мотивации. При этом будут изучаться такие составные математические термины, составляющие которых находятся в атрибутивном или общегенитивном отношениях непосредственной доминации (ОНД)<sup>1</sup>. Составные выражения с этими ОНД обладают наиболее богатым набором способов мотивации.

Явление мотивации составных выражений позволяет объяснить продуктивность речевой деятельности человека<sup>2</sup>. Именно благодаря мотивации, на основании ограниченного опыта человек понимает новые, никогда ранее не слышанные составные выражения, а также для каждого нового значительного явления он имеет возможность подобрать подходящее название.

Таким образом, мы будем искать способы предсказания определяющего описания на основании известных сведений. Эти способы будут выявляться исходя из предварительного рассмотрения и сопоставления наборов терминов и определяющих описаний для них. При этом будет уточняться вид сведений, необходимый для функционирования искомым способом мотивации.

Среди способов нахождения определяющих описаний можно выделить относительно простые способы предсказа-

<sup>1</sup> И. А. Мельчук. Автоматический синтаксический анализ. М., 1966, стр. 19.

<sup>2</sup> A. Miller. Language and Psychology. New-Directions in the Study of Language. Mass., 1964.

ния, допускающие относительно простое общее описание.<sup>3</sup> Рассмотрим пример.

«Назовем эйлеровой цепью (соответственно эйлеровым циклом) цепь (соответственно цикл), содержащую все ребра графа и притом только по одному разу»<sup>4</sup>.

Здесь, по существу, два определения для двух составных терминов: (а) «эйлерова цепь» и (б) «эйлеров цикл». При чем определение для одного из них, например для эйлеровского цикла, легко может быть получено из определения для эйлеровой цепи заменой цепи на цикл. Иначе говоря, мы можем здесь сопоставить слову «эйлеров» следующую структуру: эйлеров  $X$  есть  $X$ , содержащий все ребра графа и притом по одному разу. Под  $X$  здесь подразумевается элемент определенного класса слов. Класс этот может быть определен набором признаков, а в случае небольшого числа элементов — задан просто списком.

Если с одной из составляющих термина соотнесена такого рода структура, а другая входит в класс слов, для которого эта структура справедлива, то нахождение определяющего описания для составного термина сводится к простому заполнению тех мест структуры, где располагаются переменные.

Не всегда эту связующую структуру выяснить так легко, и не всегда она так проста, как в данном случае. Некоторые тонкости появляются, когда слово входит в определяющее описание неявно, через своего «представителя», находящегося с представляемым словом в каком-либо очевидном отношении. Вполне понятно, что чем шире класс слов, для которого справедлива структура, сопоставляемая другому слову, тем экономнее такое структурное представление по сравнению с множеством частных определений. Однако такое расширение применимости структуры часто осуществляется за счет некоторых обобщений в ней; причем эти обобщения делают и все предсказание ориентирующе обобщенным. Например, слову «дискриминант» можно сопоставить следующую структуру:

«Дискриминант  $X$  это такое математическое выражение, составленное из коэффициентов  $X$ , знак которого во многом

<sup>3</sup> См.: М. А. Козаринский. Некоторые мотивированные связи сложных терминов. В сб.: Материалы семинара по проблеме мотивированности языкового знака. Л., 1969, стр. 32.

<sup>4</sup> К. Берж. Теория графов и ее применение. М., 1962, стр. 179.

определяет характер решения  $X$ »<sup>5</sup> (здесь класс  $X$  — квадратное уравнение, кубическое уравнение, дифференциальное уравнение).

Рассмотрим теперь более сложный способ предсказания определяющего описания. К такому способу приходится прибегать, когда нет возможности опереться на простые мотивированные связи, некоторые примеры которых мы рассмотрели ранее. Сложный способ должен быть более универсален. Но расширение предсказующих возможностей потребует задания исходных сведений в особой форме. Иначе говоря, для того чтобы какой-либо автомат мог бы приблизиться к предсказующим возможностям «понимающего» человека, этот автомат должен исходить из сведений того же типа, что и знания человека. Мы имеем в виду здесь главным образом следующие исходные положения.

I. Некоторые слова не изолированы, а как бы окружены некоей сферой других слов и выражений (своеобразным ассоциативным полем), отношение которых к первичному слову известно, т. е. может быть легко восстановлено. (Существование ассоциативных полей неоднократно подчеркивал Фердинанд де Соссюр.) Мы будем символически изображать переход от слова **а** к слову **в** из семейства, окружающего **а**, стрелкой: «**а** → **в**» и будем говорить: «**а** порождает **в**».

II. Существуют классы наборов слов (в частности, пар слов), таких, что если предъявить этот набор, то можно получить в ответ отношение, связывающее слова этого набора.

III. В процессе работы «понимающего» механизма каждое из этих множеств из I и II может пополняться за счет изменения знаний.

Эти положения являются исходными, аксиоматическими для нашей модели. Вообще говоря, понятие «ассоциативного поля», его вид и способы обнаружения — проблемы отнюдь не простые и решаются они в различных областях знания по-разному. Однако для нашей модели сейчас эти вопросы являются не принципиальными, чисто техническими.

В приводимых далее примерах мы считали, что в порождаемое семейство входит все то, что легко вспоминается в связи с первичным выражением.

Поясним еще положение II. Замечено, что кроме «ассоциативных полей» для данных первичных слов существуют

<sup>5</sup> См.: И. Н. Бронштейн, К. А. Семдяев. Справочник по математике. М., 1954, стр. 476.

еще классы других слов, которые обладают следующими свойствами. Если задать пару отдельных слов, одно из которых первичное, то по этой паре, как по двойному адресу, можно легко вспомнить связывающее эту пару слов отношение. И в то же время первичное слово может и не породить второе слово этой пары, т. е., по-видимому, в мозгу человека закрепляются не только статистически связанные объекты, но и повторяющиеся отношения.

После этих предварительных замечаний, перейдем теперь к изложению сложного способа предсказания определяющих описаний, основанного на идее алгоритма встречных волн.<sup>6</sup>

Итак, пусть предъявлен некоторый составной термин  $\alpha\beta_0$ , где  $\alpha_0$  и  $\beta_0$  — обозначения его составляющих, и требуется найти определяющее описание для этого термина.

В соответствии с положением I о порождаемых семействах для каждого из выражений  $\alpha_0$  и  $\beta_0$  можно получить соответствующие семейства выражений  $A_1$  и  $B_1$ , где  $A_1$  и  $B_1$  — обозначения семейств (классов, слоев), создаваемых в результате первого шага порождения. Каждое выражение из слоя  $A_1$  может породить свое семейство, совокупность которых для всех выражений из  $A_1$  даст нам второй слой выражений  $A_2$ . Аналогично рассуждая, можно получать слои  $A_j$  из  $A_{j-1}$  и  $B_j$  из  $B_{j-1}$ . Наша дальнейшая задача состоит в установлении каких-то типов связей между элементами из слоя типа А и слоя типа В.

Пусть способы установления таких связей уже известны. Тогда суть алгоритма встречных волн состоит в том, чтобы двигаться (в смысле порождения новых слоев) навстречу друг другу поочередно со стороны  $\alpha_0$  и  $\beta_0$ , пока не обнаружится связь между каким-либо элементом из слоя типа А и каким-либо элементом из слоя типа В.

Здесь будут рассмотрены только два типа связи: 1) обнаружение тождества и 2) обнаружение отношения. Это значит, что при сопоставлении элементов встречных слоев (один элемент из слоя типа А, а другой — из слоя типа В) по отношению к этим элементам могут быть заданы два вопроса:

1) Совпадают ли они? 2) Известно ли из прошлого опыта отношение между ними?

<sup>6</sup> См.: М. А. Козаринский. Алгоритм встречных волн. В сб.: Методы вычислений. Вып. II. Изд. ЛГУ, 1963, стр. 132—139.

В случае положительного ответа на один из этих вопросов связь между сопоставляемыми элементами считается установленной. И тогда по стрелкам, за которыми стоят известные отношения, можно восстановить всю семантическую конструкцию, связывающую исходные слова или выражения  $\alpha$  и  $\beta$ , что, собственно говоря, и является нахождением определяющего описания для составного термина. (Здесь только нужно еще свернуть цепочку найденных отношений в связную конструкцию соответствующего определяющего описания.)

Перейдем теперь к рассмотрению примеров, иллюстрирующих данный сложный способ предсказания. Отметим только, что в примерах будут указаны лишь цепи рассуждений, приведшие к цели, хотя во всем рассуждении присутствуют, конечно, в гораздо большем количестве цепи, не приводящие к положительному результату сопоставления. Некоторые из этих боковых цепей могут быть устранены в результате специальных приемов, направленных на преодоление неоднозначностей в рассуждении.

### Примеры на обнаружение тождества

Пусть предъявлен термин:

#### 1. Полярная ось.

Проведем следующее рассуждение.

1.1. Полярный  $\rightarrow$  полюс  $\rightarrow$  точка.

1.2. Ось  $\rightarrow$  прямая  $\rightarrow$  точка.

1.3. Точка (1.1.)  $\equiv$  точка (1.2.).

Восстановим стрелки. Точка из (1.1.) связана с полюсом через отношение типа: полюс есть точка. А точка из (1.2.) связана с прямой отношением типа: точка принадлежит прямой или прямая проходит через точку. В силу тождества (1.3) получаем, что полярная ось — это прямая, проходящая через точку, которая в то же время и есть полюс.

Тем самым мы довольно точно, хотя и не полностью, предсказали определяющее описание для полярной оси. В действительное определение входит то, что полярная ось — заданная прямая, проходящая через полюс<sup>7</sup>.

---

<sup>7</sup> И. Н. Бронштейн, К. А. Семендяев. Справочник по математике, стр. 199.



В дальнейшем для краткости мы не будем восстанавливать отношения, стоящие за стрелками, полагая, что при необходимости это всегда можно сделать без труда. Еще пример.

## 2. Закон ассоциативности.

Рассуждение.

2.1. Закон  $\rightarrow$  свойство объектов  $\rightarrow$  свойство.

2.2. Ассоциативность  $\rightarrow$  свойство.

2.3. Свойство (2.1.)  $\equiv$  свойство (2.2.).

Следует подчеркнуть, что предсказывается не полностью определение для составного термина, а то в определении, что может быть извлечено из самих слов термина.

Касаясь всей проблемы проверки на тождество элементов, следует еще оговориться, что могут встретиться конфликтные ситуации. Например, известно заранее, что сопоставляемые элементы обозначают различные явления (не совпадают по смыслу). В этом случае предполагаемое тождество аннулируется.

## Примеры на обнаружение отношения

### 1. Кривая распределения.

Рассуждение.

1.1. Кривая  $\rightarrow$  график функции  $\rightarrow$  функция.

1.2. Функция, распределение?

Здесь вопросительное выражение (1.2) означает, что в память посылается запрос, известно ли отношение между сопоставляемыми элементами? И оттуда на основании положения II может быть получен ответ в виде:

1.3. Функция распределения.

Сравните: «Кривая распределения — графическое изображение функции распределения»<sup>8</sup>.

### 2. Угловой коэффициент.

Рассуждение.

2.1. Угол  $\rightarrow$  прямые под углом  $\rightarrow$  прямая.

2.2. Коэффициент  $\rightarrow$  коэффициент в уравнении  $\rightarrow$  уравнение.

2.3. Прямая, уравнение?

2.4. Уравнение прямой.

<sup>8</sup> О. В. Мантуров, Ю. К. Солнцев, Ю. И. Соркин, Н. Г. Федин. Толковый словарь математических терминов. М., 1955, стр. 176.

После восстановления отношений, стоящих за стрелками, сравните полученное описание со следующим определением: «Угловой коэффициент прямой на плоскости относительно декартовой прямоугольной системы координат — тангенс угла между этой прямой и положительным направлением оси  $X$  данной системы координат»<sup>9</sup>.

Мы уже говорили о том, что порождаемые «ассоциативные поля», по-видимому, не слишком велики. Во всяком случае, они меньше класса элементов, отношение которых к первичному известно. Одна из причин ограниченности «ассоциативных полей», возможно, кроется в ограниченности оперативной памяти человека. Порождать слишком большие слои человек, очевидно, не смог бы. Иначе ему пришлось бы при сопоставлении элементов из разных слоев удерживать слишком большое количество элементов в оперативной памяти (пока сопоставление элементов встречных слоев не окончено). Поэтому человек не идет по пути расширения класса порождаемых элементов, а разнообразит способы их сопоставления. Это мы и пытались здесь хотя бы частично отразить.

Итак, мы показали некоторые способы предсказания определяющих описаний для составных терминов. Такого рода мотивация позволяет объяснить (промоделировать) психологический феномен понимания ранее никогда не слышанных сочетаний. Тем самым объясняется продуктивность речевой деятельности человека. Именно благодаря мотивации становится понятным, как на основании конечного числа сведений об элементах и связях в языке достигается понимание в принципе бесконечного числа фраз.

Н. Д. Кремнева

## СЛОВОИЗМЕНЕНИЕ И СЛОВООБРАЗОВАНИЕ С ПОЗИЦИИ АВТОМАТИЧЕСКОГО АНАЛИЗА

При автоматической обработке словоформ традиционное разграничение словоизменяемых и словообразовательных

<sup>9</sup> Там же, стр. 468.

аффиксов, как и вообще разграничение лексических и грамматических элементов слова, принципиального значения не имеет. Всякая часть слова (будь то приставка, корень, суффикс или окончание) рассматривается как носитель определенной информации. Решающую роль играет фактор «регулярности—нерегулярности» этой информации. Примером регулярного словоизменения является склонение имен прилагательных, нерегулярного — образование супплетивных форм сравнительной и превосходной степеней сравнения. Пример регулярного словообразования — употребление увеличительного суффикса **-ищ** с основами имен существительных, нерегулярного — употребление того же суффикса **-ищ** в слове «жилище». Следовательно, рассматриваемый в таком плане процесс порождения словоформы имеет два пути. В одном случае две части слова сочетаются на основе совпадения индивидуального признака и общее значение сочетания выходит за рамки их частных, «самостоятельных» значений. Например, выделение в словоформе «безобразие» приставки **без-** и основы **-образ-** так же лишено смысла, как и выделение в словоформе «железо» частей **жел-** и **-ез(о)**. В другом случае две части слова сочетаются на основе совпадения группового признака<sup>1</sup> и общее значение сочетания является пересечением их частных, «самостоятельных» значений. Очевидно, что сочетания первого типа в отличие от второго не поддаются алгоритмической обработке и должны содержаться на входе алгоритма в виде неделимого целого. Именно поэтому фактору «регулярного-нерегулярного» порождения словоформы придается решающее значение при решении задачи автоматической переработки информации.

Рассмотрим ряд примеров регулярного порождения словоформы.<sup>2</sup> Словоформа «звезды».

Значения основы **звезд-** 1) небесное тело, 2) геометрическая фигура имеют одинаковые групповые признаки: substantiveная парадигма, I склон., твердый вариант, возможность образования мн. ч.

---

<sup>1</sup> Условимся считать групповым такой признак, который обеспечивает одной части слова равнозначную сочетаемость, по меньшей мере с двумя другими частями.

<sup>2</sup> Подробно о процессе регулярного порождения русских словоформ см. в статье Г. С. Ульяновой «Построение алгоритма морфологического анализа для русского языка на основе конечноавтоматной модели» (наст. сб., стр. 145).

### Значения аффикса -ы:

1) им. пад. мн. ч. (субстантивная парадигма, II склон., мужской род, твердый вариант, одушевленность, возможность образования мн. ч. на -ы); 2) им. и вин. пад. мн. ч. (субстантивная парадигма, II склон., мужской род, твердый вариант, неодушевленность, возможность образования мн. ч. на -ы); 3) родит. пад. ед. ч. (субстантивная парадигма, I склон., твердый вариант, невозможность образования мн. ч.); 4) родит. пад. ед. ч., им. и вин. пад. мн. ч. (субстантивная парадигма, I склон., твердый вариант, возможность образования мн. ч.); 5) им. и вин. пад. мн. ч. (мужской, средний, женский род (адъективная парадигма), усеченное окончание (по типу «папины», «отцовы»), твердый вариант); 6) краткость мн. ч. (адъективная парадигма, мужской, женский, средний род, твердый вариант).

Результирующая информация словоформы «звезды»: небесное тело или геометрическая фигура, им. или вин. пад. мн. ч. или родит. пад. ед. ч.

Словоформа «полы».

Значения основы пол-:

1) настил (субстантивная парадигма, II склон., мужской род, твердый вариант, неодушевленность, возможность образования мн. ч. на -ы); 2) биологический признак (субстантивная парадигма, II склон., твердый вариант, мужской род, неодушевленность, невозможность образования мн. ч. на -ы); 3) равнина; 4) возделываемый участок; 5) пространство; 6) фон; 7) край (для последних пяти значений групповые признаки одинаковы: субстантивная парадигма, II склон., средний род, мягкий вариант); 8) нижний край одежды (субстантивная парадигма, I склон., твердый вариант); 9) пустой (адъективная парадигма, твердый вариант).

Результирующая информация словоформы «полы»: настил, им. или вин. пад. мн. ч. или нижний край одежды им. или вин. пад. мн. ч. или родит. пад. ед. ч.

Словоформа «миров(ой)».

Значения основы мир-:

1) вселенная; 2) часть вселенной; 3) Земля; 4) население Земли (все четыре значения имеют признаки: субстантивная парадигма, возможность адъективной модификации на -ов); 5) общество; 6) область явлений (оба значения имеют признаки: субстантивная парадигма, невозможность адъективной модификации); 7) светская жизнь; 8) сельская община.

(оба значения имеют признаки: субстантивная парадигма, возможность адъективной модификации на **-ск**); 9) отсутствие вражды; 10) соглашение о прекращении вражды. (оба значения имеют признаки: субстантивная парадигма, возможность адъективной модификации на **-н**).

Значения аффикса **-ов-**:

1) родит. пад. мн. ч. (субстантивная парадигма, II склон., неодушевленность, возможность образования мн. ч. на **-ы**, возможность образования родит. пад. мн. ч. на **-ов**, признак конечного аффикса словоформы); 2) родит. и вин. пад. мн. ч. (субстантивная парадигма, II склон., одушевленность, возможность образования мн. ч. на **-ы**, возможность образования родит. пад. мн. ч. на **-ов**, признак конечного аффикса); 3) адъективная парадигма, усеченное окончание, твердый вариант (субстантивная парадигма, II склон., одушевленность); 4) адъективная парадигма, твердый вариант (субстантивная парадигма, возможность адъективной модификации на **-ов**).

Результирующая информация словоформы «миров(ой)»: вселенная, часть вселенной, Земля, население Земли, адъективная парадигма, твердый вариант.

Таким образом, признак невозможности адъективной модификации препятствует включению в результирующую информацию пятого и шестого значений основы **мир-**, а наличие дифференцированного признака адъективной модификации — включению значений 7—10.

Аналогичная картина наблюдается при анализе словоформ «крупозный», «крупяной» и «крупный». Признак невозможности адъективной модификации у основы **круп-** в значении 'туловище лошади' воспрещает учесть это значение во всех этих трех случаях. Дифференцированные признаки адъективной модификации данной основы в значениях 'острое воспаление дыхательных органов' и 'зерно', во-первых, обеспечивают полное разрешение многозначности основы, а во-вторых, предупреждают ложную интерпретацию словоформы «крупный».

Иначе обстоит дело с основами **асфальт**, **энтузиазм** и **вет(е)р**. Поскольку во всех этих трех случаях не возникает ни проблемы многозначности, ни проблемы омонимии, нет необходимости не только в дифференциации признака, но и в самом признаке возможности адъективной модификации (см. словоформы «асфальтовый», «ветренный», «ветряной» и «ветровой»). Оттенками значений словоформ «ветренный»,

«ветряной» и «ветровой» можно пренебречь, так как они связаны не со значением основы **вет(е)р** ('движение воздуха'), а со значениями слов, определяемых с помощью этой основы. Определяемое слово и является главным «конкретизатором» значений основы, его определяющей (ср. англ. *wind-glass* 'ветровое стекло' и *windmill* 'ветряная мельница', франц. *moteur à vent* 'ветряной двигатель' и *glace pare-vent* 'ветровое стекло').

Все вышеизложенное касалось «регулярности» образования словоформ. Теперь остановимся на вопросах, связанных с конечной результирующей информацией.<sup>3</sup>

Если с точки зрения способа автоматической обработки необходимо различать регулярное-нерегулярное образование словоформы, то с точки зрения конечного результата автоматической обработки целесообразно разграничить образования «содержательное» и «процедурное».

К первому («содержательному») типу образования словоформы относятся, например:

1) окончания основ имени существительного — показатели множественного числа и типа синтаксической связи;

2) приставки, суффиксы и окончания глагольных основ — показатели вида и времени;

3) суффиксы слов, образованных от глагольных основ, **-тель, -чик, -щик** — показатели «субъекта» («исполнитель», «резчик», «мойщик»);

4) суффиксы слов, образованных от глагольных основ, **-лив-, -чив-, -л-, -тельн-** — показатели «постоянного качества» («терпеливый», «возмутительный», «заносчивый», «гордый», «усталый»);

5) суффикс **-е-** основ имени прилагательного — показатель восприятия признака («белеть», «молодеть»);

6) окончания II спряжения глагола у слов, образованных от основ имени прилагательного, — показатели передачи признака («белить», «молодить»);

7) приставки глагольных основ — показатели начала действия («поплыть», «запеть») и т. д.

---

<sup>3</sup> Под оптимальным вариантом конечной результирующей информации понимается вариант, обеспечивающий наилучшее выполнение конкретной практической задачи. В дальнейшем автор ориентируется на описание предложения в виде дерева зависимостей как на результат первого этапа автоматической обработки текста. Совершенно очевидно, что при другой ориентации (например, на смысловую структуру предложения) конечная результирующая информация существенно изменится.

Ко второму («процедурному») типу образования словоформы относятся:

1) окончания основ имени прилагательного — показатели рода, числа, падежа;

2) суффиксы глагольных основ — показатели причастия, деепричастия, лица, числа, рода;

3) суффиксы слов, образованных от глагольных основ, **-ани(е), -ени(е), -к(а), -ти(е), -аци(я), -ств(о), -тельств(о)** — показатели субстантивной модификации («осуществление», «купание», «рубка», «закрытие», «вибрация», «строительство»);

4) суффиксы слов, образованных от основ имени прилагательного, **-ост(ь), -от(а), -ет(а), -ев(а), -ств(о), -изн(а)** — показатели субстантивной модификации («смелость», «теплота», «синева», «упрямство», «нищета», «желтизна»);

5) суффиксы слов, образованных от основ имени существительного, **-н-, -енн- (-ен-), -онн-, -ск-, -ов-, -ев-, -ьн-, -йн-, -ан-, -ян-, -озн-** — показатели адъективной модификации («схемный», «государственный», «локационный», «авансовый», «параллельный», «лучевой», «авторский», «нефтяной», «кожаный», «тифозный», «гарантийный») и т. д.

Итак, с позиций автоматического анализа даже больше оснований считать «атомный», «точность», «изготовление» формами от «атом», «точный», «изготавливать», чем общепринятые «атомы», «точнейший», «изготовил». Во всяком случае, в рамках поставленной практической задачи единственно правильным путем определения понятий «словоизменение» и «словообразование» представляется поиск определения функционального (как это все чаще признается в последнее время в отношении понятий «синонимия» и «омонимия»).

Л. Н. Смирнова

## ОБ ОДНОМ ТИПЕ СЛОВ-ЗАМЕСТИТЕЛЕЙ В СВЯЗНОМ ТЕКСТЕ

При описании поверхностной структуры связного текста, помимо рассмотрения синтаксических связей, сфера действия которых ограничена отдельным предложением, необходимо

описание и ряда других типов связей, действие которых выходит за пределы предложения. К их числу относятся связи анафорические, которые имеют место при повторном упоминании одного и того же объекта. При этом типе связи некоторый элемент текста (заместитель) указывает на другой элемент текста (антецедент). Объекты, обозначаемые антецедентом и заместителем, тождественны.

Антецедентом может быть имя существительное, именная или глагольная группа, предложение, фрагмент текста из нескольких предложений. Главным словом антецедента называем синтаксическую вершину антецедента. Если же антецедентом является фрагмент текста из нескольких предложений, то предполагается возможным представление синтаксической структуры связного текста в виде дерева зависимостей, и тогда можно говорить о синтаксической вершине некоторого фрагмента связного текста.

Заместителями в русском языке могут быть личные местоимения 3-го лица, возвратные местоимения («себя», «себе»), возвратное притяжательное местоимение («свой»), указательные местоимения («тот», «это»), местоименные наречия («здесь», «сюда», «там» и т. д.), относительные местоимения и наречия («который», «что», «где» и т. д.), имена существительные с идентифицирующими определениями и без них, субстантивированные прилагательные, причастия и числительные («последний», «сказанное», «вышеупомянутое» и т. п., «оба», «первый», «второй» и т. п.).

Предметом описания предлагаемой работы являются имена существительные с идентифицирующими определениями и без них в роли заместителей.<sup>1</sup>

Говоря о существительных с идентифицирующими определениями, мы имеем в виду следующие словосочетания: этот S, такой S, данный S, наш S, последний S, первый S, второй S, указанный S, упомянутый S и т. д. Идентифицирующими определениями могут быть указательные местоимения («этот», «такой», «тот»), притяжательное местоимение («наш»), числительные («оба», «первый», «второй» и т. п.), прилагательные и причастия («последний», «данный», «указанный», «упомянутый» и т. п.).

Функция перечисленных здесь определений состоит не в обозначении некоторого признака объекта, а лишь в иден-

---

<sup>1</sup> Материалом предлагаемого исследования являются тексты научно-го и научно-популярного характера.



тификации повторно упоминаемых объектов, в указании «адреса» предшествующего упоминания о том же объекте (последнее наиболее очевидно для заместителей «первый S», «второй S», «последний S»)².

Случаи употребления заместителей данного типа в зависимости от смыслового соотношения понятий, выражаемых главным словом antecedента и существительным в заместителе, можно разбить на четыре подтипа.<sup>3</sup>

**I. Существительное в заместителе совпадает с главным словом antecedента.** В заместителе может повторяться только главное слово antecedента:

«За последние два десятилетия созданы **автоматы**, которые выполняют сотни миллионов и даже миллиарды операций, прежде чем выдают результаты. Однако действие **этих автоматов** не последовательное»⁴.

Имя существительное без определений может рассматриваться как частный случай именной группы. Если такое имя существительное является antecedентом для некоторого заместителя, то главное слово antecedента и antecedент совпадают. Например:

«Если в **четырёхугольнике** диагонали точкой их пересечения делятся пополам, то **данный четырёхугольник** — параллелограмм».

«Котел питает паром **турбину**, **турбина** крутит генератор».

В ряде случаев в заместителе повторяется главное существительное antecedента с частью зависимых от него слов или вся именная группа antecedента.

---

² Из перечисленных выше заместителей, которые будут рассматриваться в данной работе, чаще других употребляется заместитель «этот S», поэтому в большинстве приводимых ниже примеров будет представлен именно этот заместитель.

³ Существительное в заместителе может иметь зависимые от него слова, в том числе и другие имена существительные. Говоря о существительном в заместителе, мы имеем в виду синтаксическую вершину заместителя.

⁴ В данной статье все примеры взяты из журнала «Наука и жизнь», 1964, № 4; 1965, № 3; 1966, № 6; 1967, № 2, 7, 10, а также из следующих книг: А. Тьюринг. Может ли машина мыслить? М., 1960; Л. И. Бутенмахер. Электронные информационно-логические машины. М., 1960; А. П. Киселев. Геометрия. М., 1952; А. В. Гладкий, И. А. Мельчук. Элементы математической лингвистики. М., 1969.

«В „игре в имитацию” принимают участие только **цифровые вычислительные машины**. В статье дается краткий обзор природы и свойств **этих вычислительных машин**».

В отдельных случаях повторение части слов, зависимых от главного слова антецедента, или всей именной группы необходимо для однозначной идентификации повторно упоминаемых объектов:

«В качестве примера можно сослаться на **информационный язык для элементарной геометрии, разработанный в ВИНТИ АН СССР**.

Любопытным примером семантического языка является построенный Г. Фройденталем „Линкос”, предназначенный для общения с внеземными цивилизациями.

Наконец, имеется ряд семантических языков, разработанных для целей автоматического перевода, в процессе которого семантический язык должен выступать в роли **языка-посредника**. Здесь мы упомянем семантические языки Кембриджского лингвистического кружка, а также язык „СМ-1” для представления математических текстов: в отличие от **только что упомянутого языка для геометрии, „СМ-1” должен служить для...»**.

Употребление заместителя «язык» или «этот язык» (т. е. повторение только главного слова антецедента) привело бы к тому, что в качестве антецедента был бы выбран «язык „СМ-1”».

**II. Существительное в заместителе является синтаксическим дериватом по отношению к главному слову антецедента (глаголу, причастию, прилагательному).** В данном случае речь идет о повторном упоминании некоторых фактов, ситуаций и т. п. Синтаксический дериват совпадает с главным словом антецедента по смыслу, но отличается от него синтаксически — относится к другой части речи. Например:

«При вероятностной адресации организация словаря **отличается** от ранее рассмотренных. **Это отличие** связано с характером образования сжатых кодов».

«Память тесно **связана** с мышлением. **Эта связь** доказывается тем, что в любом восприятии неизменно присутствует элемент суждения».

«Русская морфология заслуженно считается **очень сложной**. В предложенных правилах **эта сложность** гораздо более заметна».

«Подчиняясь командам с Земли, Луна-10 заблаговременно **развернулась** соплом двигателя по направлению на центр Луны. **Разворот** совершался относительно опорных направлений на Землю, Луну и Солнце».

III. **Существительное в заместителе является классификатором**<sup>5</sup> по отношению к главному слову антецедента — имени существительному или глаголу. Классификатор обозначает родовое понятие по отношению к понятию, обозначаемому главным словом антецедента — именем существительным. Например:

«География **ревматизма**, т. е. зависимость **этого заболевания** от климатических условий, в настоящее время подвергается пересмотру».

«Передвижения **леммингов** обычно бывают заметны лишь в определенных условиях, например на узком перешейке между озерами или у слияния двух рек. В таких местах можно увидеть большие скопления **зверьков**».

Подобные отношения возможны и в том случае, когда главным словом антецедента является глагол: классификатор обобщает, включает в некоторый класс действие, обозначаемое этим глаголом. Например:

«Если человек необходимо является самым совершенным существом, он может не **бояться** потерять свое доминирующее положение. Ясно, что популярность теологического возражения связана именно с **этим чувством**».

Повторное упоминание объекта может происходить путем употребления нейтральных существительных широкого обобщенного значения, не отражающих специфические свойства упоминаемого повторно объекта. Таковыми являются следующие заместители: «этот случай», «это обстоятельство», «это дело», «этот факт», «эта вещь» и т. п. Эти заместители близки указательному местоимению «это», и можно, по-видимому, говорить о прономинализации подобных заместителей:

«Пенопласты начинают теснить более традиционные материалы. **Это обстоятельство** повлекло за собой необходимость изучения различных свойств пены».

---

<sup>5</sup> Термин взят из работы: Е. В. Падучева. Выражение тождества упоминаемых объектов как одна из проблем синтеза языкового текста. Тр. III Всесоюзной конференции по информационно-поисковым системам и автоматизированной обработке научно-технической информации, т. II. М., 1967.

«Две разные грамматики порождают один и тот же язык. В этом случае грамматики называются эквивалентными».

С идентифицирующей функцией классификаторы могут совмещать оценочную функцию — передавать оценку повторно упоминаемого объекта:

«При оценке веса и размеров машины ЭНИАК мы должны помнить, что этот гигант используется для обработки 20 чисел в 10 десятичных знаков».

«У многих препарат вызывает изжогу, потерю аппетита, тошноту. С этими неприятностями приходится мириться».

«Больные не могут называть предъявленные им вещи и пытаются преодолеть этот недостаток, заменяя название описанием функции».

«Наше рассуждение окажется движущимся в замкнутом круге. Чтобы избежать этой опасности, мы приведем перечень тех средств, которые...»

При употреблении классификаторов выявление анафорических связей возможно лишь на основе некоторой информации, полученной из данного текста или представляющей элементы «знаний о мире».

Не исключено и обратное. Воспринимающий текст может не располагать соответствующими знаниями, но выявить анафорическую связь на основе интуитивного представления о построении связного текста. Тогда текст с заместителем может стать источником новой информации для воспринимающего текст.<sup>6</sup> Например:

«Лючилио Ванини понимал, что католическая церковь не простит ему вольнодумства. Над головой ученого стали сгущаться тучи».

Подобный текст может стать источником знания: Лючилио Ванини — ученый.

IV. Существительное в заместителе и главное слово антецедента находятся в отношении синонимии. Например:

«Развалины Чан-Чана сохранились чудом. Уцелеть им помогло то, что об этих руинах никто не знал».

«Каково бы ни было максимальное число знаков, на которое рассчитана машина, в процессе последовательных умножений этот максимум рано или поздно будет достигнут».

<sup>6</sup> См. об этом: Е. Фарыно. Дискуссия по вопросам семантики текста. НТИ, сер. 2. Информационные процессы и системы, № 4, 1969.

Особый интерес представляют те случаи, когда информация о контекстуальной синонимии между главным словом антецедента и существительным в заместителе может быть получена косвенным путем на основе параметральных отношений между словами, а именно существительное в заместителе представляет собой значение некоторого параметра по отношению к слову, управляющему главным словом антецедента (в частном случае именем существительным, которое является антецедентом).<sup>7</sup> Например:

«Каждый год культурные растения уносят из почвы **10 миллионов тонн фосфорной кислоты**. Эту потерю природа не в силах восполнить».

S<sub>2</sub> (уносить)=потеря, т. е. имя существительное (именная группа), которое занимает второе место при глаголе «уносить», синонимично существительному «потеря». Отсюда в нашем примере: 10 миллионов тонн фосфорной кислоты=потеря.

«В настоящей статье **Дж. Неймана** вопросам логики посвящен специальный раздел. Здесь автор обращает особое внимание на то, что...»

S<sub>1</sub> (Сaus (статья, песня, музыка, теория и т. п.)) = автор, т. е. имя существительное, которое обозначает того, кто «делает» статью, песню, музыку, теорию и т. п., синонимично существительному **автор**.<sup>8</sup> Отсюда: Дж. Нейман=автор.

«В случае равномерного кода предполагается, что все символы алфавита **равновероятны**, хотя это допущение и не отражает фактического положения о распределении вероятностей символов».

S<sub>2</sub> (предполагать, что)=допущение. Отсюда: все символы алфавита равновероятны=допущение.

При любом из рассмотренных выше подтипов (I—IV) в заместителе могут появляться определения, которые носят оценочный характер или основаны на признаках, уже выска-

---

<sup>7</sup> О параметральных отношениях между словами см.: А. К. Жолковский, И. А. Мельчук. О семантическом синтезе. В сб.: Проблемы кибернетики, вып. 19. М., 1967.

<sup>8</sup> Имя собственное лица, употребленное в форме родительного падежа при существительных типа «статья», «теория» и т. п., обозначает автора, т. е. того, кто делает «статью», «теорию» и т. п. Ср., однако, сочетание «книга Дж. Неймана», которое неоднозначно.

занных в предшествующем тексте в связи с предыдущими упоминаниями объекта:<sup>9</sup>

«Известно, что **проблема** пристрастности и границ „анализа по непосредственно составляющим“ неоднократно дискутировалась. Не рассматривая эту сложную и интересную **проблему** во всех подробностях...»

«Я полагаю, что читателю знакомы понятия о телепатии, ясновидении, способности к прорицанию и психокинезису. Эти поразительные явления, по-видимому, опровергают то...»

«Языки, порождаемые „неукорачивающими грамматиками“, являются легко распознаваемыми. Этот важный факт строго доказывается».

«Стрептококковые токсины вызывают распад собственных **белков** организма. Причем воздействие стрептококковой инфекции таково, что **эти разрушенные белки**, попадая в кровь, воспринимаются организмом как чужеродные».

«Глаза не только обнаруживают наличие предмета, но и передают все сведения о его размерах, форме, характере поверхности и расстоянии, на котором он находится. К сожалению, **этот информационный аппарат** очень ограничен в своих возможностях».

Определения, основанные на признаках, уже упоминавшихся в предшествующем тексте в связи с тем же самым объектом, не несут новой информации, но иногда они необходимы для однозначной идентификации повторно упоминаемого объекта. Например:

«Естественные языки образуют очень специфический, имеющий свои особые характеристики класс **эффективных функций**. И наш математик сразу обнаружит, что сведений об эффективных функциях, почерпнутых из теории алгоритмов, для исследования естественных языков недостаточно. Следо-

---

<sup>9</sup> Естественно считать, что заместитель не может содержать признака, который не имел места в предшествующих упоминаниях объекта (за исключением признака оценочного характера). Этим объясняется неграмматичность примера, приведенного Е. В. Падучевой в ст. «Международная конференция по семиотике в Польше» (НТИ, 1967, № 2): «Мы пили в чулане чайник. Мы вскипятили воду в этом жестяном чайнике».

По-видимому, новые признаки уже упоминавшихся ранее объектов могут вводиться в текст только в виде самостоятельных предикатов. Однако иногда в тексте имеют место подобные сокращенные способы введения новой информации, например: «Об этих титулах нам известно по найденной в Херсоне печати. Эта **подвесная свиццовая** печать скрепляла когда-то документ, от которого ничего не осталось».

Признаки «подвесная», «свиццовая» естественнее было бы ввести в текст не в виде определений, а в виде самостоятельных предикатов.

вательно, ему придется специально заняться изучением **этих особых функций**».

Учет определения «особые» позволяет считать заместитель «эти особые эффективные функции» средством повторного упоминания не об эффективных функциях, почерпнутых из теории алгоритмов, а об эффективных функциях, имеющих особые характеристики (см. «специфический, имеющий свои особые характеристики класс эффективных функций»).

Если личные местоимения 3-го лица в научно-технических текстах всегда являются средством обозначения повторно упоминаемого объекта и уже по форме их можно «узнать» как заместителей, то с рассматриваемыми здесь заместителями дело обстоит сложнее. Во-первых, очевидно, что имя существительное без идентифицирующих определений лишь в особых условиях связного текста воспринимается как заместитель. Во-вторых, имеются сочетания, которые омонимичны сочетаниям имен существительных с идентифицирующими определениями, но не являются средством повторного упоминания объекта. Например:

**«Последнее достижение фирмы — настольная электронная вычислительная машина».**

**«В любое данное время возможно использование только конечной части запоминающего устройства».**

**«Эта книга (статья, параграф, глава и т. п.) посвящена вопросам автоматике».**

Выражения типа «эта книга», «данный параграф», «предлагаемая работа», будучи употреблены до изложения соответствующей книги, работы и т. п., не являются заместителями, т. е. средством повторного названия объекта, уже упоминавшегося в тексте.

Из всего сказанного очевидно, что описание условий, при которых имена существительные с идентифицирующими определениями и без них действительно являются заместителями (иначе говоря, условий, при которых имеет место анафорическая связь между данным существительным и некоторым другим словом), представляет собой обязательную часть предлагаемого описания этих заместителей.

Опираясь на рассмотренные выше подтипы этих заместителей (I—IV), можно сформулировать условия:

Условие 1. В предшествующем тексте имеется вхождение имени существительного, совпадающего с данным существительным.

Как правило, имена существительные, служащие для предыдущего и последующего наименования одного и того же объекта, совпадают в числе. Исключение представляют те случаи, когда в первоначальном вхождении имени имеет место формально единственное число, но референт — реальная множественность. Тогда повторное вхождение этого же имени существительного имеет форму множественного числа. Например:

«Возьмем две окружности. Впишем в каждую из них правильный **многоугольник** с одним и тем же числом сторон и будем удваивать число сторон каждого из **этих многоугольников**».

«Сейчас уже всем знакомы понятия первой и второй космической скорости. Как известно, для Земли эти скорости соответственно равны 7,91 и 11,2 км/сек».

Выражение «первая и вторая космическая скорость» семантически представляет собой «склею» двух конструкций: 1) первая космическая скорость и 2) вторая космическая скорость, где «скорость» в первом случае и «скорость» во втором соответствуют разным объектам. Однако условие 1 не является достаточным для вывода о наличии анафорической связи.

Условия, устраняющие ложные анафорические связи в случае, если и предыдущее, и повторное вхождения имени существительного имеют место в пределах одной и той же предикативной группы, могут быть сформулированы в синтаксических терминах (в терминах дерева зависимостей и в терминах глубинного синтаксиса)<sup>10</sup>.

Условие IА. Если повторное вхождение имени существительного сопредикативно предыдущему вхождению того же имени существительного, то оно не заместитель для этого предыдущего вхождения. Например:

«Набор категорий изменяется от **автора к автору**».

<sup>10</sup> В попытке описать анафорические связи в виде условий, при которых эти связи могут или не могут иметь места, данная работа опирается на указанную ранее статью Е. В. Падучевой «Выражение тождества упоминаемых объектов как одна из проблем синтеза языкового текста». В частности, используется положение из этой статьи о том, что сопредикативные элементы не могут быть связаны анафорической связью. (О сопредикативности см. также: Л. Н. Кувалдина. Анафорические связи в русском языке. В сб.: Информационные вопросы семиотики, лингвистики и автоматического перевода, вып. 2. М., 1971, стр. 47—54.)



(Словоформы «автора» и «автору» являются местами одного и того же предиката «изменяется», т. е. они сопредикативны; между ними нет анафорической связи).

«Совершенно другие процессы происходят при взаимодействии **протона с протоном**». (Словоформы «протона» и «протоном» — разные места одного и того же предиката «взаимодействие», т. е. они сопредикативны и между ними нет анафорической связи).

«Присмотримся к любому атому кислорода в кристалле  $\text{CO}_2$ . С **атомом** углерода „своей“ молекулы **этот атом** связан несравненно прочнее, чем с атомами углерода соседних, „чужих“ молекул». (Словоформы «атом» и «атомом» не связаны анафорической связью; они сопредикативны, т. е. суть места одного и того же предиката «связан»).

Условие 1Б. Если повторное вхождение имени существительного синтаксически связано с предыдущим вхождением так, что находится на одной ветви дерева зависимостей (при условии, что вершина дерева — сказуемое), то оно не является заместителем для этого предыдущего вхождения. Например:

**Потомки потомков** также суть потомки<sub>2</sub>.

(Словоформы «потомки» и «потомков» не связаны анафорической связью, так как вторая словоформа синтаксически зависит от первой).

Условие 1В. Если повторное вхождение имени существительного связано с предыдущим сочинительной связью, то оно не является заместителем для этого предыдущего вхождения:

«В статье рассматриваются некоторые экономические показатели, свойственные развитым капиталистическим **странам<sub>1</sub>** и **странам<sub>2</sub>** социалистическим».

(Словоформы «странам<sub>1</sub>» и «странам<sub>2</sub>» связаны сочинительной связью, поэтому между ними не имеет места анафорическая связь).

Если предыдущее и повторное вхождения имени существительного находятся в разных предикативных группах, то однозначное выявление анафорической связи может быть достигнуто путем учета слов зависимых от предыдущего и повторного вхождения. Как мы уже отмечали, в целях однозначной идентификации в заместителе повторяется часть слов, зависимых от главного слова антецедента, или вся

именная группа antecedента или появляются определения, основанные на признаках, уже упоминавшихся в тексте в связи с вхождением того же имени существительного.

В некоторых случаях для устранения ложных анафорических связей употребляются специальные слова-дифференциаторы («другой», «один—другой», «остальные», «новый» и т. п.). Имена существительные с такими словами не могут употребляться для повторного упоминания объекта:<sup>11</sup>

«Прямая, проведенная через середину стороны треугольника, параллельно другой его стороне, делит третью сторону пополам».

«Если середины сторон какого-угодно четырехугольника взять за вершины нового четырехугольника, то последний есть параллелограмм».

Заметим, что для заместителя «этот S» такая ситуация, когда предыдущее вхождение имени существительного, совпадающего с существительным в заместителе, не являлось бы его antecedентом, очень редка.

Условие 2. В предшествующем тексте имеются глагол, причастие или прилагательное, для которых данное существительное — синтаксический дериват.

Условие 3. В предшествующем тексте есть имя существительное или глагол, по отношению к которому данное имя существительное — классификатор.

Условие 4. В предшествующем тексте имеется некоторое существительное или придаточное дополнительное предложение при глаголах со значением «мысли, речи, чувства»,<sup>12</sup> с которым данное существительное находится в отношении контекстуальной синонимии.

Очевидно, в ситуациях, для которых справедливы условия 2, 3, 4, 5, выявление анафорических связей возможно лишь на основе семантической информации. Однако можно указать некоторые способы формализации семантической информации. Например, некоторая информация о родо-видовых и синонимических отношениях может быть задана в словаре. Это относится к постоянным «словарным» родо-видовым соотношениям и случаям синонимии.

<sup>11</sup> См. об этом также: Е. В. Падучева, ук. ст. Тр. III Всесоюзной конференции, т. II. М., 1967, стр. 128.

<sup>12</sup> Интересно, что Ш. Балли указывал на возможность рассматривать подобные придаточные предложения как случаи перехода предложений в класс существительных (Ш. Балли. Общая лингвистика и вопросы французского языка. М., 1955, стр. 131).

Иногда информация о контекстуальной синонимии может быть получена косвенным путем, на основе параметральных отношений между словами. Так, соотношения типа  $S_2$  (уносить) = потеря (см. примеры на стр. 69) и соответствующее правило. Слово, занимающее второе место при глаголе «уносить» и синонимичное слову «потеря», может быть задано заранее.

Б. М. Лейкина

### ЗАДАЧИ ЭТАПА ПРЕОБРАЗОВАНИЯ В СИСТЕМЕ АП

Системы АП, построенные по принципу независимости анализа и синтеза, характеризуются наличием особого преобразовательного этапа, на котором структура переводимого отрезка текста приводится к некоторому стандартному виду, лишенному особенностей как входного, так и выходного языка. Этап преобразований входит в качестве одной из составных частей и в экспериментальную систему АП, разработанную в ВЦ ЛГУ<sup>1</sup> и проходившую в 1967—1970 гг. (в первом своем варианте) комплексную экспериментальную проверку и частично отладку.

Каждая из основных частей данной системы (анализ, преобразование, синтез) представляет собой набор специализированных грамматик, построенных по определенной модели, и универсальных алгоритмов работы с такими грамматиками. Первый и последний этапы работы системы обеспечиваются соответственно анализирующими и синтезирующими грамматиками,<sup>2</sup> промежуточный этап — трансформационными грамматиками, или грамматиками преобразования.

<sup>1</sup> Б. М. Лейкина, Т. Н. Никитина, М. И. Откупщикова, С. Я. Фитиалов, Г. С. Цейтин. Система автоматического перевода, разрабатываемая в группе математической лингвистики ВЦ ЛГУ. НТИ, 1966, № 1.

<sup>2</sup> К синтезирующим грамматикам относится, в частности, разработанная М. И. Откупщиковой грамматика позиционного синтеза, с которой работает тот же алгоритм, что и с грамматиками преобразований. — См.: М. И. Откупщикова. Грамматика порядка слов для синтеза при автоматическом переводе. Автореф. канд. дисс. Л., 1967.

В соответствии с принципом независимости анализа и синтеза этап преобразования, в свою очередь, распадается на Преобразования I и II. Основной задачей Преобразования I является переход от структуры предложения входного языка к стандартной форме представления содержания этого предложения — структуре предложения на языке-посреднике. Преобразование II вносит в структуру переводимого предложения, представленную в терминах языка-посредника, изменения, обусловленные спецификой выходного языка, и вырабатывает информацию, нужную для морфологического синтеза.

Анализирующие грамматики и грамматика Преобразования I разработаны на материале английского языка, синтезирующие грамматики и грамматика Преобразования II — на материале русского языка. Обе грамматики преобразования построены по одной и той же модели. Единым является и интерпретирующий алгоритм работы с этими грамматиками.<sup>3</sup>

Объектом преобразования является синтаксическая структура переводимого предложения — иерархически упорядоченная совокупность его элементов. В качестве элементов синтаксической структуры выступают словоформы (или «свернутые» устойчивые комплексы словоформ) входного текста, задаваемые информацией, сопоставленной им во входном словаре и подвергшейся соответствующим изменениям на этапах, предшествующих этапу преобразования.

В основу принятой системы информации о словах положена открытая пересекающаяся классификация. Информация задается в виде набора признаков, соответствующих различным элементарным разбиениям и расположенных в определенном порядке.<sup>4</sup> Структура информации об основе во входном словаре — двухъярусная. Верхний из выделяемых ярусов

<sup>3</sup> Первоначальный вариант модели грамматики преобразования был разработан Г. С. Цейтиным и Б. М. Лейкиной. Несколько измененный действующий вариант модели грамматики и интерпретирующий алгоритм, позволяющий производить нужные преобразования по условиям любой грамматики, построенной по данной модели, разработаны Н. А. Крупко — См.: Н. А. Крупко. О языке и алгоритме... В сб.: Методы вычислений. вып. 8. Изд. ЛГУ, 1972.

<sup>4</sup> Б. М. Лейкина, Т. Н. Никитина, М. И. Откупщикова, С. Я. Фитиалов, Г. С. Цейтин, ук. ст. НТИ, 1966, № 1, стр. 14; Б. М. Лейкина. Принципы построения английской грамматики для синтаксического анализа в машинном переводе. НТИ, 1964, № 11, стр. 39—40.

можно условно назвать грамматическим, нижний — лексическим. На каждом ярусе возможно ветвление, т. е. наличие нескольких вариантов (веток) информации — дизъюнкция из нескольких конъюнктивных наборов признаков. Ветвление обусловлено неоднозначностью основы: грамматической неоднозначностью на верхнем ярусе и лексической неоднозначностью на нижнем.

На этапах, предшествующих преобразованию, грамматика и алгорифмы оперируют лишь с верхним ярусом информации о слове, ее «грамматическим кустом», характеризующим слово в целом. В результате работы досинтаксических этапов анализа (морфологического анализа, анализа устойчивых комплексов, этапа предварительного решения грамматической омонимии) часть веток грамматического куста информации, соответствующих отдельным грамматическим вариантам («омонимам») текстового слова, может быть вычеркнута. В случае остаточной грамматической неоднозначности слова (т. е. при сохранении двух или более вариантов информации на верхнем ярусе) предложение на входе в синтаксический анализ расщепляется на однозначные «омофразы» — грамматические варианты предложения, не содержащие эксплицитно неоднозначных слов.

Преобразование работает с отдельными вариантами анализа таких омофраз. При этом используется не только соответствующая ветка верхнего яруса информации — общая информация, или «информация о слове», но и соответствующий данному грамматическому омониму нижний куст информации, каждая ветка которого представляет собой информацию об отдельном лексическом варианте (лексическом значении) данного слова и включает номер значения, служащий входом в выходной словарь, и «лексический дифференциатор» — признак, обозначающий тип контекстуальных условий реализации данного значения.

В основе представления синтаксической структуры предложения, вырабатываемой на предшествующем преобразовании этапе синтаксического анализа, лежат понятия дерева зависимостей и валентных «мест», заполняемых словами, зависимыми от данного. Трактовка синтаксической структуры анализируемого предложения как дерева зависимостей в известном смысле уравнивает типы синтаксической связи: все они (в том числе и отношение сочинения) представляются как бинарное иерархическое отношение доминанции и описы-

ваются в терминах «хозяина» (господствующего слова) и подчиненного слова. Дифференцированная характеристика типов синтаксической связи обеспечивается вырабатываемой в синтаксическом анализе информацией о валентных местах. Содержание (или тип) отношения доминанции (тип ОНД, по терминологии И. А. Мельчука) рассматривается как «относительное», «синтаксическое» значение и указывается в общей части информации о подчиненном члене соответствующей пары узлов дерева зависимостей в виде номера места, в счет которого данный член подчиняется своему хозяину. Общая часть информации каждого слова, кроме вершины предложения, включает также номер слова, которому оно подчинено («номер хозяина»).

Предложение на этапе преобразования представлено в виде линейно упорядоченной совокупности информации об отдельных его словах, соответствующей на входе в преобразование линейной последовательности слов входного предложения.

Граматики преобразования представляют собой набор правил («трафаретов»), применяемых в определенном (фиксированном в грамматике) порядке. В левой части трафарета задано некоторое условие на подструктуру предложения, в правой части указаны преобразование, которое следует произвести, если трафарет «наложился» на рассматриваемое предложение, т. е. если данное условие в нем выполняется, и адреса трафаретов, к которым следует перейти в случае наложения и в случае неналожения трафарета. Правила записываются на некотором формальном языке, описание которого дано в статье Н. А. Крупко.<sup>5</sup>

Каждая из грамматик распадается на ряд блоков. Правила, входящие в один блок, объединены общностью решаемой задачи и/или однотипностью предусматриваемых ими операций преобразования.

В Преобразовании I решаются три основные задачи: 1) браковка неправильных вариантов анализа, 2) интерпретация незабравованных вариантов и 3) преобразование их к языку-посреднику.

Браковка вариантов производится главным образом за счет привлечения дополнительной по сравнению с этапом синтаксического анализа информации о семантических подклас-

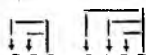
---

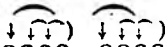
<sup>5</sup> Н. А. Крупко, ук. ст. В сб.: Методы вычислений, вып. 8. Изд. ЛГУ, 1972.

сах и лексической сочетаемости. Так, в предложении This will always result in the suppression of desired target echoes and receiver noise<sup>6</sup> один из вариантов анализа, в котором сочинительной связью объединялись словоформы suppression и noise, был отбракован как не удовлетворяющий требованию «семантической однородности» сочиненных членов при наличии «кандидатов» (echoes and noise), удовлетворяющих такому требованию.

Интерпретация распадается на несколько видов. Один из них — выявление «истинной» (реализованной в данном случае) видовой структуры в «спрессованной» родовой. Эта задача диктуется применяемым на этапе синтаксического анализа приемом «компрессии» вариантов синтаксического анализа, т. е. замены определенного набора таких вариантов одним из членов этого набора. Из практических соображений экономии времени, требуемого на автоматический перевод определенной единицы текста, в некоторых случаях структурной неоднозначности дистрибутивная модель, не являющаяся однозначной в терминах грамматики зависимостей, условно трактуется грамматикой как однозначная и получает на выходе синтаксического анализа один (статистически преобладающий) вариант анализа, служащий как бы родовым именем целого класса анализов. Выявление реализованной в данном отрезке текста видовой структуры, входящей в класс структур, представленных родовой, и является одной из задач интерпретации результата анализа. Метод «компрессии» на этапе анализа и соответствующие операции интерпретации в преобразовании применяются главным образом в тех случаях, когда известны все альтернативные варианты анализа («объем» и характер неоднозначности дистрибутивной модели) и способ решения данного типа структурной неоднозначности. В отдельных случаях спрессовываются и такие дистрибутивные модели, которые не могут быть однозначно интерпретированы. Подчиненному члену их в таком случае в Преобразовании I приписывается признак «условности хозяина». Примером спрессовываемых в анализе и требующих интерпретации неоднозначных дистрибутивных моделей могут служить дистрибутивные модели SSS, SSSS и т. д. с последовательным  $\begin{matrix} \frown & \frown & \frown \\ \downarrow & \downarrow & \downarrow \end{matrix}$  или параллельным  $\begin{matrix} \frown & \frown & \frown \\ \downarrow & \downarrow & \downarrow \end{matrix}$

<sup>6</sup> См. приведенные в конце статьи примеры экспериментальных переводов.

  
(SSS, SSSS) подчинением или же, в случае четырехэлементных (или еще более сложных) моделей, характеризуемые сочетанием последовательного и параллельного подчинения

  
SSSS, SSSS и т. п.), дистрибутивные модели ASS, ASSS, QSS, SFSP, SFSA и др.<sup>7</sup>

Другой вид интерпретации — выявление семантики грамматических форм (числа существительного, времени, вида, наклонения, залога глагола и т. п.). В случае однозначности грамматической формы интерпретация этого типа сводится к простой перекодировке формальных признаков входного языка в термины грамматических значений языка-посредника. В случае неоднозначных форм интерпретация включает и решение соответствующих типов грамматической омонимии. Ср. предложения 3 и 4, в первом из которых глагольная форма Present Indefinite интерпретируется как форма настоящего времени, а во втором (на основе учета типа придаточного предложения и соотношения форм глаголов — сказуемых главного и придаточного) — как форма будущего времени. Примером могут служить и предложения 2 и 7, в которых на основе выработанной в Преобразовании информации о многократности действия глагол переведен формой несовершенного вида.

К этому типу интерпретации относится и различение грамматических значений многофункциональных служебных глаголов — задача, решаемая в особом блоке «свертки сложных глагольных форм» (содержательно — блоке аналитической морфологии глагола). Иллюстрацией решения задач такого типа могут служить предложения 6 и 14, в первом из которых глагол *do* интерпретируется как формант отрицательной формы глагола, а во втором — как показатель определенного модального значения — подчеркнутой оценки действия как реального (*does result* «действительно приводит»).

Следующим видом интерпретации является семантическая интерпретация синтаксической структуры — выработка информации о типах синтаксической связи и об актуальном членении предложения на различных уровнях: непредикативных («адъюнктивных»), по терминологии Есперсена, подчинен-

<sup>7</sup> S означает существительное, A — прилагательное, Q — числительное, F — предлог, P — причастие.



ных членов, предикативных («нексусных») подчиненных словосочетаний и придаточных предложений.

Информация о типе синтаксической связи получается на основе учета номера валентного места и индивидуальных характеристик членов элементарного сочетания (хозяина—подчиненного). В отдельных случаях анализируется и более широкий контекст.

Информация об актуальном членении вырабатывается на основе позиционного соотношения хозяина и подчиненного; учитываются также особенности некоторых синтаксических конструкций (так называемых выделительных конструкций, непроективных структур). Ср., например, перевод предложения 5, в котором постпозиция группы подлежащего обусловлена выработанной в Преобразовании I информацией об актуальном членении первого уровня, основанной на смысловой функции непроективной структуры.

Чрезвычайно важной задачей интерпретации в Преобразовании I является «лексическая интерпретация» — выявление реализованного в данном случае лексического варианта слова, т. е. решение лексической неоднозначности. Эта задача решается в некоторых случаях средствами «внутреннего» контекста путем учета допустимых сочетаний признаков в информационной формуле неоднозначного слова, например допустимости сочетания дифференциатора лексического значения «уверен» слова *certain* лишь с синтаксическим признаком функции предикативного члена, а значения «некоторый» — лишь с синтаксическим признаком атрибутивного отношения. Сопоставление лексического дифференциатора отдельных веток лексического куста слова с соответствующим признаком общей информации о слове диктует в таких случаях выбор одной из веток и отбрасывание остальных (ср. перевод предложения 20).

Значительно чаще, однако, лексический дифференциатор диктует учет средств внешнего контекста: наличие или отсутствие определенного типа подчиненных членов, их оформление и т. п. Автоматическое решение лексической неоднозначности средствами внешнего контекста иллюстрируют, например, предложения 2 и 8 (*result* 'приводить к', *result* 'получаться'), 9 и 10 (*allow* 'позволять', *allow* 'обеспечивать'), 13 и 12 (*is* 'служит (для)', *is* 'находится'), 11 (*consist* 'заклучаться (в)', а не 'состоять (из)') — значение, также предусмотренное во входном словаре), 1 (*mean* 'означать', а не 'иметь в виду') и др.

Из более частных задач интерпретации в Преобразовании I следует отметить учет разрывных устойчивых комплексов и учет анафорических связей.

Устойчивые комплексы, анализируемые в Преобразовании I, представляют собой в основном определенные параметры или сочетания параметров некоторых ключевых слов.<sup>8</sup> Сюда относятся такие сочетания, как *pay attention (to), take place, set in motion, take into consideration* и т. п. Комплексы такого рода не могут учитываться в предшествующем синтаксическому анализу блоке обработки устойчивых комплексов, ориентирующемся лишь на линейную, а не на синтаксическую структуру и рассматривающем соответственно лишь неразрывные устойчивые комплексы. Для анализа разрывных устойчивых комплексов вследствие их грамматической и позиционной вариабильности (ср., например, *no attention has been paid...; great attention, as has been mentioned above, should be paid...*) должен учитываться не только линейный, но и синтаксический контекст, что возможно лишь на послесинтаксическом этапе анализа — в Преобразовании I. Ядром таких комплексов считается существительное. Ему во входном словаре в качестве особого омонима приписывается значение комплекса в целом и соответствующий оператор (лексический дифференциатор данного значения). «Параметральное слово» (глаголы *to have, to be, to make, to put*, и некоторые другие) в случае реализации данного значения стирается; при этом соответствующая грамматическая информация предварительно переносится в информацию, сопоставленную ядру.

Учет анафорических связей в действующем варианте системы осуществляется в очень ограниченных масштабах и лишь для субстантивных заместителей *one* и *that* и для вспомогательных глаголов. Интерпретация анафорических связей в полном объеме не может быть осуществлена в рамках пофразного перевода, так как она обязательно должна учитывать и межфразовые связи.

Следующая из отмеченных выше основных задач Преобразования I — преобразование структуры входного предложения к стандартной форме представления его содержания, т. е. к предложению на языке-посреднике, может быть разбита на ряд более частных задач. К ним относится в первую очередь устранение элементов «аналитической морфоло-

<sup>8</sup> А. К. Жолковский, И. А. Мельчук. О семантическом синтезе. В сб.: Проблемы кибернетики, вып. 19. М.—Л., 1967.

гии», слов, выполняющих чисто служебную функцию в предложении: артиклей, предлога *of*, вспомогательных глаголов. После соответствующей интерпретации, результатом которой является приписывание знаменательным словам грамматических значений, выражаемых с помощью указанных служебных элементов, последние устраняются из структуры дерева зависимостей, сопоставленного данному предложению. К операциям этого рода примыкает и устранение предлогов сильного управления, не требующее специальной интерпретации и связанное лишь с переносом информации о типе объектного отношения (первообъектное, второобъектное и т. д.) с предлога на соответствующий именной член. Преобразование этого типа отражают перевод почти любого из экспериментальных предложений (ср., например, предложения 2, 5, 6, 7).

Другого типа частной задачей преобразования к языку-посреднику является устранение специфичных для английского языка случаев тавтологичности (например, повторного выражения модального значения реальности описываемой ситуации: **This does really intensify...** и т. п.) и «расщепление» некоторых специфичных для английского языка совмещений значений, например совмещения значения предиката условия или цели со значением отрицания (*unless, lest*). Результат такого расщепления иллюстрирует перевод предложения 4.

Наконец, чрезвычайно существенной задачей при переходе к языку-посреднику является приведение к стандартному виду некоторых специфически английских синтаксических структур (конструкций с конечным предлогом, бессоюзных придаточных предложений, союзных причастных и предложных герундиальных оборотов и некоторых других). Такого рода преобразованию подвергалось, например, дерево зависимостей, выработанное в результате синтаксического анализа английских предложений 15, 16 и 17.

Преобразование I завершается стиранием всей информации, не входящей в систему языка-посредника. Остаточная информация о каждом узле преобразованного дерева зависимостей включает номер соответствующего лексического значения и (в случае невершинных узлов) номер хозяина и признак, обозначающий тип синтаксического отношения (ОНД), связывающего данный узел с его хозяином. Кроме того, в информацию о некоторых узлах могут входить зна-

чения определенных грамматических категорий, предусмотренных грамматикой языка-посредника (числа, времени, залога и некоторых других), а также определенные семантические множители.

После Преобразования I происходит обращение к выходному словарю, в котором каждому номеру лексического значения сопоставлен его лексический куст — определенное множество (от 1 до 7) переводных основ. Каждая ветка лексического куста характеризуется своим набором информационных признаков.

Преобразование II (от языка-посредника к естественному (выходному) языку) представляет собой как бы зеркальное отображение Преобразования I (преобразования от естественного (входного) языка к языку-посреднику). Задачи Преобразования II, таким образом, в самых общих чертах аналогичны описанным выше задачам, решаемым в Преобразовании I, и включают преобразование посреднической структуры предложения к соответствующей русской структуре, интерпретацию (исходя из норм русского языка) лексической и грамматической информации, сопоставленной узлам дерева зависимостей, и, наконец, браковку неправильных вариантов синтеза. Как и в Преобразовании I, интерпретация и собственно преобразование тесно переплетаются между собой.

Блоки Преобразования II распадаются на три основные группы: 1) блоки выбора основы, 2) блоки грамматической интерпретации и преобразования и 3) блок браковки результатов преобразования. Имеется также особый блок подпрограмм, учитывающий случаи, когда одному «слову» языка-посредника (одному номеру значения) в выходном языке соответствует словосочетание из двух или более слов, которое вследствие изменяемости неконечного элемента (или элементов) не может быть задано в виде отдельной лексической единицы.

Блоки выбора основы работают в тех случаях, когда номеру лексического значения в информации определенного узла дерева зависимостей переводимого предложения сопоставлен в выходном словаре куст с числом веток  $\geq 2$ . Неоднозначность выражения лексического значения обусловлена несколькими факторами: «грамматической супплетивностью» основ, «семантической супплетивностью», т. е. вариативностью, связанной с учетом семантической деривации

(одно и то же лексическое значение в сочетании с разными семантическими множителями может выражаться разными основами), синонимией и тем, что можно назвать «фразеологической супплетивностью» — различным оформлением параметральных значений при разных ключевых словах.

Супплетивность грамматическая и семантическая решается на основе информации, пришедшей из языка-посредника (информации о значении той или иной грамматической категории или об определенном семантическом множителе, предусмотренных в языке-посреднике). Так, например, в случаях супплетивности типа «человек»/«люди», «хорошо»/«лучше», «идет»/«шел», «был»/«буд-» и т. п. информация о числе, степени сравнения, глагольном времени однозначно предпрещает выбор основы. Это же относится и к семантической супплетивности (ср., например, «читать»/«чтение», «переводить»/«переводчик», «большой»/«увеличивать» и т. п.). В этих случаях задача выбора одной из конкурирующих основ решается средствами внутреннего контекста: выбирается основа, валентностные характеристики которой, фиксированные в ее словарной информации, согласуются с реализованными условиями функционирования соответствующего значения, фиксированными в информации, выработанной для данного узла в результате анализа предложения.

Выбор синонима связан обычно с учетом внешнего контекста, и информация о синонимичных основах (аналогично информации о конкурирующих лексических вариантах в Преобразовании I) включает специальные признаки, диктующие анализ тех или иных контекстуальных условий, анализ, результат которого решает судьбу данной основы.

В качестве многоосновных (синонимичных) слов задаются и некоторые из таких широких значений, как параметры А. К. Жолковского и И. А. Мельчука, — значения *Corul*, *Magh*, *Antimagh* и некоторые другие.<sup>9</sup>

Иллюстрацией результатов работы блоков выбора основы могут служить экспериментальные переводы предложений 4, 6, 16, 17, 18, 19 (выбор переводного эквивалента глагола-связки), 3 и 19 (выбор переводного эквивалента прилагательного *negligible*), 6 и 22 (выбор переводного эквивалента при-

<sup>9</sup> Параметры *Loc*, *Dir<sub>1</sub>* и *Dir<sub>2</sub>* рассматриваются как грамматические (синтаксические) значения и учитываются в одном из блоков грамматической интерпретации.

лагательного true), 21 (выбор лексической реализации параметра Magn) и ряд других.

Блоки грамматической интерпретации и преобразования решают ряд частных задач, связанных с приведением стандартизованного посреднического представления содержания переводимого предложения к виду, согласующемуся с нормами и спецификой русского языка. Здесь происходит синтаксический синтез и вырабатывается информация, нужная для морфологического синтеза.

Примером блоков этой группы могут служить: 1) Блок опорных слов, который предусматривает вставку нужных соотносительных местоимений и наречий, обобщающей соотносительной частицы **ни** (сколько бы... ни), эксплетивного отрицания (until 'пока... не') и т. п.; 2) Блоки интерпретации информации о типе синтаксической связи (решающие задачу выбора формы выражения различных синтаксических значений, в том числе и значений параметров Loc, Dir<sub>1</sub> и Dir<sub>2</sub>); 3) Блок согласования, учитывающий не только иерархические, но и некоторые анафорические связи; 4) Блок аналитической морфологии.

Завершающий Преобразование II Блок браковки результатов преобразования является лишь предварительной меткой заключительного фильтра, который не пропускал бы некоторые варианты преобразования.

Грамматики преобразования предусматривают четыре основных типа трансформаций. К **первому** типу можно отнести изменение общей части информации об отдельном слове. Операции этого типа имеют место при обработке любого предложения.

**Второй** тип составляют трансформации изменения структуры предложения без изменения числа его компонентов путем переподчинения отдельных слов. Трансформация этого рода применялась, например, при обработке предложения 2: словоформа desired, на входе в преобразование подчинявшаяся существительному echoes, в Преобразовании I переподчинилась существительному target.

**В третий** тип входят трансформации, изменяющие структуру предложения путем добавления новых узлов к исходному дереву зависимостей предложения. Иллюстрацией могут служить предложения 3 и 19, где добавленными узлами являются подчинительный союз **что** и запятая; предложение 17, в котором бессоюзное определительное придаточное

предложение в Преобразовании I заменяется на придаточное, вводимое относительным местоимением; предложение 19, в котором один узел, соответствующий английскому прилагательному *negligible*, замещается на два: «можно» и «пренебречь»; предложение 18, где добавленным (в Преобразовании II) узлом является указательное местоимение «эти».

**Четвертый** тип трансформаций составляют операции устранения, или вычеркивания, различных объектов. Разновидностями трансформаций этого типа являются: 1) вычеркивание отдельных значений неоднозначного слова в Преобразовании I; примером могут служить предложения 6, 17 и др., в которых вычеркиваются все «несвязочные» значения глагола *to be* ('равняться', 'иметься', 'иметь', 'находиться', 'служить (для)'); 2) аналогичные операции вычеркивания отдельных основ «неодноосновных» значений в Преобразовании II; 3) устранение отдельных узлов преобразуемого дерева зависимостей; 4) устранение отдельных вариантов анализа или преобразования.

Все операции вычеркивания содержательно являются операциями браковки вариантов информации. Браковка вариантов, не отбракованных в анализе, осуществляется за счет использования в преобразовании дополнительных фильтров в виде ряда семантических и структурных признаков, не учитываемых в анализе.

Число трансформаций, осуществляемых при обработке одного предложения, может быть довольно значительным. Так, например, при обработке предложений 2 (*This will always result in the suppression of desired target echoes and receiver noise*), помимо изменения общей информации об отдельных узлах соответствующего предложению дерева зависимостей, применялись следующие трансформации: 1) браковка лексических значений (*result* 'являться результатом', 'получаться'); 2) браковка синонимических основ (*desired* 'полезный', 'требуемый', 'заданный'); 3) браковка видовых вариантов основ (*result* 'привести'); 4) устранение узлов (артикля, предлогов *of* и *in*, глагола *will*); 5) добавление узлов (предлогов **к** и **от**, глагольных форм «отраженный» и «будет», запятых); 6) изменение структуры дерева зависимостей путем переподчинения узлов (переподчинение узлов, соответствующих словоформам *desired* и *target*).

Преобразование (так же, как и предшествующий этап синтаксического анализа) является в принципе множествен-

ным. Предложению с единственным вариантом синтаксического анализа на входе в преобразование могут соответствовать несколько вариантов перевода. В случае поступления на вход преобразования нескольких вариантов анализа каждый из них в преобразовании обрабатывается в отдельности и может в свою очередь породить несколько вариантов перевода.

Основной причиной неодновариантности как на этапе синтаксического анализа, так и на этапе преобразования является ограничение протяженности контекста, используемого для разрешения различных случаев неоднозначности, рамками переводимого предложения. Для однозначного перевода, как известно, в ряде случаев необходим учет значительно более широкого контекста.

Более того, для однозначного перевода нередко требуется и «экстралингвистический контекст» — определенные знания об объектах и явлениях, о которых идет речь, знания, на наличие которых у читателя ориентируется автор текста и которые поэтому эксплицитно в переводимом тексте не выражены. В случае человеческого (а не машинного) перевода этот факт иногда приводит к одновариантному, но неверному переводу. Пример из перевода книги А. С. Залманова «Тайная мудрость человеческого организма» (М.—Л., 1966): «**Разве до антибиотической эры наблюдалось столько случаев ультрахронического, ультраупорного, скрытого воспаления легких, без температуры, без кашля, без мокроты, которая выводит большую часть населения из строя?**» (стр. 152). Принцип множественного перевода исключает такого рода ошибки.

Недостаточность контекста предложения проявляется прежде всего при решении структурной неоднозначности, т. е. на этапе синтаксического анализа. В преобразовании варианты обусловлены в первую очередь тем, что контекст преобразуемого предложения может оказаться недостаточным для выбора лексического варианта неоднозначного слова в Преобразовании I (случай более редкий, не представленный ни в одном из экспериментальных переводов) и при выборе нужной основы в Преобразовании II, если одному лексическому значению соответствуют два или более переводных эквивалента (главным образом синонимические основы или видовые варианты основы—случай значительно более частый).

В преобразовании используется ряд приемов, ограничивающих число вариантов перевода там, где это может быть



сделано без ущерба для смысла перевода. В случае нескольких остаточных (т. е. не исключенных анализом контекста) синонимичных основ, совпадающих по своим синтаксическим свойствам, одна из основ задается как основная, а остальные приводятся в скобках, например: *desired* в ряде случаев переводится как 'требуемый (заданный)'.

Неодновариантность перевода, которая могла бы быть вызвана варьированием основы, обусловленным в значительной степени эвфоническим фактором (ср. «краткие» и «полные» формы предлогов **в, во, к, ко, о, об** и др.), предотвращается стандартизацией — заданием в выходном словаре одного (статистически ведущего) варианта основы. Ср. перевод предложения 6 (...в всех случаях вместо «**во** всех случаях»).

Особым приемом сокращения числа вариантов перевода (приемом «переводной компрессии»), подсказанным общеизвестным фактом избыточности языкового кода, являются отмеченная выше выработка в Преобразовании I признака недостаточности анализа для некоторых неоднозначных структур, проанализированных стандартным образом (т. е. применением метода «компрессии») в синтаксическом анализе, и приписывание в Преобразовании II на основе этой информации признака усеченной формы согласуемым словам (прилагательным, причастиям, относительным местоимениям), которые могут быть соотнесены с разными существительными, не совпадающими между собой по «согласовательным» признакам. Так, в приведенном выше предложении, неправильно переведенном человеком, предусмотренные в грамматиках преобразований операции обусловили бы перевод относительного местоимения усеченной формой «котор.».

В остальных случаях переводимое предложение расщепляется на варианты. Ср., например, три варианта предложения I, два из которых обусловлены объективной неоднозначностью английского предложения в условиях пофразного перевода, а третий — отсутствием в пределах переводимого предложения контекстуальных уточнителей видового значения глагола.

Операции браковки и «переводной компрессии» существенно сокращают число вариантов перевода. Первые комплексные эксперименты по автоматической обработке предложения от буквенного ввода его в машину до печати перевода являются в этом отношении достаточно обнадеживаю-

щими. Более одного варианта перевода получалось в основном лишь в случае структурной неоднозначности входного предложения, не разрешимой без учета контекста, выходящего за рамки данного предложения, или же за счет видовых вариантов основы, что для практического перевода не существенно, так как этот источник неединственности перевода может быть устранен путем стандартизации выбора видовой формы глагола при отсутствии соответствующего контекста.

Экспериментальная проверка действующей системы АП, часть которой составляют описанные здесь грамматики преобразования, позволяет надеяться, что содержательное наполнение отдельных частей системы (в частности, грамматик для разных этапов АП) может быть положено в основу системы практического информационного перевода. Вместе с тем последующие варианты системы строятся с учетом некоторых недостатков первого варианта.

Основным недостатком описанных грамматик преобразования следует считать алгоритмический метод их построения, не дающий, в частности, возможности в полной мере использовать при анализе и синтезе понятие правильной семантической структуры текста. Вторым чрезвычайно существенным недостатком, характеризующим систему в целом, являются принцип пофразного перевода (исключающий возможность использования более широкого контекста) и расщепление переводимого предложения на входе в синтаксический анализ на отдельные омофразы, что приводит к необходимости решения задачи выбора лексического варианта формы (решения лексической неоднозначности) или формального варианта значения (решение синонимии или супплетивности разных видов) в особых блоках выбора лексического варианта или выбора переводящей основы.

Во втором варианте системы (не завершеном и не подвергнувшемся экспериментальной проверке даже в уже готовых своих частях) нет разбиения на омофразы. Лексические варианты слова задаются на том же уровне, что и грамматические варианты, и «просеиваются» при установлении синтаксических связей. Связи устанавливаются не между словами, а между их грамматико-лексическими вариантами. Это позволит, как нам представляется, решать задачи, связанные с неоднозначным соответствием между формальными и семантическими единицами лексического и морфологиче-

ского уровнем языка, попутно с другими задачами анализа и синтеза и в значительной мере сократит работу этапа преобразования.

В запланированном третьем варианте системы предполагается еще большее сокращение, если не полное устранение, Преобразования I как отдельного этапа в системе АП за счет расширения сферы операций, осуществляемых на этапе синтаксического анализа.

### Примеры экспериментальных переводов:

1. Automation means increased production — Автоматизация означает возросшее производство. Средства автоматизации увеличивали (увеличили) производство.<sup>10</sup>

2. This will always result in the suppression of desired target echoes and receiver noise — Это всегда будет приводить к подавлению сигналов, отраженных от действительной цели, и шумов приемника.

3. All receivers of that type are supposed to have negligible noise. — Предполагают, что все приемники этого типа имеют незначительные шумы.

4. Unless the resistance of the circuit is decreased, this invention will not be particularly valuable — Если сопротивление цепи не будут снижать (не снизят), это изобретение не будет особенно ценным.

5. Two cases have occurred of serious blooming of the picture. — Имели место два случая значительного искажения изображения.

6. This does not seem to be true in all cases — По-видимому, это не в всех случаях верно.

7. The echo has often caused considerable suppression of the desired signal — Отраженный сигнал часто приводил к значительному подавлению полезного сигнала.

8. Signal damping may result in case of considerable circuit resistance — Затухание сигнала может получаться (получиться) в случае значительного сопротивления цепи.

9. This allows the viewer to determine the height and duration of the flight — Это позволяет наблюдателю определить высоту и длительность полета.

---

<sup>10</sup> Вариант анализа, отличающийся от данного лишь формой глагола-сказуемого, фиксируется здесь и далее в скобках.

10. The invention allows a considerable increase of the duration of the desired signal — Изобретение обеспечивает значительное возрастание длительности полезного сигнала.

11. The preliminary improvement consisted in decreasing the duration of the stroke — Предварительное усовершенствование заключалось в снижении длительности такта.

12. The subtract circuit is in front of the picture tube — Цепь вычитания находится перед кинескопом.

13. The subtract circuit in front of the picture tube is for color purification — Цепь вычитания перед кинескопом служит для повышения чистоты цвета.

14. The improvement of the picture tube does result in the purification of the picture — Усовершенствование кинескопа действительно приводит к повышению чистоты изображения.

15. Should the resistance decrease, the picture would be normal — Если бы сопротивление снижалось (снизилось), изображение было бы нормальным.

16. The resistance in the circuit and the duration of the signal are supposed to have been normal — Предполагают, что сопротивление в цепи и длительность сигнала были нормальными.

17. The noise the desired target echoes result in is considerable — Шумы, к которым сигналы, отраженные от действительной цели, приводят, значительны.

18. The two suppositions may be true — Эти два предположения могут быть верными.

19. The end connectors are supposed to be so large that their resistances are negligible — Предполагают, что оконечные соединители так велики, что их сопротивлениями можно пренебречь.

20. The viewer was certain that certain causes of the damage could have been prevented — Наблюдатель был уверен, что некоторые причины повреждения были можно предотвратить.

21. The mean duration of the desired signal is longer than that of the echo — Средняя длительность полезного сигнала больше чем отраженного сигнала.

22. In case of subtract circuits a decrease of true results seems to be large — По-видимому, снижение истинных результатов в случае цепей вычитания велико.

## АНАЛИЗ НЕКОТОРЫХ ЧИСЛОВЫХ ХАРАКТЕРИСТИК СТАТИСТИЧЕСКОГО ИССЛЕДОВАНИЯ РЕЧИ ПРИ ПАТОЛОГИИ МЫШЛЕНИЯ

Настоящая работа является продолжением ранее начатого статистического исследования синтаксических структур в образцах патологической речи 28 больных, длительно страдающих эпилепсией и шизофренией. Описание материала исследования, методика и программа машинной обработки, а также предварительные данные исследования опубликованы ранее.<sup>1</sup> Целью данной статьи являются прежде всего сопоставление клинических наблюдений с данными анализа результатов машинной обработки указанных образцов патологической речи, а также выявление наиболее достоверных числовых характеристик и оценка их дифференциально-диагностической значимости при определении характера расстройства мыслительной деятельности.

Значение подобного рода исследований для клиники является достаточно очевидным, поскольку оценка структуры расстройств мышления, играющая существенную роль при установлении диагноза, строится до сих пор главным образом на субъективном впечатлении врача.

Статистическое исследование синтаксических характеристик двух типов патологической речи представляет, как нам кажется, интерес не только для психиатрии, но и для лингвистики. Общеизвестно чисто лингвистическое и психолингвистическое значение исследований патологической речи как важного источника данных о системе языка и ее функционировании в норме. Наиболее очевидным при этом являются возможность определения степени устойчивости различных «узлов» механизма речи и возможность выявления наиболее

---

<sup>1</sup> См.: О. П. Красулина, М. И. Откупщикова, Ф. И. Случевский, Г. С. Цейтин, Б. А. Щербатов. Статистическое исследование синтаксических структур патологического мышления. В сб.: Психические заболевания, Л., 1970, стр. 303; Г. С. Цейтин. Статистическое исследование синтаксических структур в речевой продукции душевнобольных. В сб.: Математические методы в психиатрической клинике. Материалы конференции. Л., 1970, стр. 31.

слабых звеньев языковой системы, чаще всего страдающих при нарушении мышления.

Такая задача должна решаться в первую очередь путем сопоставления патологической и нормальной речи, хотя сопоставление отдельных разновидностей патологической речи, выявление речевой патологии, характерной для разных типов нарушения мышления, несомненно должны придать особую убедительность выводам относительно значимых отклонений от нормы.

Нас в данной работе будет, однако, в основном интересовать совершенно иной аспект проблемы лингвистического осмысления результатов исследования патологической речи.

Если подойти к исследуемым здесь образцам как к двум разновидностям патологической речи, двум «патологическим речевым стилям», то выявление признаков, наиболее диагностичных при идентификации типа патологии речи, может представить интерес и для общей типологии речевых стилей.

Существенным при этом нам представляется не фактическая характеристика исследуемых здесь «патологических стилей речи», а выявление среди большого числа более или менее случайного набора разнообразных синтаксических характеристик речи признаков, показывающих наибольшую меру различия, и, следовательно, наиболее значимых для дифференциации данных «речевых стилей».

Естественным продолжением такого исследования должна явиться попытка определения значимости выявленных «максимально диагностических» признаков для характеристики и различения стилей нормальной речи: функциональных стилей, индивидуальных стилей авторов.

В клинической традиции принято считать, что речевая продукция больных, страдающих эпилепсией с выраженными изменениями личности, характеризуется такими признаками, как многословие, обстоятельность, излишняя детализация повествования, увязание в мелких подробностях, повторение одних и тех же слов и выражений в пределах одного предложения. Эти особенности речи объясняются обеднением и застоем ассоциативной деятельности, утратой способности отделения главного от второстепенного, «прилипанием» к одному представлению. В отношении речевой продукции больных шизофренией принято считать, что она характеризуется относительной лаконичностью высказываний, совмещением явно несовместимых понятий, преобладанием номинативных функций над предикативными.

В табл. 1 приведены полученные после машинного анализа числовые характеристики 129 предложений (примерно 1600 слов) и указаны (и оценены количественно) наиболее характерные различия, наблюдающиеся в речи больных шизофренией и эпилепсией.

Следует отметить, что анализируемый материал представляет собой не случайную выборку, а совокупность предложений, отобранных психиатрами, как достаточно характерных, на их взгляд, для речи больных указанных групп. Авторы полностью осознают, что это в значительной мере снижает убедительность результатов анализа. Прделанная работа, однако, может представить интерес как опыт «объективизации» и уточнения интуиции психиатра. Кроме того, машинная обработка образцов патологической речи позволяет учесть целый ряд параметров, которые не учитывались ранее в психиатрической клинике и которые слишком трудоемки для ручной обработки лингвистами. В дальнейшем предполагается (как уже отмечалось в указанной выше статье Г. С. Цейтина<sup>2</sup>) провести машинную обработку большого случайно отобранного массива текстов с учетом требований лингвистической и статистической достоверности. Одной из дополнительных целей настоящей работы являлись экспериментальная проверка и отладка алгоритмов, которые будут использованы для такой обработки.

Таблица 1

Числовые характеристики особенностей речи больных шизофренией и эпилепсией

Признаки	Числовые характеристики		
	шизофрения	эпилепсия	мера различия*
Доля подлежащих среди всех слов** (/151)	0,14 ± 0,04***	0,088 ± 0,016	1,61
Доля местоимений в функции подлежащего (/151, 103)	0,072 ± 0,039	0,043 ± 0,007	0,51
Доля однородных групп дополнений (/152, 116)	0,0061 ± 0,0086	0,017 ± 0,01	0,62

<sup>2</sup> Г. С. Цейтин, ук. ст. в сб. Математические методы в психиатрической клинике. Материалы конференции. Л., 1970, стр. 34.

Признаки	Числовые характеристики		
	шизофрения	эпилепсия	мера различия*
Доля однородных сказуемых среди всех слов (/116, 291) . . . . .	0,0046±0,0071	0,023±0,009	2,34
Доля сказуемых среди однородных групп (116/291) . . . . .	0,11 ±0,15	0,42 ±0,2	1,33
Доля сочинительных союзов (/113) . . . . .	0,024 ±0,02	0,051±0,019	0,88
Доля подчинительных союзов (/114) . . . . .	0,031 ±0,035	0,075±0,026	1,06
Среднее количество подчиненных слов у инфинитива (107/601) . . . . .	1,65 ±1,23	3,96 ±0,27	1,08
Среднее количество подчиненных слов у причастия (108/601) . . . . .	0,75 ±0,96	1,67 ±0,36	0,8
Среднее количество слов, подчиненных фиктивному сказуемому**** (117/601) . . . . .	2,58 ±1,2	3,5 ±0,26	0,56
Доля однородных групп среди всех сказуемых (291/116) . . . . .	0,36 ±0,43	0,14 ±0,05	2,97
Среднее количество подчиненных у одного слова (/601) . . . . .	0,88 ±0,03	0,98 ±0,004	8,87
Средний размер группы местоимения (103/602)	1,26 ±0,4	2,47 ±1,3	0,79
Доля слов, прямо или косвенно подчиненных числительному (/105, 602) . . . . .	0,0046±0,0144	0,041±0,034	0,8
Доля слов, прямо или косвенно подчиненных глаголу (/106, 602) . . . . .	0,76 ±0,29	1,48 ±0,13	5,33
Средний размер финитно-глагольной группы (106/602) . . . . .	5,51 ±1,79	11,39 ±2,65	3,39
Средний размер группы инфинитива (107/601)	5,9 ±2,69	11,37 ±2,84	1,96



Признаки	Числовые характеристики		
	шизофрения	эпилепсия	мера различия*
Доля слов, входящих в состав однородной группы (/116, 602) . .	0,037 ± 0,29	1,14 ± 0,14	5,72
Среднее количество слов в составе однородной группы (116/602) . .	8,89 ± 2,29	21,3 ± 3,4	9,21
Средний размер группы у фиктивного сказуемого (117/602) . . .	6,17 ± 2,9	9,16 ± 0,58	1,02
Среднее расстояние до «вершины» (глубина в «дереве») (/702) . .	2,68 ± 0,3	4,73 ± 0,36	18,9
Сложность синтаксического дерева по гнездованию (/762, 331) .	0,084 ± 0,079	0,26 ± 0,1	1,86
Доля слов, прямо или косвенно подчиненных справа правосторонней связи (/762, 761) . .	0,3 ± 0,2	1,56 ± 0,49	5,55
Среднее количество слов, прямо или косвенно подчиненных справа правосторонней связи (762/761) .	0,32 ± 0,12	0,6 ± 0,07	4,06
Аналог глубины, по Ингве (/702, 234) . . .	1,1 ± 0,29	2,65 ± 0,33	12,3
Средняя длина связи в расчете на количество всех слов (/901) . . .	1,6 ± 0,22	2,84 ± 0,22	15,8
Средняя длина связи в расчете на все слова, кроме сказуемых (901/)	1,8 ± 0,2	2,91 ± 0,22	14,0
Среднее количество отождествлений (/603) .	0,017 ± 0,019	0,066 ± 0,033	1,7
Среднее расстояние между словом и антецедентом в расчете на все слова (/803) . .	0,056 ± 0,092	1,08 ± 0,52	3,7
Среднее расстояние между словом и антецедентом (803/) . . .	3,36 ± 1,57	16,3 ± 1,8	28,8

Признаки	Числовые характеристики		
	шизофрения	эпилепсия	мера различия*
Среднее расстояние от хозяина до подчиненного (средняя длина стрелки) (/606) . . . . .	0,82 ±0,21	2,03 ±0,13	14,5
Средний размер группы (/602) . . . . .	2,68 ±0,3	4,73 ±0,36	18,9
Доля слов, имеющих хозяина справа (/731) . . . . .	0,52 ±0,07	0,61 ±0,06	0,76
Средняя длина связи, направленной влево, в расчете на все слова (/961) . . . . .	0,74 ±0,17	1,29 ±0,18	4,8
То же, в расчете на такие связи (961/) . . . . .	2,23 ±0,26	3,68 ±0,42	5,47
Средняя длина связи, направленной вправо в расчете на все слова (/931) . . . . .	0,85 ±0,16	1,55 ±0,2	7,18
То же, в расчете на такие связи (931/) . . . . .	1,58 ±0,2	2,52 ±0,24	9,23
Количество право направленных стрелок на пути от сказуемого к данному слову (/702, 331) . . . . .	0,73 ±0,13	1,15 ±0,17	3,83

Примечания. \* Мера различия (М. Р.) вычислялась по формуле  $\frac{(a_1 - a_2)^2}{\sigma_1^2 + \sigma_2^2}$ , где  $a_1$  и  $a_2$  — средние, а  $\sigma_1$  и  $\sigma_2$  — стандартные отклонения.

\*\* В скобках приведен код признака, по которому можно установить его точный смысл. — См. статью: О. П. Красулина, М. И. Откупщикова, Ф. И. Случевский, Г. С. Цейтин, Б. А. Щербатов, ук. ст. В сб.: Психические заболевания. Л., 1970, стр. 303.

\*\*\* Первое число из пары, соединенной знаками  $\pm$ , обозначает среднее значение, второе число — стандартное отклонение от него.

\*\*\*\* Фиктивным сказуемым условно названы случаи, аналогичные закономерному опущению сказуемого в речи здоровых людей типа: «А он — в воду», «У нас уже лето» и т. д.

Приведенные числовые показатели наглядно подтверждают клинические данные, особенно в отношении речи больных эпилепсией. Так, о многословности высказываний свидетельствуют: показатель «глубины в дереве» (мера различия — М.Р. 18,9); средняя длина синтаксической связи в расчете как на количество всех слов (М.Р. 15,8), так и на все слова, кроме сказуемых (14,0); среднее расстояние от «хозяина» до «слуги» (М.Р. 14,5). Обстоятельность детализации повествования подтверждается средним размером группы (косвенные и прямые подчиненные + «хозяин» — М.Р. 18,9), средним количеством подчиненных у одного слова (М.Р. 8,87). Явное преобладание доли сочинительных (0,88) и особенно подчинительных союзов (1,06) также указывают на громоздкость синтаксических конструкций больных эпилепсией по сравнению с больными шизофренией.

Данные табл. 1 не только подтверждают клинические материалы о многословности предложений больных эпилепсией, но и указывают, какие части и члены предложения более подробно поясняются и «расшифровываются» больными. Весьма интересен, например, такой факт, что различие в «загрузке» глагольной группы оказалось наиболее высоким: М.Р. в оценках среднего размера глагольной группы — 3,39, а М.Р. среднего числа слов прямо или косвенно подчиненных глаголу — 5,33. Высока и доля однородных сказуемых среди всех слов (М.Р. 2,34), как и доля однородных групп среди всех сказуемых (М.Р. 2,97). Достаточно высоким оказался и средний размер группы фиктивного сказуемого: М.Р. 1,02 при среднем 9,16. Небезынтересна частота однородных групп, встречающихся в речи больных эпилепсией. Кроме указанных показателей однородных групп сказуемых, рассматривалась частота однородных групп дополнений, ее мера различия 0,62. М.Р. для средней доли слов однородной группы составляет 5,72, а для среднего количества слов, подчиненных однородной группе, — 9,21 (при среднем для эпилепсии 21,3 и шизофрении 8,89).

Необходимо отметить показатели насыщенности пояснений («загрузки») других частей речи и членов предложения. Так, средний размер группы инфинитива составляет для больных эпилепсией 11,37 при М.Р. 1,96, а М.Р. среднего количества подчиненных у инфинитива — 1,08, среднее количество подчиненных у местоимения для эпилепсии — 2,47, для шизофрении — 1,26 при М.Р. 0,79; М.Р. числа подчиненных слов у причастия и числительного оказалась одина-

ковой — 0,8. Последовательные правосторонние подчинения также свидетельствуют о стремлении к детальному описанию, как видно из табличных данных для признаков (/762, 761) и (762/761).

Повторение одних и тех же слов в пределах предложения, как свидетельство клинического признака «вязкости» мышления больных эпилепсией, подтверждается числовыми характеристиками отождествлений, среднее количество которых при эпилепсии — 0,066, а при шизофрении — 0,017 при М.Р. 1,7; но расстояние между словом и antecedentом в расчете на все слова при эпилепсии значительно выше (1,08), чем при шизофрении (0,056), при М.Р. 3,7. Более того, для эпилептического мышления весьма характерен такой показатель, как среднее расстояние между словом и antecedentом: 16,3 — для эпилепсии и 3,36 — для шизофрении при М.Р. 28,8 (!). Эти цифры могут интерпретироваться как объективное доказательство неизменного возвращения больного к представлениям, обозначенным словами, стоящими в начале предложения.

Немногословность, лаконичность высказываний больных шизофренией может быть подтверждена такими показателями, как среднее расстояние до «вершины дерева» (2,68), низкий показатель сложности синтаксической структуры (0,084), данные о средней длине синтаксической связи (1,6), среднем количестве подчиненных у одного слова (0,88) и другие, например доля сочинительных (0,024) и подчинительных (0,031) союзов. Удельный вес подлежащих в речи больных шизофренией значительно выше (0,14), чем у больных эпилепсией (0,088), что указывает на больший вес номинативных предложений и предложений с фиктивным (нулевым) сказуемым у больных шизофренией. Эти данные согласуются с тем, что фиктивных сказуемых в предложениях больных шизофренией — 31 (среднее 2,58), а у больных эпилепсией — всего 6.

Лингвистический интерес представляет тот факт, что признаки, прямо или косвенно отражающие основные конструкции языка, обладающие наибольшей частотностью, оказываются в наименьшей степени затронутыми патологическими изменениями мышления. Для некоторых таких признаков были подсчитаны средние и стандартные отклонения по всей группе больных (больные шизофренией и эпилепсией вместе). Параллельно было вычислено стандартное отклонение, которое получилось бы, если принять гипотезу, что появле-

ние данной конструкции в различных местах текста — независимые случайные события с одинаковой вероятностью. Обнаружилось, что стандартное отклонение, найденное по тексту, не выходит за пределы того, которое обусловлено случайностью (несмотря на то, что у больных сознательно отбирались предложения с наибольшими отклонениями от нормы).

Таблица 2

Данные для некоторых основных признаков по всей группе больных

Признаки	Средние и стандартные отклонения	Стандартные отклонения при гипотезе о независимых случайных событиях
Доля существительных (/101) . . . . .	0,26 ± 0,07	0,06
Доля глаголов (/106) . . . . .	0,132 ± 0,039	0,044
Доля наречий (/110) . . . . .	0,065 ± 0,027	0,033
Доля предлогов (/111) . . . . .	0,109 ± 0,038	0,040
Доля сказуемых среди личных глаголов (106/291) . . . . .	0,73 ± 0,16	0,16
Доля личных глаголов среди сказуемых (291/106) . . . . .	0,59 ± 0,15	0,15
Доля существительных среди дополнений (152/101) . . . . .	0,50 ± 0,18	0,17
Доля прилагательных среди согласованных определений (153/102) . . . . .	0,50 ± 0,22	0,20
Доля существительных среди несогласованных определений (154/101) . . . . .	0,87 ± 0,15	0,18

С другой стороны, признаками, показавшими наибольшую меру различия и оказавшимися наиболее диагностичными для различения исследованных разновидностей патологической речи, оказались следующие:

- 1) средняя длина анафорической связи (М.Р. 28,8);
- 2) средний размер группы (М.Р. 18,9);
- 3) глубина в дереве (среднее расстояние до вершины) (М.Р. 18,9);
- 4) средняя длина связи (М.Р. 15,8);

- 5) средняя длина стрелки (М.Р. 14,5);
- 6) аналог глубины по Ингве (М.Р. 12,3).

Интересно отметить, что все эти признаки относятся к синтаксической структуре предложения и естественно интерпретируются как меры общей сложности синтаксической структуры. Таким образом, приведенные статистические данные показывают, что характеристики синтаксической структуры речи могут иметь значение для исследования патологии мышления.

Т. Д. Логачева, М. С. Першикова

### ОПИСАНИЕ ЭКСПЕРИМЕНТА ПО «УПРОЩЕННОМУ» СИНТАКСИЧЕСКОМУ АНАЛИЗУ РУССКИХ НАУЧНО-ТЕХНИЧЕСКИХ ТЕКСТОВ

Синтаксический анализ рассматривается как один из этапов автоматического перевода текста с одного языка на любой другой (естественный или искусственный). Цель синтаксического анализа — определение синтаксической структуры входного предложения в виде «дерева зависимостей»<sup>1</sup>.

Выделяются два уровня работы алгоритма синтаксического анализа: предварительный, или упрощенный, и расширенный.

Проведенный эксперимент представляет собой машинную реализацию работы алгоритма на уровне упрощенного анализа, где производится лишь предварительная обработка предложения. Наличие этого этапа ускорит работу более сложной грамматики, которая должна быть использована при расширенном анализе.

Грамматические сведения, используемые алгоритмом, следующие: 1) система представления информации о слове, 2) правила, характеризующие допустимые связи между словами (упрощенный вариант грамматики зависимостей).

<sup>1</sup> Подробнее о «дерева зависимостей» см.: С. Я. Фитналов. Моделирование синтаксиса в структурной лингвистике. В сб.: Проблемы структурной лингвистики. М., 1962.

**Система представления информации о слове.** На входе этапа синтаксического анализа предложение задается в виде последовательности информации<sup>2</sup> о словах, которая соответствует последовательности слов в предложении. Под «словом» будем понимать множество всех его омонимов. Порядок омонимов при каждом слове — произвольный. Информация о слове представляется списком признаков. В начале каждого списка стоит признак, отмечающий класс слова. В большинстве случаев класс слова соответствует традиционной части речи, например, класс существительных, класс предлогов, класс наречий. В других случаях класс слова — это более широкое понятие, чем часть речи. В один класс мы объединяем слова, которые имеют общую синтаксическую функцию. Например, в один класс входят прилагательные, порядковые числительные, притяжательные местоимения — «мой», «твой», «наш», «ваш» и другие, т. е. слова, которые являются в предложении согласованным определением к имени. В один класс объединяем сочинительные союзы и запятую, глагол и глагольные формы. Всего выделено 10 классов слов, имеющих следующие символические обозначения: 1) S — существительные; 2) A — прилагательные, кванторные местоимения, притяжательные местоимения, порядковые числительные; 3) V — глаголы, причастия, деепричастия; 4) N — местоимения; 5) Q — количественные числительные; 6) B — наречия; 7) Ч — частицы; 8) F — предлоги; 9) E — сочинительные союзы, запятая; 10) J — подчинительные союзы.

При каждом символе, обозначающем класс слова, может быть один или несколько признаков. По функциональным особенностям эти признаки можно разделить на две группы.

I группа — признаки, являющиеся ограничителями класса слова. С их помощью выделяются подклассы классов слов. К ним относятся: а) признаки, имеющие условное название «индивидуальные номера» (символическое обозначение Ц). Индивидуальные номера выделяют подклассы, которые могут состоять как из нескольких слов, так и из одного слова (Ц1 — количественные числительные от 2 до 4; ЦЗ — относительные местоимения «кто», «что»; Ц15 — при-

<sup>2</sup> Информацию, с которой работает алгоритм синтаксического анализа, каждое слово получает в результате морфологического анализа. — См.: Г. С. Ульянова. Построение алгоритма морфологического анализа для русского языка на основе конечноавтоматной модели (наст. сб., стр. 145).

лагательное «должный» и т. д.); б) признаки с условным названием «семантические» (символическое обозначение С). К семантическим относятся признаки, отмечающие модальность (глаголов и наречий) — СЗ, притяжательные местоимения «мой», «твой», «наш», «ваш» (С1), кванторные местоимения «любой», «всякий», «каждый» (С2) и т. д.

Ко II группе относятся признаки, определяющие грамматическую форму слова. Это признаки с условным названием «синтаксические»<sup>3</sup> (символическое обозначение Г и Т). Синтаксическими являются признаки, отмечающие краткую форму прилагательных и причастий (Т10), признаки падежей (Т1—Т6), признаки числа (Г17, Г20) и т. д.

Некоторые из признаков могут встречаться одновременно при нескольких классах слов. Например, признак падежа получают существительные, прилагательные, предлоги. Другие же учитываются только при одном каком-либо классе. Например, признак времени учитывается только у глагола.

Всего был выделен 51 признак: **IA. «Индивидуальный номер»:** 1) Щ1 — количественные числительные от 2 до 4 — QЩ1; 2) Щ2 — притяжательные местоимения («его», «ее», «их») — NЩ2; 3) Щ3 — относительные местоимения («кто», «что») — NЩ3; 4) Щ4 — относительное прилагательное «который» — AЩ4; 5) Щ5 — указательное местоимение «это» — NЩ5; 6) Щ6 — глагол «есть» (как связка) — VЩ6; 7) Щ7 — глагол «быть» (и его формы в будущем, прошедшем времени) — VЩ7; 8) Щ11 — «всё» — NЩ11; 9) Щ12 — , — EЩ12; 10) Щ13 — сочинительные союзы — EЩ13; 11) Щ15 — прилагательное «должный» — AЩ15. **IB. «Семантические» признаки:** 12) С1 — притяжательность (местоимения «мой», «твой», «наш», «ваш») — AC1; 13) С2 — кванторность (местоимения «любой», «всякий», «каждый») — AC2; 14) С3 — модальность (глаголов и наречий) — VC3, BC3; 15) С4 — фазовые глаголы — VC4; 16) С5 — полусвязки («являться», «казаться», «становиться») — VC5; 17) С6 — каузативные глаголы («позволять», «решать» и т. д.) — VC6; 18) С10 — порядковые числительные — AC10; 19) С11 — количественные местоимения («много», «несколько») — NC11; 20) С12 — глаголы существования («наблюдаться», «существовать») — VC12; 21) С13 — отглагольные существительные — SC13; 22) С33 — «более», «менее» — BC33; 23) С40 —

<sup>3</sup> Следует отметить, что сюда входят и признаки, которые традиционно считаются морфологическими (время, число, падеж).



местоименные прилагательные — АС40. II. Синтаксические признаки: 24—29) T1—T6 — именительный—предложный падежи; 30) T7 — инфинитив; 31) T10 — краткая форма прилагательных и причастий; 32) T12 — причастие; 33) T13 — деепричастие; 34) T14 — мужской род; 35) T15 — женский род; 36) T16 — средний род; 37) T17 — одушевленность; 38) T17 — единственное число; 39) T20 — множественное число; 40) T21 — I лицо; 41) T22 — II лицо; 42) T23 — III лицо; 43) T24 — сравнительная степень; 44) T25 — превосходная степень; 45) T26 — настоящее время; 46) T27 — прошедшее время; 47) T30 — будущее время; 48) T31 — несовершенный вид; 49) T32 — вводность; 50) T34 — возвратность; 51) T37 — переходность.

**Описание грамматики.** Как уже говорилось выше, на первом этапе работы алгоритма синтаксического анализа использовался упрощенный вариант грамматики зависимостей (условно эту грамматику мы называем «миниграмматика») <sup>4</sup>. В соответствии с требованием проективности, предъявляемым грамматикой зависимостей, на синтаксическую структуру предложения накладывается ряд ограничений. Эти ограничения служат причиной некоторых условностей, принятых на основе договоренности. Используются следующие соглашения.

1. Вершиной «дерева» считаем сказуемое (у каждого «дерева» может быть только одна вершина) <sup>5</sup>: 1) «Головка смещается посредством каретки». В составном именном сказуемом вершиной считается связка: «Причиной этого является образование...»; в составном глагольном — смысловая часть, выраженная инфинитивом или причастием (в сложной форме будущего времени вершиной считается инфинитив): «Источник постоянного тока может **служить** для...», «Поток газа может быть **подан** через...», «Сварка будет **осуществляться** методом...».

2. Подлежащее в предложении зависит от вершины.

3. Сочинительные конструкции (однородные члены и сложносочиненные предложения, которые соединяются сочинительным союзом или запятой) представлены следующим образом: главным словом сочинительной конструкции счи-

<sup>4</sup> Грамматика строилась под руководством С. Я. Фитналова и при консультациях М. И. Откупщиковой.

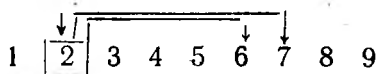
<sup>5</sup> Слова, выделенные полужирным, являются вершиной «дерева».

тается первое слово, остальные последовательно зависят от него и друг от друга (первым словом сочинительной конструкции в сложносочиненном предложении считается вершина первого предложения). Сочинительный союз или запятая в роли сочинительного союза подчиняются предшествующему слову. В сложносочиненных предложениях запятая в роли союза или союз подчиняются вершине первого предложения<sup>6</sup>:

1) быстрый и локализованный (нагрев);

2) рамы, крыши и стенки (вагонов);

3) устройство отличается отсутствием подвижных контактов и обладает хорошим быстродействием.



4. Неоднородные определения (когда отсутствует союз или запятая — формальные показатели однородности) подчиняются каждый определяемому слову:

электронные цифровые машины.

5. Контактные частицы (препозитивные и постпозитивные) подчиняются непосредственно соседнему слову. Препозитивные — слову справа, постпозитивные — ближайшему слову слева: «...молекулы взаимодействуют с уже усиленным полем». «...единство же подхода к анализу вырабатывает...»

6. Неконтактные частицы подчиняются вершине предложения:

«Однако определение имеет значение...»

<sup>6</sup> Стрелка направлена на зависимое слово.

7. В сложноподчиненных предложениях, если предложения соединяются союзом<sup>7</sup>, вершина главного предложения управляет союзом и союз управляет вершиной придаточного предложения. Если же предложения соединяются союзным словом, то слово главного предложения, заместителем которого является союзное слово, управляет вершиной придаточного предложения, остальные связи в придаточных предложениях обычны (т. е. аналогичны связям в главном предложении), союзное слово управляется по общим правилам, как полнозначное:

1            2                    3            4    5 6    7    8            9  
 «При исследовании было выяснено, что во время сдавли-  
                   10            11    12            13                            14                    15  
 вания образцов окислы с контактной поверхностью выжи-  
                   16  
 маются постепенно».

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

1            2                    3                            4                    5  
 «В современном самолетостроении используются материа-  
 6 7            8            9                    10  
 лы, которые не удается обработать».

1 2 3 4 5 6 7 8 9 10

8. Вводные слова подчиняются вершине предложения.

Принятые соглашения неполны. В данной грамматике не рассматриваются некоторые случаи непроективности (случаи связок в именном сказуемом), сравнительные конструкции (с союзом чем), не решается вопрос о выборе «хозяина» для предметных конструкций, когда они формально могут относиться и к существительному, и к глаголу и др.

Помимо направленности, связи в «дереве зависимостей» распределяются на несколько типов<sup>8</sup>, разбивающихся на две

<sup>7</sup> При двойных союзах (типа «если... то») оба компонента подчиняются вершине главного предложения. При сложных союзах (типа «с тем, чтобы») вершина главного предложения управляет первым словом, а дальше управление идет последовательно вправо (т. е. первое слово управляет вторым, второе — третьим и т. д.).

<sup>8</sup> Все типы связи: примыкание, согласование, управление считаем вариантами подчинительной связи.

группы. Первая группа типов связей выделяется на основе обычно выделяемых синтаксических связей: 1) определение (к имени), согласованное и несогласованное; 2) дополнение (выраженное беспредложной формой); 3) предложное управление (существительные, управляемые предлогом); 4) однородность (связи внутри групп однородных членов); 5) сказуемое придаточного предложения (придаточные предложения в различных функциях); 6) предикатив (именная часть сказуемого); 7) вспомогательная часть составного глагольного сказуемого (вспомогательные глаголы в личной форме и модальные слова); 8) подлежащее. Вторая группа типов связей выделена на уровне частей речи: 9) предлог (вводящий зависимую именную группу); 10) частицы («уточнение» для различных классов слов); 11) наречие («определение» к классу слова); 12) деепричастие.

На основе вышеописанной классификации связей «дерева зависимостей» была составлена миниграмматика, включающая 202 правила<sup>9</sup>. В миниграмматике рассматриваются простейшие свойства синтаксически связанной пары слов и используется приведенная выше система представления информации о слове. Миниграмматика лишь фиксирует возможность синтаксической связи между парой омонимов разных слов.

Правила грамматики представляются в виде формулы, состоящей из двух частей, соединенных стрелкой (A → B). Стрелка обозначает возможность управления одного слова другим и направлена на зависимое слово. Парная связь (т. е. связь между двумя словами) считается возможной при условии, что 1) один из омонимов включает информацию, которая должна быть у подчиняющего слова («хозяина»); 2) другой включает информацию, которая должна быть у подчиняемого слова («слуги»). Как «хозяин», так и «слуга» могут быть представлены одним символом, обозначающим класс слова, или конъюнкцией признаков и их отрицаний. Например, глагол, который может быть вершиной «дерева зависимостей», представляется конъюнкцией признаков и их отрицаний:  $V_{t12t13c3c4c6}$  — слово, относящееся к глагольному классу, но не причастие, не деепричастие, не модальный глагол, не фазовый глагол, не каузативный глагол.

Для каждого типа связи учитывается его конкретная реализация в предложении, зависящая от способов выраже-

<sup>9</sup> Грамматика занимает 1212 ячеек.

ния или от вида управляющего и управляемого слова. Например, для 7-го типа связи — вспомогательная часть составного глагольного сказуемого — отмечается, что управляющее слово может быть выражено инфинитивом и кратким причастием, а управляемое слово может быть модальным глаголом или наречием, фазовым глаголом, каузативным глаголом, формами глагола **быть**, краткой формой прилагательного **должный**.

Правила в миниграмматике на 7-й тип связи следующие: 1.  $U_{с3} \leftarrow V_{т7}$  инфинитив управляет влево модальными глаголами; 2.  $V_{с3} \leftarrow V_{т7}$  — модальными наречиями; 3.  $V_{с6} \leftarrow V_{т7}$  — каузативными глаголами; 4.  $V_{с4} \leftarrow V_{т7}$  — фазовыми глаголами; 5.  $A_{ш15 т10} \leftarrow V_{т7}$  — краткой формой прилагательного **должный**; 6.  $V_{с12 ш7} \leftarrow U_{т12 т10 г17}$ ; 7.  $V_{с12 ш7 г20} \leftarrow V_{т12 т10 г20}$  краткие причастия управляют формами глагола **быть**.

**Результаты эксперимента.** Используемый алгоритм<sup>10</sup> предварительного анализа дает следующую информацию:

1) имеет ли предложение хотя бы один допустимый<sup>11</sup> анализ;

2) если да, то какие из управлений встречаются хотя бы в одном из допустимых анализов и какие из омонимов служат вершиной хотя бы в одном из допустимых анализов (некоторое управление или вершина, допустимые по отдельности, могут не попасть ни в какой допустимый анализ);

3) какие из омонимов оказываются выбранными хотя бы в одном из допустимых анализов (если они есть); если омонима нет ни в одном допустимом анализе, то он выбрасывается (частичное решение омонимии);

Для машинного эксперимента было подготовлено 80 предложений, взятых из реферативных журналов по сварке. Из них обработано 70 предложений.<sup>12</sup>

<sup>10</sup> Данный алгоритм составлен Г. С. Цейтиным. — См.: Г. С. Цейтин. Алгоритм для упрощенного синтаксического анализа. В сб.: Проблемы кибернетики, вып. 24. М., 1971.

<sup>11</sup> Анализ считается допустимым, если выполняются следующие условия: 1) условие проективности; 2) косвенного подчинения вершине; 3) если слово  $x$  подчинено слову  $y$ , и  $y$  слова  $x$  выбран омоним  $a$ , а  $y$  слова  $y$  — омоним  $o$ , то по правилам омоним  $a$  может подчиняться омониму  $o$ .

<sup>12</sup> Причины, из-за которых не обработалось 10 предложений, были найдены, но соответствующие (уже известные) исправления не были внесены в грамматику по техническим причинам. На анализ одного предложения затрачивается около 30 сек.

В качестве примера обработанного предложения рассмотрим фразу: «Объемные резонаторы протонных линейных ускорителей должны отвечать жестким техническим условиям». Это предложение на входе до работы алгоритма предварительного анализа имело следующий вид: 1) объемные  $A_{Г20 T1}$ ; 2) объемные  $A_{Г20 T4}$ ; 3) резонаторы  $S_{Г20 T1}$ ; 4) резонаторы  $S_{Г20 T4}$ ; 5) протонных  $A_{Г20 T6}$ ; 6) протонных  $A_{Г20 T2}$ ; 7) линейных  $A_{Г20 T2}$ ; 10) линейных  $A_{Г20 T6}$ ; 11) ускорителей  $S_{Г20 T2}$ ; 12) должны  $A_{Щ15 Г20 T10}$ ; 13) отвечать  $V_{T7}$ ; 14) жестким  $A'_{Г20 T3}$ ; 15) жестким  $A_{T14 Г17 T5}$ ; 16) жестким  $A'_{T16 Г17 T5}$ ; 17) техническим  $A_{Г20 T3}$ ; 20) техническим  $T_{14 Г17 T5}$ ; 21) техническим  $A'_{T16 Г17 T5}$ ; 22) условиям  $S_{Г20 T3}$ .

После работы алгоритма на выходе синтаксического анализа получается следующий анализ: 13-й омоним — вершина предложения; 3-й омоним управляет 1-м омонимом;  $4 \rightarrow 2$ ;  $11 \rightarrow 7$ ;  $11 \rightarrow 5$ ;  $13 \rightarrow 12$ ;  $22 \rightarrow 17$ ;  $22 \rightarrow 14$ ;  $14 \rightarrow 17$ ;  $13 \rightarrow 22$ ;  $5 \rightarrow 7$ ;  $4 \rightarrow 11$ ;  $3 \rightarrow 11$ ;  $12 \rightarrow 3$ ;  $13 \rightarrow 4$ ;  $13 \rightarrow 3$ .

В этом предложении 8 «лишних» (с точки зрения «правильного» анализа) омонимов (омонимы под номерами 2, 4, 6, 10, 15, 16, 20, 21). Из них обработались 2 омонима (2-й и 4-й), остальные не вошли ни в один из допустимых анализов (частичное решение омонимии). В результате работы алгоритма получились «лишние» управления, под которыми понимаются связи, возможные с точки зрения миниграмматики, но не имеющие места на самом деле (с точки зрения принятых договоренностей). «Лишними» в этом предложении являются связи между 14-м и 17-м омонимами, так как отсутствует признак однородности, между 5-м и 7-м омонимами (также отсутствует признак однородности), между 12-м и 3-м омонимами (12-й омоним —  $AT_{10}$  может быть вершиной, но в данном предложении устанавливается другая вершина, т. е.  $AT_{10}$  — допустимая вершина, но не вошедшая ни в какой допустимый анализ).

Задача устранения «лишних» связей и «лишних» омонимов решается на следующем этапе синтаксического анализа (при расширенном анализе).

## СПОСОБЫ ВЫРАЖЕНИЯ СЕМАНТИЧЕСКИХ ОТНОШЕНИИ В ИНФОРМАЦИОННОМ ЯЗЫКЕ

Настоящая работа посвящена описанию конкретного информационного языка и способов выражения в нем семантических отношений<sup>1</sup>.

В качестве информационного массива было использовано тематическое направление «Сварка в судостроении» и индексируется 500 рефератов, соответствующих выбранной тематике; глубина индексирования в среднем составляет 30—40 терминов на реферат. На основе этого массива был создан дескрипторный словарь.

При составлении словаря использовались всевозможные справочные и учебные пособия по сварке, политехнические и терминологические словари, а также ГОСТы.

Дескрипторный словарь представляет собой список дескрипторов, расположенных в алфавитном порядке. Дескрипторы — это употребляющиеся в данной области техники более или менее стандартизированные понятия.

Ниже дается фрагмент дескрипторного словаря. В первой графе его располагается номер дескриптора или ключевого слова по порядку, во второй — пишется машинный код, в третьей — собственно словарная статья дескриптора, в четвертой — номера рефератов, в которых встретился данный дескриптор, и, наконец, в пятой — номера терминов, в определении которых встретился данный дескриптор. Данные в графах 2, 4, 5 используются для машинного эксперимента.

Рассмотрим подробнее графу 3 «Термин». Она состоит из основного дескриптора «испытание механических свойств», под которым приводятся список ключевых слов, являющихся синонимами основному дескриптору (син.), родовой дескриптор (р.) и список видовых дескрипторов (в.). От ключевых слов даны отсылки к соответствующим основным дескрипторам.

<sup>1</sup> Модель информационного языка описана в работе: Г. М. Ильин, Н. Н. Ильина, И. И. Матвеева. Модель системы поиска технической литературы по рефератам. НТИ, сер. 2, Информационные процессы и системы, 1970, № 8, стр. 24—29.

№ п. п.	Ма- шин- ный номер	Т е р м и н	№ рефера- тов, со- держа- щих термин	№ терми- нов, в опре- делении ко- торых встре- чается термин
171	4065	Испытание механических свойств син. Испытания механические син. Испытания на механиче- ские свойства  син. Испытания технологических свойств  р <sub>1</sub> Контроль качества  в. Испытания динамические  в. Испытания на длительную прочность	132, 137	169, 174, 182, 183, 189, 194, 195
178	4065	Испытания на механические свой- ства  см. Испытания механических свойств		
229	0174	Контроль качества  в. Испытания на коррозионную стойкость  в. Испытания механических свойств  в. Испытания на плотность	21, 56, 97 98, 148, 200	171, 177 181

Заметим, что термин «синоним» употребляется не в лингвистическом смысле этого слова, а в смысле сложившегося употребления терминов в данной области знания. Всего в словаре содержится 1012 дескрипторов и ключевых слов, из них последних — 136. Словарь не претендует на исчерпывающую полноту. Напротив, в ходе экспериментального опробования информационной системы дескрипторный словарь должен пополняться.

В соответствии с моделью информационного языка дескрипторы делятся на простые и сложные. В словаре сложные дескрипторы отмечены звездочкой. Простые дескрипто-

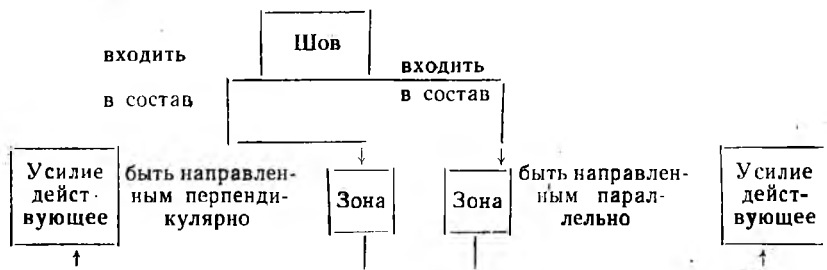


ры — это исходные понятия данной области знания, на основании которых строятся все остальные понятия данной области. Сложные дескрипторы — это такие понятия, которые получаются из простых путем терминологического определения последних. Определения сложных терминов, записанные на информационном языке, называются развертками сложных терминов.

**Пример.** Сложный дескриптор «шов комбинированный».

**Определение.** Сварной шов, часть которого расположена параллельно направлению действующего усилия, а часть — перпендикулярно направлению действующего усилия.

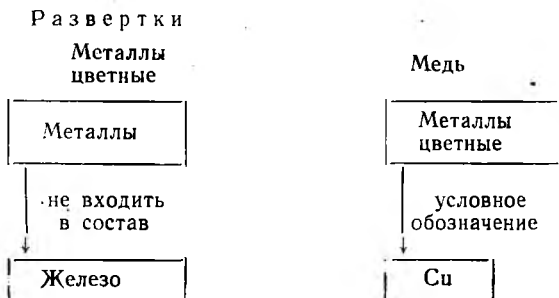
Развертка сложного дескриптора:



Сложные дескрипторы в свою очередь могут участвовать в определениях других сложных дескрипторов.

**Пример.** Металлы — простой дескриптор. Металлы цветные — сложный дескриптор. Медь — сложный дескриптор.

В определении дескриптора «металлы цветные» участвует простой дескриптор «металлы». Дескриптор «медь» определяется в свою очередь через сложный дескриптор «металлы цветные».

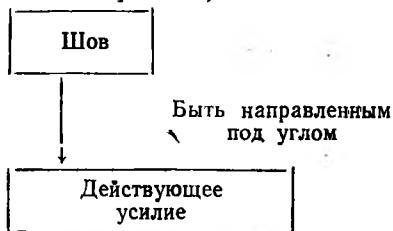


Принцип разделения дескрипторов на простые и сложные позволяет учесть достаточно полно родо-видовые отношения, существующие между понятиями в данной области знания.

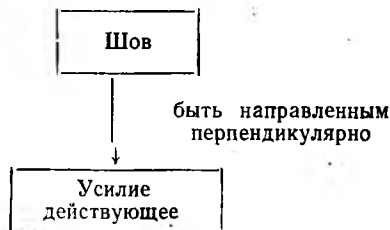
Каждое определение сложного дескриптора содержит ближайшее родовое понятие, а также набор видовых признаков, по которым данный дескриптор отличается от других дескрипторов, входящих в данное родовое гнездо.

В родовое гнездо «сварные швы» входят видовые дескрипторы «шов косой», «шов лобовой», «шов фланковый», «шов комбинированный» и т. д.

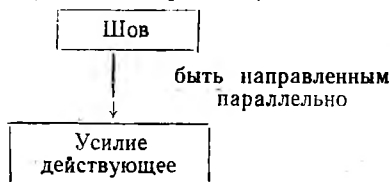
Шов косой определяется как сварной шов (родовое понятие), расположенный под углом к направлению действующего усилия (видовой признак).



Шов лобовой определяется как сварной шов (родовое понятие), расположенный перпендикулярно направлению действующего усилия (видовой признак).



Шов фланковый определяется как сварной шов (родовое понятие), расположенный параллельно направлению действующего усилия (видовой признак).



Определение шва комбинированного вместе с соответствующей разверткой приведено выше. Определение содержит два видовых признака.

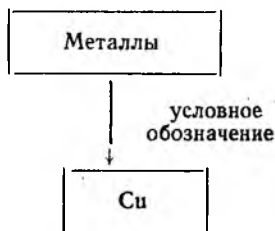
Видовые признаки дескрипторов, входящих в одно и то же родовое гнездо, могут быть названы как в вышеприведенных примерах, а может только указываться на их наличие. Для этого употребляется искусственный технический прием, заключающийся в том, что дескрипторы, входящие в родовое гнездо, различаются по видовому признаку, носящему условный характер.

Такие дескрипторы определяются через ближайшее родовое понятие с помощью условного символа, указывающего на наличие видового признака. Условный символ присоединяется к родовому понятию отношением «условное обозначение».

**Пример.** В родовое гнездо «металлы» входит целый ряд дескрипторов, различающихся по условному видовому признаку. «Железо» записывается так:



«Медь» записывается так:



Дескрипторный словарь является способом фиксирования базисных отношений (родо-видовых, синонимических и т. д.). Связи между понятиями в тексте реферата и запроса фиксируются с помощью специальных отношений, которые толь-

ко указывают на сам факт наличия связи между понятиями. Семантика же их никак не учитывается, поэтому названиям отношений не следует придавать точного значения. Используемые для записи рефератов на информационном языке отношения служат в основном для связи понятий при описании технологических процессов. Для понятий, относящихся к области технологических процессов, употребляется небольшое число характерных связей.

Примеры наиболее употребительных связей между понятиями и соответствующие им отношения информационного языка:

Субъектная связь — отношение «разработчик».

Объектная связь — отношение «исходные данные».

Результативная связь — отношение «результат».

Атрибутивная связь — отношение «характеристика».

Инструментальная связь — отношение «оборудование».

Обуславливающая связь — отношение «условие».

Связи между понятиями в текстах рефератов и запросов могут выражаться на информационном языке с помощью узла «зависимость». Узел употребляется для отражения в информационном языке факта наличия зависимости между объектами или влияния одного объекта на другой. Связи между понятиями, выражаемые узлом «зависимость», носят большей частью логический характер, соответствуя на языковом уровне различным предикатам причины, условия, тождества, следствия и т. д.

**Примеры** употребления узла «зависимость».

Реферат 153<sup>2</sup>: «Показана зависимость технологической прочности при сварке от формы кристаллитов шва».

Реферат 155: «Термическая обработка сварных соединений повышает их прочность...»

В модели информационного языка выделено несколько типов зависимостей. Конкретный материал позволяет дополнить список типов зависимостей.

**Примеры** модельных типов зависимостей.

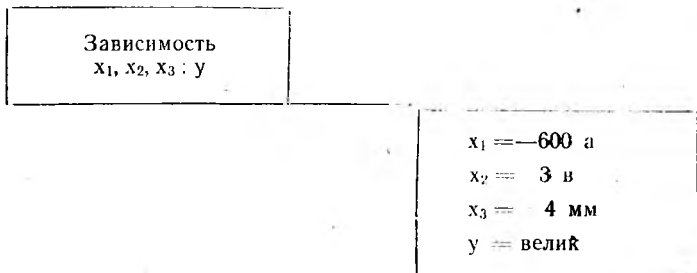
1. Значения зависимых переменных при фиксированных значениях независимых переменных.

«При автоматической сварке под флюсом наибольшего значения КПД достигает при следующем режиме сварки:

---

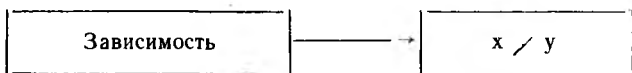
<sup>2</sup> Рефераты брались из реферативного журнала «Сварка». Номер реферата является порядковым номером в памяти ИПС.

ток 600 а, напряжение дуги 3 в, диаметр проволоки 4 мм» (Реферат 66).



2. Характер изменения одной из зависимых переменных при изменении независимой переменной:

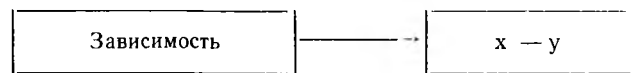
а) зависимая переменная (у) при возрастании независимой переменной (х) возрастает (х/у): «При многослойной сварке повышается пластичность металла шва» (Реферат 44).



б) зависимая переменная при возрастании независимой переменной убывает (х \ у): «При многослойной сварке разброс  $a_k$  металла шва существенно снижается» (Реферат 44).



в) зависимая переменная при изменении независимой переменной существенно не меняется: «Термообработка не оказывает существенного влияния на повышение прочности сварного соединения» (Реферат 143).

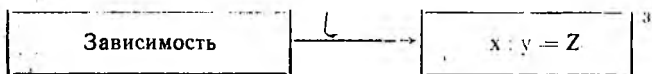


3. Взаимоотношения между двумя переменными: «Состав электродной проволоки близок к составу основного металла» (Реферат 142).

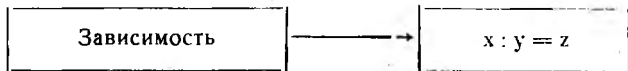


4. Сообщается о существовании некоторой зависимости между параметрами объекта, но сущность ее не раскрыва-

ется. Случай довольно-таки часто встречающийся, поскольку ИПС имеет дело с рефератами: «Влияние способа обработки поверхности на прочность сварного соединения» (Реферат 50).



«Обсуждается зависимость между остаточным напряжением и материалом сварных конструкций» (Реферат 50).



**Примеры** немодельных зависимостей. Реферат содержит несколько объектов. В этом случае может существовать зависимость между параметрами, входящими в разные объекты. Причем зависимость отражается в сравнении как параметров, входящих в разные объекты, так и самих объектов.

1. «Основное время сварки самофлюсующимся присадочным металлом сокращается на 40 процентов по сравнению со временем сварки с использованием флюсов» (Реферат 116). В реферате 116 говорится о двух объектах — «сварке самофлюсующимся присадочным металлом» и «сварке под обычным флюсом», в которых сравнивается параметр «время сварки».

2. «По сравнению с автоматической дуговой сваркой под флюсом производительность сварки с электродом повысилась в 2 раза» (Реферат 158).

3. «Рассмотрены преимущества плазменной сварки по сравнению с газозлектрической сваркой» (Реферат 129).

**Т. Н. Никитина, М. И. Откупщикова**

### СЛЕДОВАНИЕ ИЗ ПРОСТОГО ПРЕДЛОЖЕНИЯ (К ВОПРОСУ О СЛАБОУПРАВЛЯЕМЫХ ЭЛЕМЕНТАХ В ПРОСТОМ ПРЕДЛОЖЕНИИ)

При построении информационной системы «запрос—ответ» установление набора содержательных выводов, «следующих»

<sup>3</sup> Символ  $Z$  заменяет конкретное значение некоторых поддереьев в случае, если в реферате это значение не приведено, но упоминается, что в первичном документе оно имеется.

из каждого предложения, служит той основой, на базе которой конструируются ответы на запрос.

Задача исчерпывающего описания следования, даже если ее ограничить рамками простого предложения, предполагает наличие подробно разработанных концепций и исследований по многим вопросам семантики языка, чего на самом деле еще нет. Поэтому авторы в настоящей статье ограничились рассмотрением довольно узкого вопроса — о выводах из простого предложения, содержащего так называемые слабоуправляемые члены предложения.

В современной грамматике русского языка нет объективных критериев для противопоставления сильного управления слабому. Обычно при разграничении этих двух видов дополнений (или обстоятельств) предлагается один из двух принципов: статистический или семантический. Статистический — элемент, зафиксированный при данном глаголе чаще других или превышающий некоторый «порог частотности» (например, элемент, встретившийся более чем в 50% случаев употребления данного глагола в некотором тексте), считается сильноуправляемым.

Семантический — сильноуправляемым считается элемент, наиболее тесно связанный семантически с данным глаголом, т. е. тот, который «предсказывается» семантикой данного глагола.

Нетрудно видеть, что ни один из этих подходов не дает нам объективной процедуры различения сильного и слабого управлений. В первом случае сама процедура различения строгая, но выбор «порога встречаемости» совершенно произволен, во втором — понятие большей или меньшей семантической близости никак не определяется, однако оно и не является интуитивно достаточно ясным. Поясним последнее положение примером. При глаголе «посылать» семантикой этого глагола «предсказывается» довольно много элементов (мест): кто, кого, куда, зачем (с какой целью), в качестве кого, на какой срок и т. д.: «Его послали в командировку на десять месяцев», «Его послали в Москву делегатом на конференцию», «Его послали со студентами руководить их практикой» и т. д. Какие из этих элементов считать сильноуправляемыми, исходя из указанных выше критериев, сказать трудно, тем более, что словари, отмечающие глагольные управления, отличают разное число мест при глаголе «посылать».

Еще одним способом описания слабого управления может быть составление списка слабоуправляемых мест каждого конкретного глагола. Однако это представляется равносильным попытке перечислить все возможные смысловые связи данного глагола со всеми другими словами из словаря, которые могут семантически с ним сочетаться.

Предлагается иной подход к проблеме слабого управления: слабоуправляемым элементом можно считать такой элемент, который присоединяется к глаголу, являющемуся ядром некоторой глагольной конструкции, с помощью особой логической связки (коннектора). Тогда сильноуправляемым элементом будет такой элемент, который непосредственно занимает «место» того или иного глагола (предиката) без особой промежуточной связки.

В смысловой записи, т. е. в представлении текста во внутренней структуре, простое предложение со слабоуправляемым элементом может, таким образом, рассматриваться как сложная конструкция, содержащая как минимум 3 предиката — сказуемое простого предложения, предикат, соответствующий логической связке, и предикат, восстановленный в слабоуправляемом члене.

Лингвистически это можно интерпретировать как развертывание слабоуправляемого члена в придаточное предложение, а всего простого предложения — в сложное. Тогда возможность трансформации группы слабоуправляемого элемента в придаточное предложение будем считать языковым приемом определения слабоуправляемого члена при условии, что союз придаточного предложения выражает определенную логическую связь. Например, оборот с предлогом «из-за» может быть трансформирован в придаточное предложение причины. Поэтому его можно считать «сверткой» причинного предложения в простом предложении. У каждого типа придаточного предложения есть свои «свертки», эквивалентные придаточному предложению по смыслу и по правилам вывода.

Следует отметить, что возможность трансформации слабоуправляемого члена в придаточное предложение мы рассматриваем лишь как способ его выявления. Вопрос о целесообразности развертки того или иного типа предложений со слабоуправляемыми элементами для перевода их во внутреннюю структуру должен быть рассмотрен отдельно, особенно нецелесообразна развертка для элементов, которые из предложения не следуют (например, части сложного предложе-



ния с придаточным условия, часть, соответствующая второму месту предиката цели, и т. п.).

Сильное управление безусловно может и должно учитываться в словаре (например, в системе Мельчука—Жолковского, в системах АП). Что касается слабого управления, то целесообразнее включить в грамматику анализа стандартные правила, позволяющие обрабатывать любой слабоуправляемый элемент, а также списки логических связей, которые «вводят» слабоуправляемые элементы.

При таком подходе практическая задача лингвиста сводится к составлению списков логических связей слабого управления, к описанию грамматических средств выражения в языке каждой логической связки и к составлению правил решения омонимии элементов, вводящих слабоуправляемые члены предложения.

Так, если в числе логических связей будут выделены предикаты: причина (и следствие как ее конверсия), условие, цель, сравнение, последовательность во времени (и предшествование как ее конверсия), то далее нужно перечислить для каждого из этих предикатов все возможные способы его выражения в тексте: для причины — предлоги «из-за», «по», «в силу» и т. д., для условия — «под», «при», «на случай» и т. д.

Следующая задача — выработка процедуры различения значений омонимов, так как почти все предлоги, вводящие слабоуправляемые члены, представляют собой грамматические омонимы. Например, предлог «из-за» обозначает причину только в случае, если после него не стоит существительное со значением места.

Слабоуправляемые элементы чаще всего представлены в предложении сочетанием предлога с отглагольным существительным, предлога с существительным, имеющим при себе определение (в обоих этих случаях развертывание в придаточное предложение или в предикатную запись сравнительно легко). Если же слабоуправляемый элемент представлен просто сочетанием предлога с существительным, то требуется более сложное преобразование структуры этого элемента.

Приведем примеры анализа предложений со слабоуправляемыми элементами.

### 1. Предложения с предикатом «причина».

1. «У Аполлона-11 также были осложнения с посадкой главным образом из-за небольшого запаса топлива».<sup>1</sup>

Слабоуправляемый член, вводимый предлогом «из-за», трансформируется в придаточное предложение «запас топлива был небольшим». Из предложения следует, что а) у Аполлона-11 были осложнения с посадкой, б) у Аполлона-11 был небольшой запас топлива, в) второй факт есть причина первого.

2. «Современные подземные системы обречены в силу самого принципа их действия».

Слабоуправляемый член, вводимый предложным оборотом «в силу», трансформируется в предложение «так как сам принцип их действия неудачен». Из предложения следует, что а) современные подземные системы обречены, б) принцип действия современных подземных систем неудачен, в) второй факт — причина первого.

3. «Ассистент профессора Алена Паркса по забывчивости налил семенную жидкость в пробирку с глицерином».

Слабоуправляемый элемент, вводимый предлогом «по», трансформируется в предложение «был человек забывчивый». Из предложения следует, что а) ассистент профессора Алена Паркса налил семенную жидкость в пробирку с глицерином, б) ассистент профессора Алена Паркса был человек забывчивый, в) второй факт — причина первого.<sup>2</sup>

## II. Предложения с логической связкой «условие».

1. «При увеличении расстояния в 10 раз приемлемое время возрастало в 4 раза».

Слабоуправляемый элемент, вводимый предлогом «при», трансформируется в условное предложение «если расстояние увеличивалось в 10 раз, то...».

2. «Исследователи изучают изменения в клетках под воздействием уже известных препаратов».

Слабоуправляемый элемент, вводимый предлогом «под», преобразуется в предложение «если имеет место воздействие на клетки уже известных препаратов».

3. «Процедура подъема космонавтов на авианосец „Хорней“ предусматривает обеспечение их биологической изоля-

<sup>1</sup> Примеры взяты из журнала «За рубежом» за 1965—1971 гг.

<sup>2</sup> О выводах из сложного предложения см.: Н. Н. Ильина. Об одном способе смысловой классификации связей между частями сложного предложения (Сб. Информационные вопросы семиотики, лингвистики и автоматического перевода. М., 1971, стр. 14).

ции на случай занесения ими на Землю лунных микроорганизмов».

Слабоуправляемый элемент, вводимый предложением «на случай», трансформируется в условное предложение «если ими будут занесены на Землю лунные микроорганизмы».

Из простого предложения, содержащего слабоуправляемый член, развертываемый в придаточное предложение условия, по отдельности не следуют части, соответствующие местам предиката условия.

**III. Предложения с логической связкой «следование во времени».**

«После перехода в основной блок космонавты Армстронг и Олдрин с удовольствием отметили, что теперь они могут присесть».

Слабоуправляемый элемент, вводимый предложением «после», развертывается в предложение после того, как они перешли в основной блок. Из предложения следует, что а) космонавты Армстронг и Олдрин перешли в основной блок, б) космонавты Армстронг и Олдрин с удовольствием отметили, что они могут присесть, в) второй факт произошел после первого.

**IV. Предложения с логической связкой «цель».**

«Петров поступил на курсы с целью повышения своих знаний». Это предложение эквивалентно предложению «Петров поступил на курсы, чтобы повысить свои знания». Из такого предложения отдельно следует только часть, соответствующая первому месту предиката цели: «Петров поступил на курсы».

**V. Предложения с «сопутствующими» обстоятельствами.** Существуют некоторые типы слабоуправляемых элементов, которые можно развернуть в так называемые присоединительные предложения, а сами эти слабоуправляемые элементы принято называть сопутствующими обстоятельствами.

Логическая связка, присоединяющая сопутствующее обстоятельство, аналогична конъюнкции, но особого рода: оба факта имеют место, но они обязательно связаны между собой; эта связь может подчеркиваться обстоятельствами (условиями), общими для обоих предикатов, соответствующих местам предиката «конъюнкция».

1. «В США сейчас наблюдается некоторый ажиотаж в связи с использованием на животных препарата под названием „Полл I : C”». Из этого предложения следует, что а) в США сейчас наблюдается некоторый ажиотаж, б) в США

сейчас используют (испытывают) на животных препарат под названием «Поли I : С», в) первый факт связан со вторым.

2. «При лейкемии происходит чрезмерное разрастание белых кровяных телец с преобладанием незрелых форм». Выводы из этого предложения будут следующие: а) при лейкемии происходит чрезмерное разрастание белых кровяных телец, б) (при лейкемии) преобладают незрелые формы (белых кровяных телец).

3. «Вулканизация каучука производится химическим путем с использованием серы». Из этого предложения следует, что: а) вулканизация каучука производится химическим путем, б) (при вулканизации каучука) используется сера.

Из состава слабоуправляемых членов имеет смысл исключить способы выражения в предложении так называемого абсолютного времени<sup>3</sup> и места. Элементы, характеризующие абсолютное время действия и место действия, возможны при любом предикате, но, в отличие от слабоуправляемого члена, непосредственно занимают определенное место глагола-сказуемого, аналогично сильноуправляемому элементу, а не присоединяются к сказуемому с помощью особой логической связки.

И. М. Новицкая

### ЦЕЛЫЕ ПРЕДЛОЖЕНИЯ В ФУНКЦИИ СРЕДСТВ МЕЖФРАЗОВОЙ СВЯЗИ

Проблема связности текста включает в себя в качестве одного из обязательных аспектов подробное исследование средств межфразовой связи.

Однако степень исследованности различных средств межфразовых связей весьма различна — более полно рассмотрены в соответствующих работах слова и словосочетания в

---

<sup>3</sup> Абсолютным временем будем считать указание даты, а также такое описательное время, которое можно заменить датой или указанием на некоторый датированный отрезок времени (например, «когда цветет черемуха»).

функции средств связи<sup>1</sup>, возможность же употребления в этой функции целых предложений пока не получила достаточного освещения.<sup>2</sup> В данной статье рассматриваются некоторые виды целых предложений, способных выступать в функции средств межфразовых связей в текстах научной литературы.

Назовем предложение, выступающее в функции средства межфразовой связи, связочным предложением. Связочное предложение употребляется собственно тогда, когда появляется необходимость перейти от одной части изложения к другой.

Этот переход осуществляется в форме заявления о намерении совершить в связи с вышеизложенным какое-то действие, а именно: рассмотреть вопросы, обсудить или исследовать проблему, сделать выводы, привести примеры или доказательства, описать или исследовать процесс, суммировать наблюдения и т. д. Средствами связи такие предложения являются прежде всего потому, что они связывают, объединяют в относительно законченное целое ряд предложений последующего текста, которые раскрывают содержание названного действия. Кроме того, связочные предложения увязывают это относительно законченное целое с предыдущим текстом.

Связь и с предыдущим, и с последующим текстом имеет свои средства выражения в связочном предложении. Собственно, все связочное предложение и состоит из средств выражения межфразовых связей. Из суммы этих компонентов выкристаллизовывается общая союзная (связочная) функция такого предложения. Связочное предложение не несет никакой содержательной информации, а служит лишь связкой между другими содержательными высказываниями.

---

<sup>1</sup> Укажем лишь некоторые из этих работ: М. М. Макарова. Средства связи самостоятельных предложений в научно-технической речи современного немецкого языка. Автореф. канд. дисс. М., 1960; М. А. Овсянникова. Бессоюзная связь предложений в современном немецком языке. Автореф. канд. дисс. Л., 1960; Т. И. Сильман. Проблемы синтаксической стилистики. Л., 1967; И. А. Фигуровский. Синтаксис целого текста и ученические письменные работы. М., 1961, и др.

<sup>2</sup> Эта проблема затрагивается в весьма немногочисленных работах. См., например: В. А. Ицкович. К вопросу об отношениях между самостоятельными предложениями в тексте. В сб.: Вопросы языкознания, кн. 1. Изд. Львовского ун-та, 1955; В. Н. Скибо. О средствах связи самостоятельных предложений в английской научной прозе. Автореф. канд. дисс. М., 1966, и др.

Весьма характерной для связочного предложения является особая форма повелительного наклонения, в которой стоит сказуемое: «Рассмотрим принципы действия этого автомата...», «Докажем сначала, что...», «Проиллюстрируем на примере...», «Остановимся на случае...». Однако это не является отличительным признаком связочного предложения, так как, во-первых, связочное предложение может иметь и другое оформление, т. е. быть личным (с подлежащими **мы**, **я**) и безличным: «Для доказательства мы рассматриваем...», «Для доказательства необходимо рассмотреть...» Во-вторых, та же форма повелительного наклонения употребляется в предложениях, не являющихся связочными: «Обозначим эту величину буквой **l**», «Внесем эти результаты в формулу, тогда...», «Запомним это число...», «Представим себе большое число систем...».

Определяющим для связочного предложения является наличие в нем специального глагола, относящегося к некоему классу глаголов, которые можно условно назвать «операторными», так как они называют процесс, операцию, характерную для научного изложения, причем операторный глагол стоит в форме настоящего или будущего времени, но не прошедшего.<sup>3</sup> Подобный глагол будем считать основным компонентом связочного предложения.

В зависимости от оттенков значения входящих в них компонентов связочные предложения можно разделить на следующие группы.

1. Предложения перехода от одной части изложения к другой. Основными компонентами таких предложений являются:

а) глаголы, которые можно объединить в семантическую группу глаголов «перехода» (здесь имеется в виду переход от одной части изложения к другой): «перейти», (от чего-то

---

<sup>3</sup> Предложения, в которых операторный глагол стоит в форме прошедшего времени, мы не относим к собственно связочным предложениям, так как в них лишь констатируется факт совершения какого-то процесса в предшествующем тексте без связи с последующим текстом. Ср.: связочные предложения «Теперь изложим основные принципы этой концепции». «Приступаем к изложению основных принципов этой концепции» и невязочное предложение «В предыдущем параграфе мы изложили основные принципы концепции...».

Предложения с операторным глаголом в форме прошедшего времени в дальнейшем рассматриваются в качестве первой части связочного комплекса (см. стр. 130).

к чему-то), «вернуться» (к чему-то), «обратиться» (к чему-то), «приступить» (к чему-то);

б) управляемые ими предложные группы существительных, обозначающие от чего и к чему (или только к чему) совершается переход. Например: «От этих общих замечаний **перейдем** теперь к той форме принципа необратимости, которую он получил в работах Гиббса». <sup>4</sup> «**Вернемся** теперь к соотношению между сохранением энергии и сохранением импульса». «**Обратимся**, однако, к принципу наименьшего принуждения». Связь с предшествующим текстом выражается в этих предложениях либо различными формами повторов (понимая повтор в весьма широком плане), либо так называемыми средствами логической связи, либо тем и другим одновременно (например, в первом предложении «от этих общих замечаний» — повтор, «теперь» — средство установления логического отношения последовательности событий, «однако» — средство установления противительно-уступительного логического отношения). Связь с последующим текстом данных предложений обуславливается тем, что выделенные глаголы перехода «перейдем к...», «вернемся к...» и т. д. означают не что иное, как 'начнем рассмотрение чего-то...' и предполагают описание процесса рассмотрения в последующем тексте.

2. Во вторую группу входят связочные предложения, где называется действие, описание которого содержится в последующем тексте. Основными компонентами таких предложений являются:

а) Глагол, обозначающий процесс, содержание которого должно быть раскрыто в последующем тексте: рассмотреть, анализировать, взять (в смысле рассмотреть), выяснить, вспомнить, доказать, заняться чем-либо (в смысле рассмотреть), изложить, интерпретировать, исследовать, напомнить, объяснить, обосновать, обсудить, описать, определить, охарактеризовать, познакомиться, показать, пояснить, проследить, сравнить, сопоставить и т. д. Например: «**Рассмотрим**, при помощи каких операций получают эти 2<sup>2</sup> высказываний». «**Возьмем** частицу, движущуюся по некоторой поверхности по инерции. Эта частица...». «**Докажем**, что формула (1.87) в самом деле верна». «**Коснемся** двух последних статей Энгель-

<sup>4</sup> Все примеры в этой статье взяты из книг: Б. Г. Кузнецов. Принципы классической физики. М., 1958; А. И. Китов и М. А. Крицкий. Электронные цифровые машины и программирование. М., 1964.

са». «Теперь мы можем интерпретировать полученные результаты». «Для этого необходимо вспомнить некоторые элементарные сведения о характере проводимости полупроводников».

б) Вместо глагола в качестве основного компонента связочного предложения могут выступать сочетания глагола с существительным, типичные для текстов научной литературы, также обозначающие процесс действия, содержание которого раскрывается в последующем изложении: привести примеры (доказательства, распределения, способы), добавить (сказать) несколько слов, коснуться проблемы (вопроса), осветить вопрос (проблему), остановиться на вопросе (случае, понятии, проблеме), сформулировать закон (правило), проиллюстрировать примерами и т. д. Например: «Не вдаваясь в логические тонкости, **приведем** простой и наглядный способ геометрического пояснения алгебро-логических операций». «**Остановимся** сперва на случае, когда запятая фиксирована после одного старшего разряда сумматора. Пусть...». «Не проводя этих рассуждений, **проиллюстрируем** наше утверждение примерами». «Прежде всего необходимо **остановиться на вопросе** о знаках логических связей».

Сюда же отнесем и сочетания глаголов «перехода» с существительными, означающими процесс действия: перейти к рассмотрению (доказательству, изложению, описанию, сравнению и т. д.), вернуться к рассмотрению (изложению, обсуждению, описанию и т. д.), обратиться к рассмотрению (выяснению, сравнению, обоснованию и т. д.), приступить к выяснению (доказательству, изложению, описанию и т. д.).

Например: «После предварительных замечаний **перейдем** непосредственно к описанию основных типов запоминающих устройств».

«Теперь **вернемся к рассмотрению** блок-схемы устройства сложения и вычитания».

Связь с предшествующим текстом здесь осуществляется посредством повторов и средств логической связи.

Необходимо заметить также, что связочные предложения обычно выступают в виде простых предложений, что, по-видимому, обусловлено определенностью и ограниченностью их функции. В тех случаях, когда в состав связочного предложения включаются придаточное предложение или оборот, они представляют собой, как правило, более или менее подробное обоснование для совершения называемого действия



в связи с предшествующим текстом. (Особенно это касается начальных придаточных предложений и оборотов, находящихся в так называемой пограничной области, характерной для местонахождения средств межфразовой связи.) Например: **«Для того чтобы понять принцип действия полупроводниковых диодов, проанализируем некоторые данные о характере проводимости полупроводников».** **«Подводя итоги настоящего параграфа, рассмотрим общую классификацию современных быстродействующих вычислительных машин».**

В начальных придаточных и оборотах, входящих в связочное предложение, может также констатироваться факт окончания процесса описания или рассмотрения, изложенного в предшествующем тексте, в то время как в главном объявляется о переходе к следующей операции.

И в этом случае в придаточных и оборотах связочного предложения осуществляется связь с предшествующим текстом. Например: **«После того, как мы охарактеризовали принцип номинализации, остановимся подробнее на некоторых наиболее употребительных видах композиции алгоритмов».** **«Рассмотрев обобщенные нормальные алгоритмы, перейдем к характеристике собственно нормальных алгоритмов».**

3) Если две предыдущие группы предложений являлись средствами связи из-за общего содержания, вытекающего из суммы составляющих их компонентов, то третья группа предложений служит средством межфразовой связи из-за своей внешней формы. Здесь имеются в виду вопросительные предложения.<sup>5</sup>

Вопросительные предложения выполняют роль средств межфразовой связи в силу того, что они связывают предшествующий им текст с одним или несколькими последующими предложениями, в которых содержится ответ на заключенный в них вопрос. Например: **«Имеют ли эти линии физическую реальность? Фарадей решительно утверждал реальность физического существования силовых линий. В статье „О фи-**

---

<sup>5</sup> Некоторые виды вопросительных предложений в качестве средств межфразовой связи выделяются, в частности, П. А. Рестаном в статье «Вопросительное предложение, его формы и функции» (Scando-Slavica, т. XII, 1966, стр. 132—148). Также и В. В. Суренский в качестве одного из типов лексико-синтаксических отношений между смежными предложениями называет вопросно-ответные отношения (В. В. Суренский. Логико-синтаксические отношения между смежными предложениями связного текста и их выражение в устной речи. «Русский язык в школе», 1949, № 2).

зическом характере линий магнитной силы"... он пишет: „Что касается важного вопроса, подлежащего рассмотрению, то он заключается только в том, имеют ли линии магнитной силы физическое существование или нет". Ответ Фарадея положительный».

Связь с предшествующим текстом выражается в вопросительном предложении обычно разнообразными повторами. Например: «Сохраняется ли при этом ковариантность уравнений?». «Совпадает ли такое среднее распределение с наиболее вероятным?». «Но не противоречат ли такие обратимые процессы второму началу термодинамики?».

В научно-технических текстах несколько вопросительных предложений могут следовать друг за другом, при этом в вопросительной форме осуществляется постановка ряда проблем, решение которых приводится в последующем тексте. Например: От чего же зависит экономичность такого идеального двигателя? Ограничена ли она? Остается ли она неизменной при переходе от водяного пара к другим рабочим телам?

Весь этот комплекс вопросительных предложений как единое целое можно считать средством связи между предшествующей и последующей частями текста.

Однако вопросительные предложения, выступающие в качестве средств межфразовой связи, являются не единственным случаем использования уже не одного, а двух и более предложений для связи между частями текста.

Назовем несколько предложений, составляющих единое средство межфразовой связи, связочным комплексом. В связочном комплексе, как и в связочном предложении, должна быть выражена связь как с предшествующим, так и с последующим текстом, при этом основной функцией такого комплекса должна быть функция межфразовой связи.

Связочные комплексы могут встречаться и внутри параграфа, т. е. они могут связывать и ограниченные по объему части текста. Но чаще всего они связывают части текста больших объемов — параграфы и главы.

Наиболее прозрачную структуру связочного комплекса представляет собой комплекс из двух предложений, первое из которых содержит «операторный» глагол в форме прошедшего времени, осуществляя тем самым связь с предшествующим текстом, а второе является связочным предложением, осуществляющим непосредственную связь с последующим.

текстом. Находясь в начале параграфа или главы, такой комплекс служит для связывания данного параграфа (главы) с предшествующими параграфами (главами), т. е. отвечает вышеперечисленным требованиям к связочному комплексу. Например: «В предыдущем параграфе мы рассмотрели примеры блок-схем сумматоров, предназначенных для сложения и вычитания чисел с фиксированным положением запятой. Рассмотрим теперь порядок работы устройства для сложения и вычитания чисел, заданных в нормальной форме с плавающей запятой» (Китов, 166). «Выше мы в общих чертах описали некоторые известные способы программного контроля правильности работы машины и вычислений. Перейдем теперь к детальному рассмотрению каждого из этих способов» (там же, 168).

Обе вышеназванные части являются основополагающими при образовании связочного комплекса. Они обязательные, но не единственные составляющие такого комплекса. Кроме них, в комплекс могут входить предложения с самыми различными дополнительными сведениями, но при этом ни общий характер, ни основная функция связочного комплекса не меняются.

Для иллюстрации этого утверждения приведем следующий пример краткого вступления к главе, представляющего собой связочный комплекс с промежуточными предложениями: «Во введении (§ 2) была приведена упрощенная блок-схема электронной цифровой программно-управляемой машины и рассмотрен принцип программного управления. Мы уже знаем, что любая электронная программно-управляемая машина состоит из трех основных частей: арифметического устройства, запоминающего устройства и устройства управления. Кроме того, в состав машины входят специальные устройства для ввода информации (чисел и команд) в машину и для вывода информации из машины. В настоящей главе мы познакомимся с техническими принципами построения и работы устройств машины».

Таким образом, система средств межфразовой связи может быть значительно расширена за счет включения в нее некоторых видов целых предложений, а также комплексов предложений, выступающих в функции средств связи.

## СПОСОБЫ РАЗВЕРТЫВАНИЯ КОНСТРУКЦИЙ С УПОДОБЛЯЮЩИМИ МЕСТОИМЕНИЯМИ ВО ВНУТРЕННЕЙ СТРУКТУРЕ ТЕКСТА

В информационной системе «запрос—ответ»<sup>1</sup> выделяется особый способ представления естественного текста — внутренняя структура, которую предполагается строить так, чтобы выполнялись по крайней мере два требования: а) максимальная близость внутренней структуры к естественному языку, в) способность ее обеспечивать выдачу системой ответа на любой вопрос по тексту.

Элементами внутренней структуры<sup>2</sup> считаются простые и сложные объекты, а также функторы разного типа, среди которых прежде всего выделяются предикаты, служащие для связи объектов, а также коннекторы, служащие для связи двух ситуаций (ситуация — предикат с местами, заполненными объектами). Таким образом, разница между функторами «предикат» и «коннектор» заключается в различной структуре (заполнении) их мест.

Подобное различие двух типов функторов оказалось, например, удобным для описания связи между элементами текста, которую мы назвали «подобие». В зависимости от структуры мест функтора подобия выделяются два типа этого функтора.

1. Трехместный предикат подобия, который условно можно изобразить: «А подобно В в отношении С», где А и В — простые или сложные объекты, а С — некоторое их свойство (аспект, признак и т. д.), относительно которого утверждается их подобие. При этом следует отметить, что на содержательном уровне сочетания «свойство (аспект, признак и т. п.) А», «свойство (аспект) В» образуют в свою очередь сложные объекты, а не ситуации. Эквивалентным преобразованием ситуации, выраженной формулой «А подобно

<sup>1</sup> См., например: Г. М. Ильин, Б. М. Лейкина, Т. Н. Никитина, М. И. Откупщикова. Модель семантики текста и система «запрос—ответ». НТИ, 1969, № 1.

<sup>2</sup> Г. М. Ильин, Б. М. Лейкина, Т. Н. Никитина, М. И. Откупщикова. Лингвистический подход к задаче построения информационных систем. В сб.: Информационные вопросы лингвистики, семиотики и автоматического перевода. Приложение к НТИ, 1971, № 2.

В в отношении С» во внутренней структуре будет преобразование по симметричности: «В подобно А в отношении С».

2. Коннектор подобия «А подобно В», где А и В — ситуации, т. е. представляют собой предикаты с заполненными местами. Преобразование по симметричности—«В подобно А». В этом случае основание подобия обычно выражается предикатами, образующими ситуации А и В. Набор всех выводов (а следовательно, значение коннектора подобия) из «А подобно В» будет следующим: 1.1. А; 2. В; 3. «А подобно В». 4. «В подобно А», в то время как из «А подобно В в отношении С» набор выводов будет другим: II.1. «А подобно В в отношении С». 2. «В подобно А в отношении С». 3. «Свойство» С(А) подобно «свойству» С(В). 4. И наоборот. 5. «А обладает свойством С». 6. «В обладает свойством С»<sup>3</sup>.

Нетрудно видеть, что на уровне смысла II сводимо к I (это очевидно при переобозначении 3, 4, 5, 6 выводов из II, однако при условии неразличения структуры мест). Единобразное представление функтора подобия во внутренней структуре может привести к усложнению правил преобразования внешней структуры текста во внутреннюю, которые и без того достаточно сложные. Поэтому в данной статье в виде рабочей гипотезы во внутренней структуре допускается существование структурной синонимии особого типа.

Приведем примеры из текста, в которых так или иначе выражен функтор подобия:

1. «Лигнин, подобно саже, может служить усилителем каучука»<sup>4</sup>.

2. «Колебания уровня сигналов в обоих пунктах расположения антенны аналогичны по структуре».

3. «Ультракороткие волны обладают свойствами, очень схожими со свойствами световых волн».

4. «Ребенок, как и его мать, был голубоглазый и светловолосый».

5. «Но спелость у почвы, словно спелость у груши, — это состояние мимолетное».

---

<sup>3</sup> Здесь не учитываются выводы, связанные с кванторной информацией текста.

<sup>4</sup> Большинство приведенных в статье примеров взято из журналов: «Вопросы языкознания» (за 1969 г.), «Наука и жизнь» (за 1970 г.), «Элементарный учебник физики» под ред. академика Г. С. Ландсберга. М., 1967.

6. «...Гидролитическое равновесие количественно характеризуется степенью гидролиза подобно тому, как равновесие диссоциации слабого электролита характеризуется степенью диссоциации».

В предложениях 1, 4, 5, 6 различными языковыми средствами выражен коннектор подобия, в 2, 3 — предикат. Для иллюстрации перечислим набор выводов из 1-го и 5-го предложений.

1. а) лигнин может служить усилителем каучука (А),  
б) сажа может служить усилителем каучука (В),  
с) лигнин может служить усилителем каучука — коннектор подобия — сажа может служить усилителем каучука (А подобно В),  
д) и наоборот (В подобно А).
2. а) колебания уровня сигналов в одном пункте... — предикат подобия — колебания уровня сигналов в другом пункте... в отношении структуры (А подобно В в отношении С),  
б) и наоборот (В подобно А в отношении С),  
с) структура колебаний уровня сигналов в одном пункте — предикат подобия — структура колебаний уровня сигналов в другом... (свойство С (А) подобно свойству С (В)),  
д) и наоборот (свойство С (В) подобно свойству С(А)),  
е) колебания уровня сигналов в одном пункте... обладают структурой (А обладает С),  
ф) колебания уровня сигналов в другом пункте обладают структурой (В обладает С).

В естественном языке функтор подобия выражается различными средствами — словами и сочетаниями типа: «подобный (же)», «аналогичный (же)», «подобно», «как и», «как», «словно», «будто», «схожий», «соответствующий», «уподобить(ся)», «похожий», «походить», «подобно тому как», «наподобие» и т. д.

Остановимся на средстве выражения этого функтора при помощи особого разряда местоимений русского языка, которые мы назвали уподобляющими. К разряду уподобляющих местоимений относим следующие: «тот (та, то, те) же», «то же» (местоименное сущ.), «этот (эта, это, эти) же», «это же» (сущ.), «такой (-ая, -ое, -ие) же», «таков (-а, -о, -ы) же», «тот (та, то, те) же самый (-ая, -ое, -ые)», «то же самое»

(сущ.), «один и тот же» («одна и та же» и т. д.), «так же», «столько (столь) же», «тогда же», «туда же», «оттуда же», «там же». Как и всякие местоимения (не считая кванторных), уподобляющие местоимения служат заместителями предшествующих отрезков текста (или предваряют последующие его отрезки), однако это замещение «отягчено» значением уподобления содержания, выраженного замещаемым отрезком текста какому-либо другому содержанию, которое может в частном случае быть тем же самым. Сказанное в полном объеме не может быть приложимо к любому из употреблений (или, иначе, значений) уподобляющих местоимений в естественном тексте. Поэтому значение их удобнее описывать при помощи частично иерархического набора бинарных признаков (оппозиций):  $a$  — замещение,  $a_1$  — не замещение;  $b$  — уподобление,  $b_1$  — «не уподобление»<sup>5</sup>;  $c$  — уподобление содержания, выраженного отрезком текста, самому себе;  $c_1$  — уподобление не самому себе;  $d$  — предварение,  $d_1$  — не предварение. Примеры:

1) «Один ряд звуков **столь же** немотивирован, как и другой». Функцию (значение) местоимения (здесь: местоименного наречия) **столь же** можно охарактеризовать как  $a_1, b, c_1, d_1$ , т. е. здесь местоимение **столь же** выражает функтор подобия (коннектор).

2) «Он шагов в двадцать пять длиною и **столько же** в ширину». Характеристикой употребления **столько же** здесь будет:  $a, b, c_1, d_1$  (**столько же** — здесь «чистый» заместитель).

3) «Открытия случайные, подобно мутациям генотипа. И точно **так же**, как мутации, они могут привести к радикальным изменениям». Характеристика **так же** —  $a, b, c_1, d_1$ .

4) «...при помощи **одной и той же** радиолокационной станции следить за целью и наводить на эту цель снаряд не всегда возможно». Характеристика **одной и той же** —  $a_1, b, c, d_1$  (частный случай предиката подобия, когда  $A$  и  $B$  совпадают)<sup>6</sup>.

<sup>5</sup> Признак «не уподобление» введен условно для случаев, не подходящих под описанные выше значения функтора «подобия». Этот признак (см. пример 2) приписывается в случаях «чистого» замещения, которое обычно выражает частный случай подобия — полное отождествление, в этих случаях значение предиката подобия будет выражено утверждениями  $A \approx B, B \approx A$  (так называемый предикат тождества).

<sup>6</sup> Предикат подобия в случае, если на его местах стоят объекты  $A$  и  $B$ , где  $A=B$ , очевидно, становится двуместным, ибо  $C$  тогда равно 0 (см. прим. 5).

5) «Электрический сигнал от датчиков записывается на **ту же** дорожку магнитофонной ленты, что и звук». Характеристика **ту же** —  $a_1, b, c_1, d$  (предикат подобия с  $A=B$ ).

6) «Заимствованный книжный слой лексики имеется во многих языках; в русском языке, кроме отдельных выражений **того же типа**, встречаются книжные формы...» Характеристика **того же (типа)** —  $a, b, c_1, d_1$ .

Те употребления уподобляющих местоимений, которые можно охарактеризовать сочетанием признаков  $a$  и  $b_1$  (замещение и «неуподобление»), будем считать синонимами обычных заместителей, выражающими частный случай функтора подобия. Остальные значения уподобляющих местоимений служат одним из самых распространенных средств выражения функтора подобия в тексте.

Далее нас будут интересовать различные способы развертывания во внутренней структуре конструкций с уподобляющими местоимениями. Необходимость развертывания предложений во внутренней структуре связана с указанной выше задачей системы — давать ответы на любые запросы в пределах заданных текстов. Возможность же такого развертывания связана с особым свойством организации структуры естественного текста — его экономией. В частности, высокой «сокращающей» силой обладают местоимения, и это в полной мере относится к уподобляющим местоимениям, которые в большинстве случаев употребления свидетельствуют о сокращениях текста, что можно видеть, если, например, обратиться к приведенным ранее примерам.

В зависимости от вышеуказанных признаков классификации функций уподобляющих местоимений можно выделить несколько типов развертывания высказываний с уподобляющими местоимениями во внутренней структуре, а также попытаться предварительно определить языковые различия, выражающие эти типы. И то и другое является задачей «глубинного» синтаксиса.

**I тип.** Развертывание при замещении (восстановление замещения). Примеры: 1. «...книга снабжена Приложением 1, где собраны все математические утверждения..., **там же** даны ссылки на источники...». Здесь уподобляющее местоимение является чистым заместителем. По особым правилам **там же** при «глубинном» синтаксическом анализе должно быть замещено на «в Приложении 1». 2. «Разделив работу на время в секундах, найдем мощность (далее идет формула).



Такой же результат можно получить иначе». Сочетание такой же результат замещает формулу предыдущего предложения. После того как восстановление произведено, выводы из подобных высказываний будут иметь вид А, В, где А — часть, содержащая антецедент, В содержит заместитель. Применяя различные преобразования, можем получить также вывод о тождестве двух объектов (предикат тождества).

3. «Определенный артикль женского рода la теряет (а) перед гласным или дифтонгом. Прилагательное мужского рода «се» приобретает дополнительное /t/ перед теми же звуками». Здесь уподобляющее местоимение теми же замещает гласным и дифтонгом и т. д.

Иногда развертывание при замещении связано с дополнительными преобразованиями: «История вьетнамского языка нам совершенно неизвестна — остается только догадываться, как звучали вьетнамские слова 400, 500, 1000 лет назад. То же относится и к языкам Западной Африки, индейским языкам Америки». Чтобы произвести правильно отождествление в случаях с заместителями вида «то же», «это же» (относится, наблюдается, делает и т. п.), необходимо преобразовать первое предложение так, чтобы в нем на соответствующих местах стояли объекты или предикаты из второй части высказывания, расположенной после уподобляющего местоимения (в нашем примере вместо «вьетнамский» — «языки Западной Африки и индейские языки Америки»). В этом случае уподобляющее местоимение переводится во внутренней структуре в коннектор подобия.

Одним из признаков чистого замещения является позиция уподобляющего местоимения по отношению к замещаемому антецеденту, а также несопредикатность с антецедентом<sup>7</sup>, что чаще всего выражается в том, что местоимение находится или в другом предложении, или в какой-либо самостоятельной части целого предложения, расположенной после той части, которая имеет замещаемый антецедент.

Следует отметить, однако, что формальное решение восстановления заместителей уподобляющих местоимений не

---

<sup>7</sup> Е. В. Падучева подобное явление называет сопредикативностью.— См.: Е. В. Падучева. Выражение тождества упоминаемых объектов как одна из проблем синтеза языкового текста. Труды III Всесоюзной конференции по информационно-поисковым системам и автоматической обработке научно-технической информации, т. II. М., 1967.

всегда возможно. Здесь встает та же сложная проблема, что и в случае разворачивания заместителей другого рода, когда структурно-синтаксических признаков недостаточно для установления правильной связи между замещаемым отрезком текста и заместителем, и поэтому необходимо обращение к смыслу высказывания.

**II тип.** Разворачивание при уподоблении. В этом случае одно из мест функтора в первой ситуации или сам функтор уподобляются соответствующему структурно месту такого же функтора или подобному функтору другой ситуации. Примеры: 1. «Один контрпример отвергает догадку так же эффективно, как и десять». 2. «Признаки высоты тона широко используются в той же функции, что и жесты...» 3. «Один ряд звуков столь же немотивирован, как и другой». 4. «И лишь глаза, такие же светлые и прозрачные, как у Тимофея, дерзко смотрели и независимо». 5. «Они поедут туда же, куда и мы».

В этих случаях часть предложения, соответствующая второй ситуации, обычно сокращена. Явно выражено только то место ситуации, которое отличается по содержанию и выражению от соответствующего ему структурно места первой ситуации. Поэтому при разворачивании подобных высказываний во внутренней структуре вторая часть будет соответствовать первой ситуации («накладываться на нее») с заменой содержания того места, которое явно выражено во второй части (на этом этапе эллипсисы другого типа считаются уже восстановленными).

В примере вторая ситуация «развернется» так: «десять контрпримеров отвергают догадку эффективно». В набор выводов у предложения 1, следовательно, войдут первая ситуация «один контрпример отвергает догадку эффективно», вторая, а также третья, в которой коннектором подобия соединены предикаты «эффективно».

Из предложения 2 следует, что вторая ситуация — «жесты используются в некоторой функции». Предикат подобия соединяет места, заполненные словами «функции».

Из предложения 3 следует: вторая ситуация — «другой ряд звуков немотивирован». Коннектор подобия соединяет предикаты первой и второй ситуаций — «немотивирован».

Из предложения 4 получаем вторую ситуацию — «У Тимофея светлые и прозрачные глаза». Отношения подобия связывают функтор «светлые и прозрачные». Вообще, если

уподобляются функторы, то в результате мы имеем коннектор подобия и можем получить соответствующие выводы (см. первую часть статьи); если уподобляются места, то мы имеем соответственно предикат подобия.

Характерной особенностью уподобляющих местоимений в естественном тексте является наличие их непосредственной синтаксической связи со словом, заполняющим то место второй ситуации, которое уподобляется соответствующему месту первой ситуации. В связи с этим на этапе синтаксического анализа легко формально связать два места отношением подобия.

Если уподобляющее местоимение заполняет место предиката, не являясь синтаксическим его определителем, то для получения выводов, соответствующих по содержанию исходному тексту, уподобляемые места первой ситуации следует заполнить неопределенным содержанием (см. пример 5). Первая ситуация: «они поедут(а) куда-то». Вторая ситуация: «мы поедем(в) куда-то»: (а) и (в) — здесь «одно и то же место».

Формальным средством обнаружения необходимости применить второй тип развертывания может служить препозиция уподобляющего местоимения и с точки зрения линейного расположения его относительно второй ситуации (в этом, в частности, выражается предваряющая функция местоимения). Вторым формальным признаком может служить также наличие в предложении конструкций с относительными словами, которые зачастую коррелируют с уподобляющими местоимениями: там же, где (и); тот же, что (и); так же, как (и); туда же, куда (и); такой же, как (и); столько же, сколько (и); тогда, когда (и); тот же самый, что (и) и т. п.

Употребление местоимений изучаемого типа в функции уподобления с предварением является очень характерным для этих местоимений. Уподобляющие местоимения, помимо выражения функтора подобия, несут на себе функцию замещения и служат поэтому одним из способов сокращения текста. В связи с этим зачастую требуются более сложные преобразования для получения выводов, соответствующих значениям функтора подобия.

**III тип.** Уподобление объекта самому себе. Сфера действия уподобления ограничена обычно рамками одного предложения. Эта функция выражается, как правило, местоиме-

нием «один и тот же», реже «тот же самый» (когда оно не несет предваряющей функции) и другими. Смысл, выражаемый этими местоимениями, — тот же самый, что и в типе II, но употребляются они обычно для отождествления не двух или более ситуаций (число их обычно явно задано в тексте), а для многих ситуаций, количество которых точно выявить невозможно, а следовательно, трансформация развертывания в этом случае становится невозможной. Примеры: «...многократно употребляется **один и тот же** корень, например, в словах „жизнь“, „живое“, „жить“».

«На этом основании вполне возможно обнаружить **один и тот же** состав „звуков“ в языках с абсолютно различной фонематической структурой».

«Текстовые единицы находятся в свободном чередовании, если они встречаются в **одних** и тех же окружениях, не различая значений».

Это интересный случай языковой компрессии, который мы назовем условно абсолютным сокращением, при котором развертывание нецелесообразно, и поэтому для внутренней структуры следует искать какой-то адекватный естественному тексту способ представления, например ввести одноместный предикат «одинаковый».

Следующим этапом исследования можно считать выявление омонимичных средств внутри выделенных трех типов, а также описание всех синонимичных способов выражения функтора подобия. Однако это не входит в задачу данной статьи.

Л. А. Старченко

#### НЕКОТОРЫЕ СПОСОБЫ ВЫРАЖЕНИЯ В ТЕКСТЕ «ИСТОЧНИКА» ИНФОРМАЦИИ

Начальным этапом на пути построения языка «смысла» является комплексное изучение внешней языковой структуры с точки зрения содержания сообщения.

Следует различать «содержательную» информацию (на основе которой будут делаться выводы) и «оценочную» (так

или иначе характеризующую содержание). Одной из возможных характеристик «содержательной» информации является указание на ее «источник».<sup>1</sup> Настоящая работа ставит своей целью описать в общих чертах те элементы внешней структуры текста, которые в той или иной форме указывают на такой источник. Исследование проводилось на материале научно-технических текстов (научные статьи, доклады на конференциях).

Условимся понимать под сообщением любое отдельное предложение. В зависимости от того, кто является автором сообщения, любое предложение можно предварить словами: «Я сообщаю, что...» или «Ученые сообщают, что...» Но обычно эксплицитно выраженное указание на того, кто является ответственным за вводимую информацию, встречается в текстах редко (особенно в тех случаях, когда таким «ответственным» является сам автор) и выражается ограниченным числом чисто «декларативных»<sup>2</sup> глаголов («говорить», «сообщать», «передать» и т. п.). Большинство же глаголов, вводящих информацию об источнике информации, всегда содержит какой-то дополнительный компонент (чаще всего модальный), который мы здесь рассматривать не будем (утверждать = говорить, что веришь; предполагать = говорить, что считаешь возможным и т. д.).

«Источник» информации вводится не только глаголами. Случаи эксплицитного выражения авторской ссылки на источник сообщаемой информации можно расклассифицировать следующим образом.

1. Ссылка на собственное авторское знание: «Я говорю, что...», «Мы сообщаем, что...».

Примеры:

1) В заключение я скажу, что неудовлетворительность и примитивность анализа результатов наблюдений и исследований являются одной из главных и частых причин диагностических ошибок.

2) Мы сообщаем, что первые эксперименты по лечению злокачественных опухолей с помощью лазера дали обнадеживающие результаты.

---

<sup>1</sup> Т. Н. Никитина, Л. А. Старченко. Некоторые вопросы соотношения внешней и внутренней структуры текста. В сб.: Информационные вопросы семиотики, лингвистики и автоматического перевода, вып. 2. М., 1971, стр. 23.

<sup>2</sup> Термин «декларативный» заимствован у Ш. Балли.— См.: Ш. Балли. Общая лингвистика и вопросы французского языка. М., 1955, стр. 60.

## II. Ссылка на коллективное знание.

### A. Общая (недифференцированная) ссылка.

а) Ссылка на сам факт коллективного знания: Мы<sup>3</sup> знаем, что...; Нам известно, что...; Известно, что... (я и некоторые или все знаем, что...); Общеизвестно, что... (я и все знаем, что...); Мало кому известно, что... (я и некоторая часть всех знаем, что...). Примеры<sup>4</sup>:

«Как мы знаем, положение окружности на плоскости полностью определяется любой тройкой ее точек, не лежащих на одной прямой». «Нам известно, что такая система крайне неустойчива». «Известно, что прозрачность интерферометра меняется для каждой длины волны при прохождении света от одной отражающей поверхности к другой...». «Общеизвестно, что в современном английском языке наблюдается смещение понятия переходности-непереходности». «Всем известно, что если посеять семена картофеля или яблони, то признаки хорошего сорта в потомстве не сохраняются».

б) Ссылка на некоторый вид информационного материала, обусловившего факт коллективного знания. Примеры:

1. «Судя по данным зарубежной печати, подобная система устанавливается на истребителе, вооруженном управляемыми реактивными снарядами».

2. «Как стало известно из газет, самоходный аппарат „Луноход-1” хорошо перенес условия лунной ночи».

### B. Частная (дифференцированная ссылка).

а) Указание на область науки. Примеры:

«Согласно квантовой механике, момент количества движения электрона тоже квантуется...»

«Из физики известно, что двигающиеся электроны... испытывают воздействие сил, направление которых определяется правилом левой руки».

б) Ссылка на специалиста в некоторой области науки. Примеры:

«С точки зрения этнографа, не существует общества, не имеющего своей культуры».

«Но кроме нашей Галактики, как утверждают астрономы, существуют сотни миллиардов других Галактик».

<sup>3</sup> В данном случае «мы» — это способ выражения неопределенно-личного субъекта в отличие от случая 1, где «мы» — группа авторов.

<sup>4</sup> Примеры, приведенные в статье, взяты из журнала «Наука и жизнь» за 1966—1970 гг.

«Многие лингвисты высказывали мнение, что фонетические различия должны определяться через дифференциальное значение».

III. Ссылка на конкретный источник знания.

А. С определенным указанием на:

а) конкретное лицо:

«Резерфорд предположил, что атом состоит из положительно заряженного ядра».

«По мнению Кеплера, свет распространяется мгновенно».

«По С. К. Всехсвятскому, из продуктов таких извержений образуются кометы».

б) конкретный вид информационного материала:

«Из опубликованных газетой „Нью-Йорк таймс“ документов Пентагона явствует, что месяцы, прошедшие с начала 1964 года и до провокации в Тонкинском заливе... были решающим периодом, когда составлялись планы последующих этапов расширения американской войны во Вьетнаме».

в) конкретный научный закон, теорию, формулу:

«Из формулы Больцмана видно, что число частиц, находящихся в возбужденном состоянии, зависит прежде всего от температуры».

«По теории Бора, электрон вращается вокруг ядра атома только по вполне определенным орбитам».

Б. С неопределенным указанием на:

а) конкретное лицо (которое известно автору, но которое автор не хочет назвать):

«Один немецкий поэт сказал, что с тех пор как существуют женщины, существует искусство украшения».

«Один французский астроном в шутку заметил, что этот эксперимент можно сравнить с метким выстрелом в глаз мухи с расстояния 5 км, когда муха летит со скоростью 100 км/ч».

б) конкретное сочинение:

«Как видно из статей В. И. Ленина о Л. Н. Толстом, стремление художника ...правдиво, реалистически отобразить действительность может даже в какой-то степени выводить его за пределы классовой ограниченности его мировоззрения».

в) конкретный научный закон:

«Один из постулатов... гласит, что скорость света в вакууме есть величина постоянная...».

Следует отметить также, что в одном и том же сообщении может содержаться не одно, а несколько указаний на «источник» информации. Это случаи так называемой «ссылки на ссылку» и цепочка таких ссылок в принципе неограничена. Примеры:

«А. В. Исаченко, полемизируя с А. А. Реформатским, утверждает, что с точки зрения логики антонимами следует считать только противоречительные понятия, поддающиеся диалектической интерпретации».

«Космонавты корабля „Джеминай-7“, как сообщает иностранная печать, установили связь с Землей по лазерному лучу и передали, что вся бортовая аппаратура работает нормально».

Важно отметить также роль кванторных слов при классификации ссылок. Так, например, «один», «один из», «некий» являются прямыми указателями на случай ИБ. На этот же случай указывают количественные числительные в отличие от слов «многие», «некоторые», указывающие на случай ПБ.

Наблюдения над текстом показали, что основными средствами эксплицитного выражения ссылки на источник сообщения являются следующие:

1. Глаголы речи в различных выражениях: мы говорим, нам сказали, нас известили, как сказано в..., нами утверждается, говоря словами автора и т. п.

2. Предикативное наречие «известно».

3. Конструкция «предлог **по** или **согласно** + одно из нижеследующих существительных: мнение, сведение, теория, концепция или существительных, образованных от глаголов речи (например, сообщение, предположение и т. п.)».

4. Выражения «с точки зрения, судя по + существительное».

5. Выражение с предлогом «из»: из ...видно (следует, вытекает, явствует и т. п.).

Интересно отметить, что выражения «судя по...» и «из... видно (следует, вытекает, явствует)» имеют оттенок интерпретации передаваемого сообщения в отличие от других выражений, носящих характер пересказа или цитирования.

Дальнейшая разработка данной проблемы, как это видно из примеров, тесно связана с решением вопроса о представлении кванторного и модального компонентов высказывания.



## ПОСТРОЕНИЕ АЛГОРИТМА МОРФОЛОГИЧЕСКОГО АНАЛИЗА ДЛЯ РУССКОГО ЯЗЫКА НА ОСНОВЕ КОНЕЧНОАВТОМАТНОЙ МОДЕЛИ<sup>1</sup>

Цель словарно-морфологической обработки текста — заменить каждую словоформу входного предложения на морфологическую информацию о ней. Словоформой считается набор букв от пробела (или дефиса) до пробела (или дефиса).

Исходными данными для работы алгоритма являются: словоформа входного предложения и словарь-таблица<sup>2</sup>, описывающая морфологию русского языка в рамках конечноавтоматной модели.

Алгоритм представляет собой перебор различных возможных путей порождения словоформы в указанной модели. Для начальной части словоформы (т. е. основы или основы с некоторым количеством присоединенных аффиксов) предполагается, что некий путь порождения цепочки уже пройден (имеется в виду последовательное присоединение аффиксов с определенными значениями), и зафиксировано определенное состояние автомата (т. е. информация при начальной части словоформы). Из остатка словоформы выделяются одна или несколько начальных букв (аффикс), определенная часть информации которой (которых) полностью совпадает с определенной частью информации, задающей состояние автомата. Так получается предположение о порождении более длинной цепочки и фиксируется новое состояние автомата. Если таким образом доходим до конца словоформы, то это возможный вариант ее порождения;

<sup>1</sup> Данная статья является обобщением работы по созданию морфологического анализа — начальной ступени системы автоматического перевода текстов научно-технических публикаций на язык экспериментальной информационно-поисковой системы, разрабатываемой в группе математической лингвистики ВЦ ЛГУ. В подготовке лингвистического материала и в проведении машинного эксперимента большую помощь оказали Н. Д. Кремнева и Н. А. Крупко.

<sup>2</sup> Для машинной реализации оказалось целесообразным разделить словарь-таблицу на словарь основ («Словарь основ») и таблицу аффиксов («Словарь аффиксов»).

если нет, то это значит, что на одном из предыдущих шагов было сделано неправильное предположение. В первом случае делаем шаг назад в поисках другого варианта: другого набора информации при той же последовательности букв (при последнем аффиксе), а если его нет — в поисках другой последовательности букв (аффикса другой длины). Во втором случае делаем шаг назад в поисках аффикса другой длины. Если на этом шаге перебраны все варианты, с которыми мы доходим или не доходим до конца словоформы, то делаем еще шаг назад и т. д. Так доходим до самого первого предположения, относительно основы слова, и здесь также ищем другой вариант: либо другой набор информации при той же начальной части словоформы, либо другой вариант самой начальной части словоформы (начальную цепочку другой длины). Все промежуточные результаты (предположения на разных шагах) хранятся в специально организованном массиве.

#### Описание конкретного автомата для русской морфологии

Прежде всего отметим, что русская морфология в нашем словаре-таблице учтена не полностью. Под неполнотой понимается совокупность разного рода ограничений.

Ограничение I. Объектом описания являлась письменная речь, в которой е и ё не различаются и где не маркируется ударение. По этой причине словоформы типа: «все» и «всё», «синеет» и «синее», «разрѣзать» и «разрѣзѣть», «получите» и «получѣте», «стоѣщий» и «стоѣщий» рассматриваются как омонимы.

Ограничение II. Пока исключены из рассмотрения: 1) обработка равнозначных основ с целью их отождествления (чередования гласных и согласных, беглые гласные); 2) членение сложного слова на две или более самостоятельные основы; 3) стандартная обработка словообразовательных аффиксов<sup>3</sup>; 4) обработка неотделяемой частицы *-ся* (*-сь*), например «кажется»; 5) значения родительного и предложного падежей у аффиксов *-у*, *-ю*, например «нет сахару», «нет чаю», «на берегу», «на краю».

---

<sup>3</sup> Все разночтения основ (кроме разночтений типа: брат—братья, время—времени), сложные слова и все комбинации основ со словообразовательными аффиксами имеют в настоящее время самостоятельные входы в «Словарь основ».

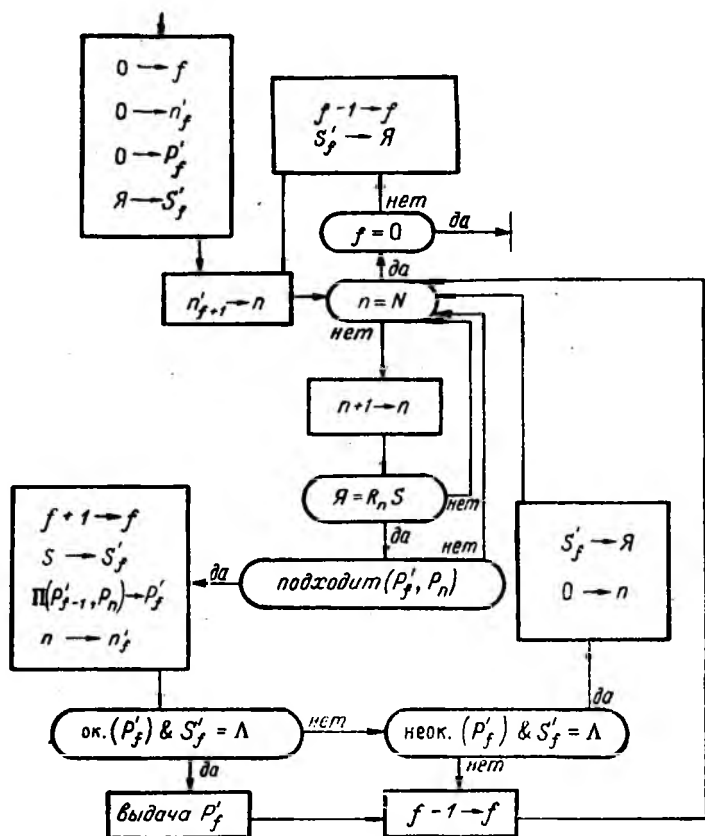


Схема алгоритма морфологического анализа (предложена Г. С. Цейтиним).  
Перечень символов, использованных для записи алгоритма,  
см. Приложение I.

При  $i = 0$  этот оператор имеет только положительный выход.

Словарь-таблица<sup>4</sup> задает множество правил перехода автомата из одного состояния в другое. Она организована следующим образом: строчки от I до N с информацией  $R_n$  в первом столбце,  $P_n$  во втором столбце (состоит из двух

<sup>4</sup> «Словарь основ» только иллюстрируется несколькими входами (см. Приложение II), «Словарь аффиксов» приводится полностью (см. Приложение III).

«подстолбцов»), где R — основа слова (или аффикс), P — информация об основе (аффиксе). В случае неоднозначности основы (аффикса) приводятся последовательно все варианты информации. Информация при основе (аффиксе) состоит из двух частей: сравниваемая информация (информация, которая должна проверяться на полное совпадение со сравниваемой информацией, выработанной на данный момент) — подстолбец P1 и приписываемая информация — подстолбец P2. Подстолбцы состоят из рубрик. Сравнимые признаки<sup>5</sup> (P1) могут быть существенными (P1.1) и несущественными (P1.2)<sup>6</sup> для присоединения и интерпретации данного аффикса. Приписываемые признаки (P2) разделены на три группы: собственно приписываемые признаки — они остаются как результат работы данного алгоритма в отличие от всех других признаков, которые стираются в процессе его работы (P2.1); признаки, которые на определенном этапе становятся сравниваемыми (P2.2)<sup>7</sup>; признак окончательности анализа (P2.3).

Таким образом, каждая строка словаря-таблицы должна описывать: 1) порождаемую цепочку (R); 2) предшествующее состояние автомата; 3) последующее состояние автомата. Фактически каждая строка описывает некоторое множество таких переходов с одинаковой порождаемой цепочкой.

Предшествующее состояние автомата во всех этих переходах характеризуется признаками из рубрик P1.1 и P1.2. Сравнимая информация должна содержать признаки, присутствие которых существенно (P1.1), и не должна содержать признаков, отсутствие которых существенно. Так как последних большинство, то вместо рубрики для этих признаков заведена рубрика для признаков, присутствие или отсутствие которых вообще несущественно (P1.2).

Последующее состояние автомата получается из предшествующего в результате следующих операций: признаки из P1 стираются, если а) рубрика P2.2 в предшествующем состоянии непустая или б) получен вариант анализа словоформы; признаки из P2.1, P2.2, P2.3 приписываются к информации; признак окончательности анализа из P2.3 стирается, когда получен вариант анализа словоформы.

<sup>5</sup> Каждый признак представлен комбинацией буквы и цифры.

<sup>6</sup> В «Словаре основ» сравниваемые признаки не разделяются на группы.

<sup>7</sup> Такая рубрика в «Словаре основ» не используется.

Список морфологических признаков<sup>8</sup>:

**M1** — признак субстантивного словоизменения. Этот признак задается при всех существительных; числительных: «пять»<sup>†</sup> «двадцать», «тридцать»; местоимениях: «я», «ты», «себе».

**M2** — признак адъективного словоизменения. Задается при всех прилагательных, в том числе местоименных; при количественных, относительных местоимениях, при личных местоимениях «он» (-а, -о, -и); при собирательных числительных.

**M3** — признак нестандартного словоизменения числительных. Задается при числительных: «два», «три», «четыре», «сорок», «девяносто», «сто».

**M4** — признак нестандартного словоизменения местоимений. Задается при местоимениях «мы», «вы».

**M5** — признак глагольного словоизменения. Задается при всех глаголах.

**M6** (не маркируется) — признак первого типа субстантивного словоизменения. Задается при существительных, относящихся в традиционной грамматике ко второму и третьему склонениям; при существительных типа «время», «чернила», «граждане»; при числительных, получивших признак **M1**.

**M7** — признак второго типа субстантивного словоизменения. Задается при существительных, относящихся в традиционной грамматике к первому склонению; при местоимениях, получивших признак **M1**.

**M10** — признак первого подтипа первого типа субстантивного словоизменения (-а — T2/M35.T4; -е — T6; -и — Г20Т1/М35.Г20Т4; -й — Т1/М35.Т4; -я — Т2/М35.Т4)<sup>9</sup>.

**M11** — признак второго подтипа первого типа субстантивного словоизменения (-а — T2/М35.Т4/Г20Т1/М35.Г20Т4; -е —

<sup>8</sup> Остальные, неморфологические, признаки перечислены в статье Т. Д. Логачевой, М. С. Першиковой «Описание эксперимента по упрощенному синтаксическому анализу русских научно-технических текстов» (наст. сб., стр. 102).

<sup>9</sup> В работе не дается содержательного описания подтипов, они заданы сугубо формально. Признак подтипа задается слову в том случае, если перечисленные в скобках аффиксы либо не встречаются в его парадигме, либо встречаются именно с теми значениями и теми ограничениями (ограничительные признаки отделены от значения аффикса точкой), которые указаны при аффиксах. Черточка над признаком означает 'не'.

Т6; -и — Т2/Т3/Т6/М40.Г20Т1/М40М35.Г20Т4; -й — Т1/М35.Т4; -я — Т2/М35.Т4/Г20Т1/М35.Г20Т4).

М12 — признак третьего подтипа первого типа субстантивного словоизменения (-и — Т6/М40.Г20Т1/М40М35.Г20Т4; -й — Т1/М35.Т4; -я — Т2/М35.Т4).

М13 — признак четвертого подтипа первого типа субстантивного словоизменения (-а — Т1/М35.Т4; -и — Т2/Т3/Т6; -я — Т1/М35.Т4).

М14 — признак пятого подтипа первого типа субстантивного словоизменения (-а — Т2/Г20Т1/М35.Г20Т4; -е — Т1/Т4/Т6; -я — Т2/М40.Г20Т1/М40.Г20Т4).

М15 — признак шестого подтипа первого типа субстантивного словоизменения (-е — Т1/М35.Т4; -и — Т6; -й — Г20Т2; -я — Т2/Г20Т1/Г20Т4).

М16 (не маркируется) — признак первого подтипа второго типа субстантивного словоизменения (-ей — Т5/М40.Г20Т2/М40М35.Г20Т4; -и — Т2/М40.Г20Т1/М40М35.Г20Т4; -я — Т1).

М17 — признак второго подтипа второго типа субстантивного словоизменения (-ей — Т5; и — Т2/Г20Т1/М35.Г20Т4; -я — Т1).

М20 — признак третьего подтипа второго типа субстантивного словоизменения (-ей — Т5; -и — Т2/Т3/Т6/М40.Г20Т1/М40М35.Г20Т4; -я — Т1).

М21 — признак четвертого подтипа второго типа субстантивного словоизменения (-я — Т2/Т4).

М22 (не маркируется) — признак первого типа адъективного словоизменения. Задается при качественных и относительных прилагательных.

М23 — признак второго типа адъективного словоизменения. Задается при притяжательных, местоименных прилагательных; при местоимениях и числительных, получивших признак М2.

М24 (не маркируется) — признак первого подтипа первого типа адъективного словоизменения (-ой — Т15Т2/Т15Т3/Т15Т5/Т15Т6).

М25 — признак второго подтипа второго типа адъективного словоизменения (-ой — Т14Т1/Т14Т4/Т15Т2/Т15Т3/Т15Т5/Т15Т6).

**М26** — признак первого подтипа второго типа адъективного словоизменения (-е — Т16Т1/Т16Т4; -его — Т14Т2/Т14Т4/Т16Т2; -ей — Т15Т2/Т15Т3/Т15Т5/Т15Т6; -ем — Т14Т6/Т16Т6; -ему — Т14Т3/Т16Т3; -и — Г20Т1/Г20Т4; -им — Т14Т5/Т16Т5/Г20Т3; -о — Т16Т1/Т16Т4; -ого — Т14Т2/Т14Т4/Т16Т2; -ом — Т14Т6/Т16Т6; -ому — Т14Т3/Т16Т3; -ым — Т14Т5/Т16Т5/Г20Т3).

**М27** — признак второго подтипа второго типа адъективного словоизменения (-е — М36.Т16Т1/М36.Т16Т4/Г20Т1/Г20Т4; -его — Т14Т2/Т16Т4/Т16Т2; -ей — Т15Т2/Т15Т3/Т15Т5/Т15Т6; -ем — М36.Т14Т6/М36.Т16Т6/Т14Т5/Т16Т5/Г20Т3; -ему — Т14Т3/Т16Т3; -о — Т16Т1/Т16Т4; -ого — Т14Т2/Т14Т4/Т16Т2; -ом — Т14Т6/Т16Т6; -ому — Т14Т3/Т16Т3).

**М30** — признак третьего подтипа второго типа адъективного словоизменения (-е — Т1/Т4; -его — Т2; -ем — Т5/Т6; -ему — Т3; -им — Т5; -о — Т1/Т4; -ого — Т2; -ом — Т6; -ому — Т3).

**М31** — признак четвертого подтипа второго типа адъективного словоизменения (-е — Т1; -ем — Т5; -им — Г20Т3; -о — Т1/Т4; -ого — Т2/Т4; -ом — Т6; -ому — Т3; -ым — Г20Т3).

**М32** — признак пятого подтипа второго типа адъективного словоизменения (-его — Т14Т2/Т14Т4/Т16Т2/Т16Т4; -ей — Т15Т3/Т15Т5/Т15Т6; -ем — Т14Т6/Т16Т6; -ему — Т14Т3/Т16Т3; -и — Г20Т1; -им — Т14Т5/Т16Т5/Г20Т3; -о — Т16Т1).

**М33** (не маркируется) — признак первого типа нестандартного словоизменения числительных (-а — Т14Т1/Т14Т4/Т16Т1/Т16Т4).

**М34** — признак второго типа нестандартного словоизменения числительных (-а — Г20Т2/Г20Т3/Г20Т5/Г20Т6).

**М35** — признак одушевленности.

**М36** — признак мягкости основы. Задается при существительных, имеющих в форме Г20Т3 аффикс -ям; при прилагательных, имеющих в форме Т16Т1 аффикс -ее или -е.

**М37** — признак твердости основы. Задается при существительных, основа которых не является мягкой и не является смешанной<sup>10</sup>.

**М40** — признак единичности. Задается при существительных, имеющих формы только единственного числа.

**М41** — признак возможности сравнительной степени. Задается при прилагательных, если: а) форма сравнительной степени образуется при помощи аффикса -ее; б) форма срав-

<sup>10</sup> Основа существительного считается смешанной, если слово в форме Г20Т3 имеет не -ям, а в остальной парадигме есть хотя бы один из аффиксов: -ев, -ей, -ем, -и.

нительной степени образуется при помощи аффикса **-е** и не попала в словарь целиком.

**M42** — признак возможности образования наречия от данного прилагательного.

**M43** — признак возможности образования наречия на **-о**.

**M44** — признак возможности образования наречия на **-е**.

**M45** — признак возможности образования наречия на **-и**.

**M46** — признак имперфективной интерпретации аффиксов **-а, -я**. Задается при глаголах, если при образовании видовой пары эти аффиксы служат показателями несовершенного вида.

**M47** — признак возможности образования причастия на **-т**.

**M50** — признак невозможности образования страдательного причастия настоящего времени, а также признак возможности нулевого аффикса в формах **T1** и **T4** у существительных, получивших признак **M10** или **M11**.

**M51** — признак невозвратности глагола, а также признак возможности нулевого аффикса в форме **T2Г20** у существительных, получивших признак **M10** или **M11**.

**M60** — признак окончательности анализа.

**M700** — признак возможности аффикса **-ся**.

#### Пример обработки словоформы описанным алгоритмом <sup>11</sup>

Словоформа «выполняемый». Заносим ее в массив для предположений (см. массив **B**, строка 0 на стр. 155).

В «Словаре основ» (строка 3) находим начальную цепочку **выполн-** со своей информацией. Заносим ее и остаток в массив **B** (строка 1).

Из остатка словоформы выделяем первую букву **-я**. В «Словаре аффиксов» (строка 144) находим аффикс **-я**, при котором информация подстолбца **P1** совпадает с информацией подстолбца **P1** при основе (в последней не должно встречаться признаков, не указанных в **P1** при аффиксе). Новую начальную цепочку и новый остаток заносим в массив **B** (строка 2). Информация при ней формируется следующим образом: в подстолбец **P1** заносятся признаки **M5Г31** из рубрики **P2.2** «Словаря аффиксов»; в рубрику **P2.1** заносится признак **V** из рубрики **P2.1** при основе и признак

<sup>11</sup> Пример здесь излагается таким образом, чтобы читатель смог вручную проверить работу алгоритма и таблиц.



Г31 из рубрики P2.1 при аффиксе. Графа S<sup>1</sup> непустая (есть остаток), и в выработанной информации нет признака окончательности анализа M60.

Выделяем из остатка следующую букву: **-е**. В «Словаре аффиксов» не находим **-е**, допустимого выработанной информацией. Выделяем еще одну букву: получаем аффикс **-ем**. На строке 182 находим допустимый аффикс. Новую начальную цепочку, сформированную информацию при ней и новый остаток заносим в массив В (строка 3, вариант 1). Остаток в S' есть, и нет признака M60.

Выделяем следующую букву: **-ы**. Ни одного допустимого **-ы** нет. Выделяем последнюю букву, получаем аффикс **-ый**. Ни одного допустимого **-ый** также нет.

Делаем шаг назад в поисках другого допустимого варианта аффикса **-ем**. Находим его на строке 184. Новую информацию при **выполняем-** смотрим в массиве В, строка 3, вариант 2. Остаток еще есть, но есть признак M60, поэтому еще раз ищем вариант аффикса **-ем**. Находим его на строке 186. Новую информацию при **выполняем-** смотрим в массиве В, строка 3, вариант 3. Есть остаток, и нет признака M60.

Снова выделяем **-ы**. На строке 116 находим допустимый аффикс (M2M24 не маркируются). Цепочку **выполняем-**, информацию при ней и остаток заносим в массив В (строка 4, вариант 1). Есть остаток, но есть признак M60. Другого допустимого **-ы** нет.

Снова выделяем последнюю букву остатка и получаем аффикс **-ый**. На строке 252 находим **-ый**, допустимый информацией при **выполняем-**. Цепочку **выполняемый** и информацию при ней заносим в массив В (строка 4, вариант 2). Остатка больше нет, и есть признак M60.

Таким образом, получен вариант анализа словоформы «**выполняемый**» — VT1T12T14G26G31G35 (признаки подстолбца P1 и рубрики P2.3 стираются) — глагол, причастие настоящего времени, страдательного залога, несовершенного вида, мужского рода, именительного падежа.

Делаем шаг назад в поисках другого допустимого варианта **-ый**. Находим его на строке 253. Цепочку **выполняемый** и новую информацию при ней заносим в массив В (строка 4, вариант 3). Это второй вариант анализа словоформы «**выполняемый**» — VT4T12T14G26G31G35 — глагол, причастие настоящего времени, страдательного залога, несовершенного вида, мужского рода, винительного падежа. Других **-ый** нет.

Убеждаемся также, что нет других -ем и нет больше допустимых -я.

Делаем последний шаг назад в поисках другого варианта информации при основе выпол- или другого варианта начальной цепочки. Ни того, ни другого не находим. На этом работа над словоформой «выполняемый» заканчивается.

#### Описание машинного эксперимента

Эксперимент проводился на материале словаря основ в 260 основ и входного текста в 982 словоформы (80 предложений), взятого произвольно.<sup>12</sup> 518 словоформ обработались правильно. 442 словоформы не обработались совсем, так как в словаре основ не оказалось соответствующих входов. 22 словоформы (4 из них встретились дважды), наряду с правильным вариантом, получили 1—2 лишних варианта (в ряде случаев тоже правильных, в ряде случаев неправильных). Рассмотрим эти случаи.

Словоформы: «выполненные», «выполненных», «данные», «использовались», «обеспечивает», «обеспечивают», «предложенной», «предложенные» — получили второй правильный вариант из-за отсутствия в настоящий момент у выделенных аффиксов признака, по которому они должны присоединяться только к определенным глагольным основам.

По этой же причине получили неправильные варианты словоформы: «обеспечение» (причастие), «предложенной» (глагол в повелительном наклонении), «применение» (причастие), «применением» (глагол второго лица множественного числа, причастие настоящего времени), «применении» (глагол в повелительном наклонении).

Словоформы: «данные», «даны», «использован», получили неправильный вариант (причастие несовершенного вида) из-за отсутствия у основы признака невозможности образования такого причастия.

Словоформа «можно» получила неправильный вариант (причастие): у основы мож- должен быть признак невозможности причастия.

Словоформа «именно» получила неправильный вариант (причастие прошедшего времени): у основы име- должен быть признак невозможности такого причастия.

<sup>12</sup> На БЭСМ-3М на обработку текста такой длины затрачивается около 20 минут.

Массив В

f	R'	S'	P			n'	
			P1	P2			
				P2.1	P2.3		
0		выполняемый	—	—	—		
1	выполн-	-яемый	M5M46	V	—	3	
2	выполня-	-емый	M5Г31	VG31	—	144	
3	выполняем-	-ый	*M700	VG20Г21Г26Г31	—	182	
			**M5Г31	VT10T12T14Г26Г31Г35	M60	184	
			***M2M22M24	VT12Г26Г31Г35	—	186	
4	*выполняемы-	-й	M2M22M24	VT10T12Г20Г26Г31Г35	M60	116	
			**выполняемый	M2M22M24	VT1T12T14Г26Г31Г35	M60	252
			***выполняемый	M2M22M24	VT4T12T14Г26Г31Г35	M60	253

Примечание. \*Вариант 1; \*\*вариант 2; \*\*\* вариант 3.

Словоформы «дает», «дается» получили неправильный вариант (будущее время); этой основе нужен индивидуальный признак.

Таким образом, эксперимент показал, что усовершенствование описанного морфологического анализа должно пойти по пути расширения системы признаков.

#### Приложение I

Перечень символов, использованных для записи алгоритма

- Я — рабочее слово (остаток).
- В — массив для предположений (см. пример).
- R' — графа массива В для хранения порождаемых цепочек.
- S — остаток.
- S' — графа массива В для хранения остатка.
- P' — графа массива В для хранения промежуточной информации.
- n — номер рабочей строки словаря-таблицы.
- n' — графа массива В для хранения n (в первой строке номер строки «Словаря основ», в остальных — «Словаря аффиксов»).
- i — номер строки массива В.
- R — основа слова (аффикс).
- P — информация при основе (аффиксе).
- N — номер последней строки словаря-таблицы.
- П(P, P') — новый вариант P'.
- Λ — информация об отсутствии остатка.
- ок. — информация об окончательности анализа.
- неок. — информация о неокончателности анализа.

#### Приложение II

##### «Словарь основ»

n	R	P		
		P1	P2	
			P2.1	P2.3
1	анализ	M1M10M37M50	ST14	—
2	без	—	F	M60
3	выполн.	M5M46	V	—
4	газов	M2	A	—
5	дву(x)	M3	Q	—

«Словарь аффиксов»

n	R	P					
		P1		P2			
		P1.1	P1.2	P2.1	P2.2	P2.3	
1	-a	M1	M10M11M12M13M14M15M35M37 M50M51		T2	—	M60
2	-a	M1M35	M10M11M12M13M14M15M37M50		T4	—	M60
3	-a	M1	M11M13M35M37M50M51		T1Г20	—	M60
4	-a	M1	M11M13M37M50M51		T4Г20	—	M60
5	-a	M1M7	M16M17M35M37M40		T1	—	M60
6	-a	M2	M24M25M36M41M42M43M44M45 M51		T10T15	—	M60
7	-a	M2M23	M26M27M30M31M32		T1T15	—	M60
8	-a	M3		—	T1T14	—	M60
9	-a	M3		—	T4T14	—	M60
10	-a	M3		—	T1T16	—	M60
11	-a	M3		—	T4T16	—	M60
12	-a	M3M34		—	T2Г20	—	M60
13	-a	M3M34		—	T3Г20	—	M60
14	-a	M3M34		—	T5Г20	—	M60
15	-a	M3M34		—	T6Г20	—	M60
16	-a	M5Г27		—	T15Г27	M700	—
17	-a	M5Г27Г31		—	T15Г27Г31	M700	—
18	-a	M5Г31	M47M50		T13Г31	M700	—
19	-a	M1M14	M35M36		T1Г20	—	M60

n	R	P				
		P1		P2		
		P1.1	P1.2	P2.1	P2.2	P2.3
20	-a	M5M46M50	—	Г31	M5M50Г31	—
21	-a	M5M46	—	Г31	M5Г31	—
22	-a	M5	Г31	—	M5	—
23	-a	M1M14	M36	T4Г20	—	M60
24	-b	M5	M47M50Г31	T13	—	M60
25	-e	M1	M6M7M10M11M12M16M17M20 M21M35M36M37M40M50M51	T6	—	M60
26	-e	M1M14	M10M11M12M13M15M35M36M40	T1	—	M60
27	-e	M1M14	M10M11M12M13M15M35M36M40	T4	—	M60
28	-e	M1M14	—	T6	—	M60
29	-e	M1M7	M16M17M20M21M35M36M37M40	T3	—	M60
30	-e	M2M36	M41M42M44M51	T10T16	—	M60
31	-e	M2M36M41 M42M44	—	B	—	M60
32	-e	M2M41	M24M25M36M42	Г24	—	M60
33	-e	M2M41	M24M25M36M42	B Г24	—	M60
34	-e	M2M23M36	M26M27M30	T1T16	—	M60
35	-e	M2M23M36	M26M27M30	T4T16	—	M60
36	-e	M2M23M27	M36	T1Г20	—	M60
37	-e	M2M23M27	M36	T4Г20	—	M60
38	-e	M2M23	M26M27M30M31M32M36	T1	—	M60
39	-e	M2M23M30 M36	—	T4	—	M60
40	-e	M3	—	T1T15	—	M60

41	-e	M3	—	T4T15	—	M60
42	-e	M5	M46M50	—	M5	—
43	-и	M1	M6M7M10M11M12M16M17M20 M21M35M36M50M51	T1Г20	—	M60
44	-и	M1	M6M7M10M11M12M16M17M20 M21M36M50M51	T4Г20	—	M60
45	-и	M1	M6M7M11M16M17M20M21M35 M36M40	T2	—	M60
46	-и	M1	M6M7M11M16M20M35M36M40	T3	—	M60
47	-и	M1	M6M7M10M11M12M13M14M15 M16M20M35M36M40	T6	—	M60
48	-и	M1M13	—	T2	—	M60
49	-и	M1M13	—	T3	—	M60
50	-и	M1M13	—	T6	—	M60
51	-и	M2	M24M25M36M41M42M43M44M45 M51	T10Г20	—	M60
52	-и	M2M42M45	—	B	—	M60
53	-и	M2M23M26	M31M36	T1Г20	—	M60
54	-и	M2M23M26	M36	T4Г20	—	M60
55	-и	M5	M46M47M50	Г33	M700	—
56	-и	M5Г31	M47M50	Г31Г33	M700	—
57	-и	M5Г27	—	Г20Г27	M700	—
58	-и	M5Г27Г31	—	Г20Г27Г31	M700	—
59	-и	M5	M46	—	M5	—
60	-и	M5Г31	—	Г31	M5	—
61	-й	M1M36	M10M11M12M35M40	T1	—	M60
62	-й	M1M36	M10M11M12M40	T4	—	M60
63	-й	M1M36	M6M7M10M11M12M13M14M15 M16M17M20M21M35	T2Г20	—	M60

n	R	P				
		P1		P2		
		P1.1	P1.2	P2.1	P2.2	P2.3
64	-й	M1M7M20 M35M36	—	T4Г20	—	M60
65	-й	M2M23M26 M36	—	T1T14	—	M60
66	-й	M2M23M26 M36	—	T4T14	—	M60
67	-й	M5	M46M47M50	Г33	M700	—
68	-й	M5Г31	M47M50	Г31Г33	M700	—
69	-л	M5	M47M50	T14Г27	M700	—
70	-л	M5Г31	M47M50	T14Г27Г31	M700	—
71	-л	M5	M47M50	Г27	M5Г27	—
72	-л	M5Г31	M47M50	Г27Г31	M5Г27	—
73	-м	M3	—	T3Г20	—	M60
74	-м	M4	—	T3	—	M60
75	-н	M5	M46M50	T10T12T14Г27 Г35	—	M60
76	-н	M5Г31	M50	T10T12T14Г27 Г31Г35	—	M60
77	-н	M5	M46M50	T12Г27Г35	M2M22M24	—
78	-н	M5Г31	M50	T12Г27Г31Г35	M2M22 M24	—
79	-н	M5	M46M50	—	M5	—
80	-о	M1	M10M11M12M37M51	T1	—	M60
81	-о	M1	M10M11M12M37M51	T4	—	M60



82	-o	M2	M24M25M31M36M41M42M43M51	T10T16	—	M60
83	-o	M2M42M43	M24M25M36M41	B	—	M60
84	-o	M2M23M26	M31	T1T16	—	M60
85	-o	M2M23M26	—	T4T16	—	M60
86	-o	M2M23M30	—	T1	—	M60
87	-o	M2M23M30	—	T4	—	M60
88	-o	M3M34	—	T1	—	M50
89	-o	M3M34	—	T4	—	M60
90	-o	M5Г27	—	T16Г27	M700	—
91	-o	M5Г27Г31	—	T16Г27Г31	M700	—
92	-o	M5	Г31	—	M5	—
93	-o	M2M23M27	—	T1T16	—	M60
94	-o	M2M23M27	—	T4T16	—	M60
95	-o	M2M23M31	—	T4	—	M60
96	-o	M2M23M31	—	T1	—	M60
97	-r	M5M47	M50	T10T12T14Г27	—	—
				Г35	—	M60
98	-r	M5M47M50Г31	—	T10T12T14Г27	—	—
				Г31Г35	—	M60
99	-r	M5M47	M50	T12Г27Г35	M2M22M24	—
100	-r	M5M47M50Г31	—	T12Г27Г31Г35	M2M22M24	—
101	-y	M1	M10M11M12M13M14M15M35M37			
			M50M51	T3	—	M60
102	-y	M1M7	M16M17M35M37M40	T4	—	M60
103	-y	M2M23	M26M27M30	T4T15	—	M60
104	-y	M5	M46M47M50	Г21Г30	M700	—
105	-y	M5Г31	M47M50	Г21Г26Г31	M700	—
106	-y	M5	M47M50Г31	Г31	M5	—
107	-y	M5	M47M50	—	M5	—

n	R	P					
		P1			P2		
		P1.1	P1.2		P2.1	P2.2	P2.3
108	-x	M3		—	T2Г20	—	M60
109	-x	M3		—	T4Г20	—	M60
110	-x	M3		—	T6Г20	—	M60
111	-ш	M5	M50		T12Г27	M2M22 M24M36	—
112	-ш	M5Г31	M50		T12Г27Г31	M2M22M24 M36	—
113	-ы	M1	M6M7M10M16M17M35M37M50		T1Г20	—	M60
114	-ы	M1	M6M7M10M16M17M37M50		T4Г20	—	M60
115	-ы	M1M7	M16M17M35M37M40		T2	—	M60
116	-ы	M2	M24M25M36M41M42M43M44M45 M51		T10Г20	—	M60
117	-ы	M2M23M26		—	T1Г20	—	M60
118	-ы	M2M23M26		—	T4Г20	—	M60
119	-ы	M5	M47M50		—	M5	—
120	-ы	M5M50Г31	M47		Г31	M5	—
121	-ь	M1	M10M11M12M35M36M40		T1	—	M60
122	-ь	M1	M10M11M12M35M36M40		T4	—	M60
123	-ь	M1M10		—	Г20	M1M6M13	—
124	-ь	M1M7M17 M36	M35		T2Г20	—	M60
125	-ь	M1M7M17 M35M36		—	T4Г20	—	M60

126	-б	M5	M46M47M50	G33	M700	—
127	-б	M5Г31	M47M50	Г31Г33	M700	—
128	-б	M5M50Г31	—	T7Г31	M700	—
129	-б	M5M50	—	T7	M700	—
130	-ю	M1M36	M10M11M12M13M14M15M35M40	T3	—	M60
131	-ю	M1M7M36	M16M20M35M40	T4	—	M60
132	-ю	M2M23M36	M26M27M30	T4T15	—	M60
133	-ю	M5	M46M47M50	Г21Г30	M700	—
134	-ю	M5Г31	M47M50	Г21Г26Г31	M700	—
135	-ю	M5Г31	—	Г31	M5	—
136	-я	M1	M10M11M12M13M14M15M35M36 M37M40	T2	—	M60
137	-я	M1M35	M10M11M12M13M14M15M36M37	T4	—	M60
138	-я	M1	M11M13M35M36	T1Г20	—	M60
139	-я	M1	M11M13M36	T4Г20	—	M60
140	-я	M1M7M36	M16M17M20M21M35M40	T1	—	M60
141	-я	M2M36	M41M42M44M51	T10T15	—	M60
142	-я	M2M23M36	M26M27M30	T1T15	—	M60
143	-я	M5Г31	M47M50	T13Г31	M700	—
144	-я	M5M46	M47M50	Г31	M5Г31	—
145	-я	M5	Г31	—	M5	—
146	-я	M1M14M36	M11M35	T1Г20	—	M60
147	-я	M1M14M36	M11	T4Г20	—	M60
148	-я	M1M7M21	M10M11M12M13M14M15	T2	—	M60
149	—	—	—	—	—	—
150	-аш	M5Г31	M50	T12Г26Г31	M2M22 M24M36	—

и	R	P						
		P1			P2			
		P1.1	P1.2		P2.1	P2.2	P2.3	
151	-ая	M2	M24M25M36M41M42M43M44M45					
			M51		T1T15	M700	—	
152	-ая	M2M36	M51		T1T15	M700	—	
153	-ва	M5	—		Г31	M5Г31	—	
154	-вш	M5	M46M47M50		T12Г27	M2M22M24		
						M36	—	
155	-вш	M5Г31	M47M50		T12Г27Г31	M2M22M24		
						M36	—	
156	-ев	M1	M10M11M12M13M14M15M35M36					
			M50		T2Г20	—		M60
157	-ев	M1M35	M10M11M12M13M14M15M36M50		T4Г20	—		M60
158	-ев	M1M36	—		T2	—		M60
159	-ее	M2M36	M41M42M43M44M45M51		T1T16	M700	—	
160	-ее	M2M36	M41M42M43M44M45M51		T4T16	M700	—	
161	-ее	M2M41	M24M25M36M42M43M44M45		Г24	—		M60
162	-ее	M2M41M42	M24M25M36M43M44M45		В Г24	—		M60
163	-ее	M2M23M32						
		M36	—		T2T15	—		M60
164	-ее	M2M23M32						
		M36	—		T4T15	—		M60
165	-ей	M1	M6M7M10M11M12M13M14M15					
			M35M36M50		T2Г20	—		M60

166	-ей	M1M35	M6M7M10M11M12M13M14M15 M36M50	T4Г20	—	M60
167	-ей	M1M7	M16M17M20M21M35M36	T5	—	M60
168	-ей	M2M36	M22M23M26M27M30M41M42M43 M44M45M51	T2T15	M700	—
169	-ей	M2M36	M22M23M26M27M30M31M32M41 M42M43M44M45M51	T3T15	M700	—
170	-ей	M2M36	M22M23M26M27M30M31M32M41 M42M43M44M45M51	T5T15	M700	—
171	-ей	M2M36	M22M23M26M27M30M31M32M41 M42M43M44M45M51	T6T15	M700	—
172	-ем	M1	M10M11M12M13M14M15M35M36 M40M50	T5	—	M60
173	-ем	M1M13	—	T5	—	M60
174	-ем	M2M36	M22M23M26M27M30M31M32M41 M42M43M44M45M51	T6T14	M700	—
175	-ем	M2M36	M22M23M26M27M30M31M32M41 M42M43M44M45M51	T6T16	M700	—
176	-ем	M2M23M27	M36	T5T14	—	M60
177	-ем	M2M23M27	M36	T5T16	—	M60
178	-ем	M2M23M27	M36	T3Г20	—	M60
179	-ем	M2M23	M26M27M30M31M32M36	T5	—	M60
180	-ем	M2M23M30 M36	—	T6	—	M60
181	-ем	M5	M46M47M50	Г20Г21Г30	M700	—
182	-ем	M5Г31	M47M50	Г20Г21Г26Г31	M700	—
183	-ем	M5	M46M47	T10T12T14Г26 Г35	—	M60

n	R	P				
		P1		P2		
		P1.1	P1.2	P2.1	P2.2	P2.3
184	-ем	M5Г31	—	T10T12T14Г26 Г31Г35	—	M60
185	-ем	M5	M46M47	T12Г26Г35	M2M22M24	—
186	-ем	M5Г31	—	T12Г26Г31Г35	M2M22M24	—
187	-ен	M1M13	—	—	M1M6M13	—
188	-ет	M5	M46M47M50	Г23Г30	M700	—
189	-ет	M5Г31	M47M50	Г23Г26Г31	M700	—
190	-ех	M2M23M27	M36	T2Г20	—	M60
191	-ех	M2M23M27	M36	T4Г20	—	M60
192	-ех	M2M23M27	M36	T6Г20	—	M60
193	-ие	M2	M24M25M36M41M42M43M44M45 M51	T1Г20	M700	—
194	-ие	M2	M24M25M36M41M42M43M44M45 M51	T4Г20	M700	—
195	-ий	M2	M22M23M26M36M41M42M43M44 M45M51	T1Г14	M700	—
196	-ий	M2	M22M23M26M36M41M42M43M44 M45M51	T4Т14	M700	—
197	-им	M2	M22M23M24M25M26M30M31M32M36 M41M42M43M44M45M51	T5Т14	M700	—
198	-им	M2	M22M23M24M25M26M30M31M32M36 M41M42M43M44M45M51	T5Т16	M700	—

199	-ИМ	M2	M22M23M24M25M26M30M31M32M36 M41M42M43M44M45M51	T3Г20	M700	—
200	-ИМ	M2M23M30	—	T5	—	M60
201	-ИМ	M5	M46M47M50	Г20Г21Г30	M700	—
202	-ИМ	M5Г31	M50	Г20Г21Г26Г31	M700	—
203	-ИМ	M5	M46M47	T10T12T14Г26 Г35	—	M60
204	-ИМ	M5Г31	—	T10T12T14Г26 Г31Г35	—	M60
205	-ИМ	M5	M46M47	T12Г26Г35	M2M22M24	—
206	-ИМ	M5Г31	—	T12Г26Г31Г35	M2M22M24	—
207	-ИМ	M5M46	—	—	M5	—
208	-ИМ	M5M46M47	—	—	M5M47	—
209	-ИТ	M5	M46M47M50	Г23Г30	M700	—
210	-ИТ	M5Г31	M50	Г23Г26Г31	M700	—
211	-ИХ	M2	M22M23M24M25M26M30M31M32 M36M41M42M43M44M45M51	T2Г20	M700	—
212	-ИХ	M2	M22M23M24M25M26M30M31M32 M36M41M42M43M44M45M51	T4Г20	M700	—
213	-ИХ	M2	M22M23M24M25M26M30M31M32 M36M41M42M43M44M45M51	T6Г20	M700	—
214	-МИ	M4	—	T5	—	M60
215	-МЯ	M3	—	T5Г20	—	M60
216	-НН	M5	M46M50	T12Г27Г35	M2M22M24	—
217	-НИ	M5Г31	M50	T12Г27Г31Г35	M2M22M24	—
218	-ОВ	M1	M10M11M12M35M37M50	T2Г20	—	M60

n	R	P				
		P1		P2		
		P1.1	P1.2	P2.1	P2.2	P2.3
219	-об	M1M35	M10M11M12M37M50	T4Г20	—	M60
220	-оe	M2	M22M23M24M25M26M27M30M41 M42M43M44M45M51	T1T16	—	M60
221	-оe	M2	M22M23M24M25M26M27M30M41 M42M43M44M45M51	T4T16	—	M60
222	-ођ	M1M7	M16M17M20M21M35M37M40	T5	—	M60
223	-ођ	M2	M22M23M24M25M26M27M30M41 M42M43M44M45M51	T2T15	—	M60
224	-ођ	M2	M22M23M24M25M26M27M30M41 M42M43M44M45M51	T3T15	—	M60
225	-ођ	M2	M22M23M24M25M26M27M30M41 M42M43M44M45M51	T5T15	—	M60
226	-ођ	M2	M22M23M24M25M26M27M30M41 M42M43M44M45M51	T6T15	—	M60
227	-ођ	M2M25	M41M42M43M51	T1T14	—	M60
228	-ођ	M2M25	M41M42M43M51	T4T14	—	M60
229	-ом	M1	M13M14M15M35M37M50M51	T5	—	M60
230	-ом	M2	M24M25M41M42M43M44M45M51	T6T14	—	M60
231	-ом	M2	M24M25M41M42M43M44M45M51	T6T16	—	M60
232	-ом	M2M23M30	—	T6	—	M60
233	-ом	M5M50Г31	—	T10T12T14Г26 Г31Г35	—	M60
234	-ом	M5M50Г31	—	T12Г26Г31Г35	M2M22M24	—



235	-ом	M2M23M26	—	T6T14	—	M60
236	-ом	M2M23M26	—	T6T16	—	M60
237	-ом	M2M23M27	—	T6T14	—	M60
238	-ом	M2M23M27	—	T6T16	—	M60
239	-ом	M2M23M31	—	T6	—	M60
240	-ти	M5	M46M47M50	T7	M700	—
241	-ти	M5Г31	M47M50	T7Г31	M700	—
242	-ть	M5	M46M47M50	T7	M700	—
243	-ть	M5Г31	M47M50	T7Г31	M700	—
244	-ут	M5	M46M47M50	Г20Г23Г30	M700	—
245	-ут	M5Г31	M47M50	Г20Г23Г26Г31	M700	—
246	-ущ	M5Г31	M47M50	T12Г26Г31	M2M22M24 M36	—
247	-ую	M2	M24M25M36M41M42M43M44M45 M51	T4T15	M700	—
248	-ши	M5	M50	T13	M700	—
249	-ши	M5Г31	M50	T13Г31	M700	—
250	-ые	M2	M24M25M36M41M42M43M44M45 M51	T1Г20	—	M60
251	-ые	M2	M24M25M36M41M42M43M44M45 M51	T4Г20	—	M60
252	-ый	M2	M36M41M42M43M44M45M51	T1Г14	—	M60
253	-ый	M2	M36M41M42M43M44M45M51	T4Т14	—	M60
254	-ым	M2	M22M23M24M25M26M36M41M42 M43M44M45M51	T5T14	—	M60
255	-ым	M2	M22M23M24M25M26M36M41M42 M43M44M45M51	T5T16	—	M60

n	R	P				
		P1		P2		
		P1.1	P1.2	P2.1	P2.2	P2.3
256	-ым	M2	M22M23M24M25M26M31M32M36 M41M42M43M44M45M51	T3Г20	—	M60
257	-ыя	M2	M22M23M24M25M26M31M32M36 M41M42M43M44M45M51	T2Г20	—	M60
258	-ых	M2	M22M23M24M25M26M31M32M36 M41M42M43M44M45M51	T4Г20	—	M60
259	-ых	M2	M22M23M24M25M26M31M32M36 M41M42M43M44M45M51	T6Г20	—	M60
260	-ью	M1M11	M35M36M40	T5	—	M60
261	-ют	M5	M46M47M50	Г20Г23Г30	M700	—
262	-ют	M5Г31	M47M50	Г20Г23Г26Г31	M700	—
263	-ющ	M5Г31	M47M50	T12Г26Г31	M2M22M24 M36	—
264	-юю	M2M36	M41M42M44M51	T4Т15	—	M60
265	-ям	M1M36	M6M7M10M11M12M13M14M15 M16M17M20M21M35	T3Г20	—	M60
266	-ят	M5	M46M47M50	Г23Г30	M700	—
267	-ят	M5Г31	M50	Г23Г26Г31	M700	—
268	-ях	M1M36	M6M7M10M11M12M13M14M15 M16M17M20M21M35	T6Г20	—	M60
269	-ящ	M5Г31	M50	T12Г26Г31	M2M22M24 M36	—
270	-яя	M2M36	M41M42M44M51	T1Т15	—	M60

271	-айш	М2	М51	Г25	М2М22М24 М36	—
272	-ами	М1	М6М7М10М11М12М13М14М15 М16М17М35М37М50М51	Т5Г20	—	М60
273	-вши	М5	М46М47М50	Т13	М700	—
274	-вши	М5Г31	М47М50	Т13Г31	М700	—
275	-ева	М5	М47М50	Г31	М5Г31	—
276	-ева	М5	—	—	М5	—
277	-его	М2М36	М22М23М26М27М30М31М32М41 М42М43М44М45М51	Т2Т14	М700	—
278	-его	М2М36	М22М23М26М27М30М31М32М41 М42М43М44М45М51	Т2Т16	М700	—
279	-его	М2М36	М22М23М26М27М30М31М32М41 М42М43М44М45М51	Т4Т14	М700	—
280	-его	М2М23М30 М36	—	Т2	—	М60
281	-его	М2М23М32 М36	—	Т4Т16	—	М60
282	-ейш	М2	М24М25М36М41М42М43М44М45	Г25	М2М22М24 М36	—
283	-еми	М2М23М27	М36	Т5Г20	—	М60
284	-ему	М2М36	М22М23М26М27М30М31М32М41 М42М43М44М45М51	Т3Т14	М700	—
285	-ему	М2М36	М22М23М26М27М30М31М32М41 М42М43М44М45М51	Т3Т16	М700	—
286	-ему	М2М23М30 М36	—	Т3	—	М60

n	R	P				
		P1		P2		
		P1.1	P1.2	P2.1	P2.2	P2.3
287	-ете	M5	M46M47M50	Г20Г22Г30	M700	—
288	-ете	M5Г31	M47M50	Г20Г22Г26Г31	M700	—
289	-ешь	M5	M46M47M50	Г22Г30	M700	—
290	-ешь	M5Г31	M47M50	Г22Г26Г31	M700	—
291	-ива	M5	M47Г31	Г31	M5Г31	—
292	-има	M5M47	—	Г31	M5Г31	—
293	-ими	M2	M22M23M24M25M26M30M31M32 M36M41M42M43M44M45M51	T5Г20 Г31	M700 M5Г31	—
294	-ина	M5	M47Г31	Г31	M5Г31	—
295	-ите	M5	M46M47M50	Г20Г22Г30	M700	—
296	-ите	M5Г31	M50	Г20Г22Г26Г31	M700	—
297	-ишь	M5	M46M47M50	Г22Г30	M700	—
298	-ишь	M5Г31	M50	Г22Г26Г31	M700	—
299	-ова	M5	Г31	Г31	M5Г31	—
300	-ова	M5	—	—	M5	—
301	-ого	M2	M24M25M41M42M43M44M45M51	T2T14	—	M60
302	-ого	M2	M24M25M41M42M43M44M45M51	T4T14	—	M60
303	-ого	M2	M24M25M41M42M43M44M45M51	T2T16	—	M60
304	-ого	M2M23M31	—	T2	—	M60
305	-ого	M2M23M31	—	T4	—	M60
306	-ого	M2M23M26	—	T2T14	—	M60
307	-ого	M2M23M26	—	T4T14	—	M60
308	-ого	M2M23M26	—	T2T16	—	M60

309	-Его	M2M23M27	—	T2T14	—	M60
310	-ого	M2M23M27	—	T4T14	—	M60
311	-ого	M2M23M27	—	T2T16	—	M60
312	-ого	M2M23M30	—	T2	—	M60
313	-ючи	M5Г31	M50	T13Г31	—	M60
314	-яни	M1M36	M6M7M10M11M12M13M14M15 M16M17M20M21M35	T5Г20	—	M60
315	-сыва	M5	—	Г31	M5Г31	—
316	-овыва	M5	—	Г31	M5Г31	—
317	нулевой аффикс	M1	M10M11M12M35M37M40M50	T1	—	M60
318	»	M1	M10M11M12M37M50	T4	—	M60
319	»	M1	M6M7M10M11M12M13M14M15 M16M17M35M37M51	T2Г20	—	M60
320	»	M2M23M26	M37	T1T14	—	M60
321	»	M2M23M26	M37	T4T14	—	M60
322	»	M700	Г27Г31	—	—	M60
323	-ам	M1	M6M7M10M11M12M13M14M15 M16M17	T3Г20	—	M60
324	-ат	M5	M46M47M50	Г20Г23Г30	M700	—
325	-ат	M5Г31	M50	Г20Г23Г26Г31	M700	—
326	-ах	M1	M6M7M10M11M12M13M14M15 M16M17M35M37M50M51	T6Г20	—	M60
327	-от	M2M23	M26M27M30	T1T14	—	M60
328	-от	M2M23	M26M27M30	T4T14	—	M60
329	-сь/ся	M700M51	Г27Г31	—	—	M60
330	-сь/ся	M700	Г27Г31	Г34	—	M60

n	R	P				
		P1		P2		
		P1.1	P1.2	P2.1	P2.2	P2.3
331	-ге	M5	—	Г20	M700	—
332	-ге	M5	Г31	Г20Г31	M700	—
333	-ому	M2	M24M25M41M42M43M44M45M51	T3T14	—	M60
334	-ому	M2	M24M25M41M42M43M44M45M51	T3T16	—	M60
335	-ому	M2M23 M30	—	T3	—	M60
336	-ому	M2M23M26	—	T3T14	—	M60
337	-ому	M2M23M26	—	T3T16	—	M60
338	-ому	M2M23M27	—	T3T14	—	M60
339	-ому	M2M23M27	—	T3T16	—	M60
340	-ому	M2M23M31	—	T3	—	M60
341	-учи	M5M50Г31	—	T13Г31	—	M60
342	-ыва	M5	M47Г31	Г31	M5Г31	—
343	-ыми	M2	M22M23M24M25M26M30M31M32 M36M41M42M43M44M45M51	T5Г20	—	M60
344	-ьми	M1M11M36	M35	T5Г20	—	—
345	с	M4	—	T2	—	M60
346	с	M4	—	T4	—	M60
347	с	M4	—	T6	—	M60

## СОДЕРЖАНИЕ

Предисловие . . . . .	3.
Е. С. Антипенко. Скобочные конструкции в системе автоматического анализа . . . . .	7
В. Д. Буторов. Отождествление и связывание переменных при семантическом анализе математических текстов . . . . .	11
А. Г. Давтян. Опыт сегментации сложных предложений на основе ограничительных знаков препинания . . . . .	26
Г. М. Ильин. Семантика информационных языков . . . . .	33.
Г. М. Кириллова. Анализ формул при автоматическом переводе научно-технических текстов . . . . .	37
Н. Л. Кириченко. Предикат тождества и языковые способы его выражения . . . . .	43.
М. А. Козаринский. Мотивация составных математических терминов . . . . .	52
Н. Д. Кремнева. Словоизменение и словообразование с позиций автоматического анализа . . . . .	58
Л. Н. Смирнова. Об одном типе слов-заместителей в связанном тексте . . . . .	63
Б. М. Лейкина. Задачи этапа преобразования в системе АП . . . . .	75
Б. М. Лейкина, М. И. Откупщикова, Ф. И. Случеский, Г. С. Цейтин, Б. А. Щербатов. Анализ некоторых числовых характеристик статистического исследования речи при патологии мышления . . . . .	93
Т. Д. Логачева, М. С. Першикова. Описание эксперимента по «упрощенному» синтаксическому анализу русских научно-технических текстов . . . . .	102
И. И. Матвеева. Способы выражения семантических отношений в информационном языке . . . . .	111
Т. Н. Никитина, М. И. Откупщикова. Следование из простого предложения (К вопросу о слабоуправляемых элементах в простом предложении) . . . . .	118

И. М. Новицкая. Целые предложения в функции средств межфразовой связи . . . . .	124
М. И. Откупщикова. Способы развертывания конструкций с уподобляющими местоимениями во внутренней структуре текста.	132
Л. А. Старченко. Некоторые способы выражения в тексте «источника» информации . . . . .	140
Г. С. Ульянова. Построение алгоритма морфологического анализа для русского языка на основе конечноавтоматной модели.	145

**Лингвистические проблемы функционального моделирования речевой деятельности**

**Выпуск I**

Редактор **Л. А. Карлова**

Техн. редактор **Г. С. Орлова**

Корректоры **Е. К. Терентьева, С. К. Школьников**

---

М-13885. Сдано в набор 26-ХІІ 1972 г. Подписано к печати 28-ІІІ 1973 г.  
Бумага тип. № 3. Формат бум. 60×84<sup>1</sup>/<sub>16</sub>. Бум. л. 5,5. Печ. л. 11.  
Уч.-изд. л. 10,01. Тираж 1380 экз. Заказ № 3041. Цена 72 коп.

Издательство ЛГУ им. А. А. Жданова  
199164, Ленинград, Университетская наб., 7/9.

---

Тип. № 7 Ленуприздат