

**МАТЕМАТИЧЕСКИЕ  
ВОПРОСЫ  
КИБЕРНЕТИКИ**

**10**

**О. С. Кулагина**

**Об одном подходе к  
установлению  
отношений между  
простыми  
предложениями в  
составе сложного при  
автоматическом  
анализе текстов**

**Рекомендуемая форма библиографической ссылки:**  
Кулагина О. С. Об одном подходе к установлению отношений между простыми предложениями в составе сложного при автоматическом анализе текстов // Математические вопросы кибернетики. Вып. 10. – М.: ФИЗМАТЛИТ, 2001. – С. 15–34. URL: <http://library.keldysh.ru/mvk.asp?id=2001-15>

# **ОБ ОДНОМ ПОДХОДЕ К УСТАНОВЛЕНИЮ ОТНОШЕНИЙ МЕЖДУ ПРОСТЫМИ ПРЕДЛОЖЕНИЯМИ В СОСТАВЕ СЛОЖНОГО ПРИ АВТОМАТИЧЕСКОМ АНАЛИЗЕ ТЕКСТОВ\*)**

**О. С. КУЛАГИНА**

(МОСКВА)

## **§ 1. Введение**

При автоматическом анализе текстов на естественных языках может представлять интерес установление не только синтаксических отношений (или, что то же, синтаксических связей) между отдельными словами, но также отношений другого уровня, в первую очередь, между простыми предложениями в составе сложного.

В системе анализа русских текстов (системе АРТ) имеются некоторые правила установления отношений между простыми предложениями в составе сложного одновременно с установлением синтаксических связей между словами.

В данной статье рассматривается несколько иной подход. Эти два подхода не исключают, а дополняют друг друга. Дело в том, что естественный язык настолько сложен, что никакая система точных формальных правил не может охватить все возможные ситуации; для любой системы таких правил можно построить противоречащий пример. Поэтому представляется полезным рассмотрение одних и тех же ситуаций под разными углами зрения. Это иногда может дать лучший результат, чем неограниченное уточнение одной системы правил. Поясняющей, хотя и поверхностной аналогией может служить известный факт, что местоположение удаленного объекта удается проще и точнее установить, если засекают его из двух разных точек.

В разработке описываемого подхода, кроме автора, участвовали Т. Б. Агранат и Е. Л. Козьмина.

Следует оговорить, что описываемые в данной работе правила и словари, как и система АРТ, рассчитаны на тексты научно-технического характера, т. е. того, что иногда называется деловой прозой. Некоторые ситуации, которые допустимы по нормам языка, но практически не употребляются в деловой прозе и не охвачены правилами, оговорены при формулировке правил.

Вообще говоря, язык не лимитирует число простых предложений в составе сложного. Чтобы получить представление о положении дел в научных текстах, было составлено два массива по 500 предложений (в каждом 10 текстов по 50 предложений из математики, лингвистики и других научных

---

\*) Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект 99-01-01191).

областей). Понятно, что такое количество нельзя считать достаточным для точных оценок, но тот факт, что оценки, полученные по этим двум массивам, получились весьма близкими, дает основание думать, что полученные цифры могут служить в качестве ориентира. Ниже приводятся результаты подсчета (в процентах с точностью до одного процента) для этих двух массивов. Одно простое предложение без союзов и знаков препинания: 16% и 12%, а с союзами или знаками препинания 30% и 30%. Сложное предложение из двух простых: 32% и 35%; из трех простых: 14% и 14%; из четырех простых: 5% и 6%; из пяти и более простых: 3% и 3%.

Из приводимых цифр видно, что в обоих массивах наиболее частыми были простые предложения, причем такие, которые содержат причастные или деепричастные обороты или однородные члены. Некоторые правила, позволяющие установить, является ли предложение простым или сложным, разработаны Т. Б. Агранат и описаны в [1].

Около половины от общего числа предложений составили сложные предложения, состоящие из двух или трех простых, тогда как сложные предложения, содержащие четыре и более простых, дали лишь около одной десятой от общего числа. Естественно, что наше внимание было сосредоточено на сложных предложениях из двух или трех простых, хотя некоторые приводимые ниже правила членения сложных предложений на простые учитывают и ситуацию четырех простых.

Установление отношений между простыми предложениями в составе сложного базируется как на общих закономерностях, так и на свойствах отдельных слов. К числу общих закономерностей относятся возможные схемы сложных предложений. В данной работе рассматриваются два вида схем, описывающих строение сложных предложений из простых: линейные и структурные.

В *линейных схемах* (ЛС) указан линейный порядок простых предложений и их границ. Границы между простыми предложениями могут оформляться по-разному: знаками препинания без союзов, с сочинительными или подчинительными союзами. В *структурных схемах* (СТС) указаны отношения сочинения и подчинения, но без детализировки типов отношения (только направление). Одной линейной схеме может соответствовать несколько структурных. *Структурой* естественно называть структурную схему, в которой указаны типы отношений. Очевидно, что одной структурной схеме может соответствовать множество структур.

При переходе от линейной схемы к структурной и затем к структуре учитываются различные факторы. Прежде всего это расположение и вид союзов, союзных слов и соответствующих им соотносительных слов. Релевантным для рассматриваемой проблемы свойством конкретных слов является возможность слова иметь сентенциальные актанты или сирконстанты. Такие возможности (причем с градацией степени) зафиксированы в синтаксическом словаре, описанном в параграфе 5. На эти оценки опирается выбор наиболее вероятного из числа потенциальных управляющих. При этом учитывается сочетаемость типов отношений между простыми предложениями в составе сложного. Некоторые данные о такой сочетаемости вошли в базу, составленную Т. Б. Агранат (см. [3]).

Ниже приводится каталог линейных и структурных схем для сложных предложений из двух и трех простых, указываются соотношения этих схем. Проблема членения сложного предложения на простые, не опирающегося на синтаксические связи между словами рассмотрена в параграфе 4.

## § 2. Сложное предложение из двух простых

В приводимых схемах используются следующие обозначения:  $P$  — простое предложение;  $зп$  — знак препинания;  $сс$  — союз сочинительный;  $сп$  — союз подчинительный. Числовые индексы — их порядковые номера. Следует оговорить, что знаки препинания показаны в схемах только при бессоюзном сочленении простых предложений, а обязательные знаки препинания перед подчинительным или сочинительным союзом для упрощения схем не указываются. Иными словами, точнее было бы сказать, что  $сп$  обозначает пару: "знак препинания+подчинительный союз", аналогично для  $сс$ .

Простые предложения могут либо следовать друг за другом, либо вставляться одно внутрь другого. Вставка показывается в линейных схемах фигурными скобками, т. е. если  $P_j$  вставлено в  $P_i$ , то записывается  $P_i\{P_j\}P_i^*$ , где  $P_i^*$  — это продолжение  $P_i$ .

При этом предложение, которое в тексте стоит в скобках или в кавычках, можно всегда считать вставленным в предыдущее даже при отсутствии продолжения объемлющего.

В структурных схемах отношение " $P_i$  и  $P_j$  — сочиненные" обозначено  $P_i \& P_j$ , а отношение " $P_i$  подчиняет  $P_j$ " обозначено  $P_i > P_j$ .

Заметим, что сложное предложение может начинаться с союза, обычно сочинительного, который связывает его с предшествующим текстом, такой союз не влияет на схему данного предложения.

Приведем сначала обобщенные схемы, в которых указаны только  $P$ , а затем рассмотрим их варианты при разных видах сочленения. Следует оговорить, что вставка одного предложения в другое оформляется знаками препинания с обеих сторон вставляемого, но для упрощения схем ниже в явном виде указывается только левый, открывающий знак. Если левый знак тире, скобка или кавычки, то правый знак такой же, в остальных случаях правый знак обычно запятая. На типе отношений правый знак не указывается.

Для двух простых предложений возможны две обобщенные линейные схемы:

$P_1 P_2$  — следование;

$P_1\{P_2\}P_1^*$  — вставка.

и три обобщенные структурные схемы:

$P_1 > P_2$  —  $P_1$  подчиняет  $P_2$ ;

$P_1 \& P_2$  — сочиненные предложения;

$P_1 < P_2$  —  $P_2$  подчиняет  $P_1$ .

**2.1. Союз отсутствует.** Возможны две линейные схемы:

**2.1.1.**  $P_1 зп P_2$

**2.1.2.**  $P_1\{зп P_2\}P_1^*$

Как было сказано выше, при вставке, т. е. при ЛС 2.1.2, всегда  $P_1 > P_2$ .

При ЛС 2.1.1 СТС зависит от знака препинания. Если  $зп$  = точка с запятой, то  $P_1$  и  $P_2$  в сущности независимы, однако для связности структурной схемы удобно считать, что в этом случае имеет место  $P_1 > P_2$  при специальном типе подчинения. Если  $зп$  = двоеточие, то  $P_2$  поясняет  $P_1$ , т. е. СТС =  $P_1 > P_2$ . Если  $зп$  = запятая, то возможно как подчинение ( $P_1 > P_2$ ), так и сочинение ( $P_1 \& P_2$ ). Различение этих случаев требует привлечения семантики.

Если  $зп$  = тире, то также возможны две СТС:  $P_1 > P_2$  и  $P_1 \& P_2$ . Тире часто означает пояснение, по смыслу примерно равное «а именно». Оно может также обозначать соединение условия и следствия: например, «Лектор пришел — началась лекция.». В этих случаях СТС =  $P_1 > P_2$ . Тире может показывать противопоставление, заменяя сочинительный союз (в частности, «а»), например, «я дал Пете тетрадь — он мне не дал ничего». Сочета-

ние запятой и тире представляет собой чисто семантическое усиление разделения, его можно анализировать как тире. Например, «Что он думал, — не нам судить.»:  $СТС = П_1 > П_2$ .

Заметим, что тире является наиболее трудным для анализа знаком препинания ввиду множества его ролей. Кроме отмеченных случаев, оно может выступать в роли замены опущенного слова, в перечислениях, в качестве выделителя для именной группы или целого предложения, между датами (1975–1990), именами собственными (самолет Москва — Париж) и т. д. Некоторые правила работы с тире описаны в параграфе 4.

**2.2. Сочинительный союз.** Возможны три линейные схемы:

**2.2.1.**  $П_1 сс П_2$

**2.2.2.**  $сс_1 П_1 сс_2 П_2$

**2.2.3.**  $П_1 \{ сс П_2 \} П_1^*$

Вторая ЛС возникает при повторяющихся или парных союзах: «либо...либо», «ни...ни», «то...то», «не только...но и» и т. п.

В первых двух случаях  $СТС = П_1 \& П_2$ , а линейной схеме 2.2.3 соответствует  $СТС = П_1 > П_2$ . Следует оговорить, что в текстах рассматриваемого типа схема 2.2.3 маловероятна, она носит скорее разговорный характер. Например, «Техник, а он всегда опаздывал, опоздал и сегодня».

**2.3. Подчинительный союз.** Возможны три линейные схемы:

**2.3.1.**  $П_1 сп П_2$

**2.3.2.**  $сп П_1 зп П_2$

**2.3.3.**  $П_1 \{ сп П_2 \} П_1^*$

Схемам 2.3.1 и 2.3.3 соответствует  $СТС = П_1 > П_2$ . При ЛС 2.3.2 имеем  $СТС = П_1 < П_2$ , причем здесь наиболее часто употребляются союзы «если», «чтобы», «хотя».

**2.4. Подчинительный и сочинительный союзы.** Для двух предложений это сочетание маловероятно. При нем возможны две линейные схемы:

**2.4.1.**  $сп П_1 сс П_2$

**2.4.2.**  $П_1 сс сп П_2$

При ЛС 2.4.1. имеем  $СТС = П_1 < П_2$ , например, «хотя  $П_1$ , но  $П_2$ ». При ЛС 2.4.2. имеем  $СТС = П_1 > П_2$ , например, « $П_1$ , а поэтому  $П_2$ », здесь  $сс+сп$  эквивалентно одному  $сп$ .

Возможна еще последовательность  $П_1 сп П_2 сс П_1^*$ , но в ней  $сс$  связывает однородные члены внутри  $П_1$ , т. е. это случай ЛС 2.3.3.

Заметим, что наиболее частыми способами оформления вставных предложений являются следующие. При бессоюзном соединении это обычно простые скобки. При сочинительном союзе возможно вставное предложение типа вводного, см. пример, приведенный выше. Как мы уже отметили, для рассматриваемого вида текстов такая конструкция маловероятна. При подчинительном союзе чаще всего встречаются следующие сочетания. Определительное придаточное, вводимое союзом «который». Сентенциальный актант, вводимый союзом «что» («Опасение, что Иванов опоздает, оправдалось»). Часто в научно-технических текстах встречаются предложения-ссылки, вводимые союзом «как». Например, «как было показано выше» — это ссылка по тексту назад, «как будет видно из дальнейшего» — ссылка по тексту вперед; «как известно» — ссылка за пределы текста.

### § 3. Сложное предложение из трех простых

В структурных схемах для фразы из трех простых, имеющей вид  $П_1 П_2 П_3$ , кроме обозначений отношений, введенных выше, используются еще следующие.

Если после  $\Pi_1$  стоят два знака, то первый указывает отношение с  $\Pi_2$ , второй — с  $\Pi_3$ . Так запись  $\Pi_1 >>$  показывает, что  $\Pi_1$  подчиняет и  $\Pi_2$ , и  $\Pi_3$ . Аналогично из двух знаков перед  $\Pi_3$  первый указывает отношение с  $\Pi_1$ , второй — с  $\Pi_2$ .

Восклицательный знак после знака  $>$  или  $\&$  указывает, что данное отношение связывает не соседние  $\Pi$ , а  $\Pi_1$  с  $\Pi_3$ . Например, запись  $\Pi_1 >! \Pi_3$  означает, что  $\Pi_1$  подчиняет  $\Pi_3$ , но не  $\Pi_2$ ; а запись  $<! \Pi_3$  показывает, что  $\Pi_3$  подчиняет  $\Pi_1$ , но не  $\Pi_2$ . Заметим, что поскольку сочинительное отношение симметрично, записи  $\Pi_1 \&! \Pi_3$  и  $\&! \Pi_3 \Pi_1$  равносильны и означают, что  $\Pi_1$  и  $\Pi_3$  — сочиненные. Например,  $\Pi_1 > \&! \Pi_2 \Pi_3$  означает, что  $\Pi_1$  подчиняет  $\Pi_2$  и сочинено с  $\Pi_3$ .

Знаком  $+$  показывается, что два сочиненных предложения в отношениях с третьим предложением выступают как единое целое. Например, запись  $\Pi_1 > \Pi_2 + \Pi_3$  означает, что  $\Pi_1$  подчиняет пару.

Знак  $\$$  показывает, что следующее за ним предложение связано сочинительной связью со сложно-подчиненным предложением, образованным двумя другими. Возможны структурные схемы  $\Pi_1 > \Pi_2 \$ \Pi_3$  и  $\Pi_1 < \Pi_2 \$ \Pi_3$ , а при другом порядке  $\Pi$  подобная ситуация маловероятна.

Для трех простых возможны шесть обобщенных линейных схем:

$$\begin{aligned} &\Pi_1 \Pi_2 \Pi_3 \\ &\Pi_1 \{ \Pi_2 \} \Pi_1^* \Pi_3 \\ &\Pi_1 \Pi_2 \{ \Pi_3 \} \Pi_2^* \\ &\Pi_1 \{ \Pi_2 \Pi_3 \} \Pi_1^* \\ &\Pi_1 \{ \Pi_2 \{ \Pi_3 \} \Pi_2^* \} \Pi_1^* \\ &\Pi_1 \{ \Pi_2 \} \Pi_1^* \{ \Pi_3 \} \Pi_1^* \end{aligned}$$

Для них возможны двадцать обобщенных структурных схем:

$\Pi_1 > \Pi_2 > \Pi_3$	$\Pi_1 > \Pi_2 + \Pi_3$	$\Pi_1 > \&! \Pi_2 \Pi_3$
$\Pi_1 >> \Pi_2 \Pi_3$	$\Pi_1 < \Pi_2 + \Pi_3$	$\Pi_1 \&! \Pi_2 < \Pi_3$
$\Pi_1 >! \Pi_2 < \Pi_3$	$\Pi_1 \& \Pi_2 + \Pi_3$	$\Pi_1 < \Pi_2 \& \Pi_3$
$\Pi_1 < \Pi_2 > \Pi_3$	$\Pi_1 + \Pi_2 > \Pi_3$	$\Pi_1 \& \Pi_2 > \Pi_3$
$\Pi_1 < \Pi_2 < \Pi_3$	$\Pi_1 + \Pi_2 < \Pi_3$	$\Pi_1 > \Pi_2 \$ \Pi_3$
$\Pi_1 \Pi_2 << \Pi_3$	$\Pi_1 + \Pi_2 \& \Pi_3$	$\Pi_1 < \Pi_2 \$ \Pi_3$
$\Pi_1 > \Pi_2 <! \Pi_3$		$\Pi_1 \& \Pi_2 \& \Pi_3$

Непроективные схемы:  $\Pi_1 >! < \Pi_2 \Pi_3$  и  $\Pi_1 \Pi_2 > <! \Pi_3$  из рассмотрения исключены

**3.1. Союзы отсутствуют.** Возможны шесть линейных схем:

- 3.1.1.  $\Pi_1 \text{зп}_1 \Pi_2 \text{зп}_2 \Pi_3$
- 3.1.2.  $\Pi_1 \{ \text{зп}_1 \Pi_2 \} \Pi_1^* \text{зп}_2 \Pi_3$
- 3.1.3.  $\Pi_1 \text{зп}_1 \Pi_2 \{ \text{зп}_2 \Pi_3 \} \Pi_2^*$
- 3.1.4.  $\Pi_1 \{ \text{зп}_1 \Pi_2 \text{зп}_2 \Pi_3 \} \Pi_1^*$
- 3.1.5.  $\Pi_1 \{ \text{зп}_1 \Pi_2 \{ \text{зп}_2 \Pi_3 \} \Pi_2^* \} \Pi_1^*$
- 3.1.6.  $\Pi_1 \{ \text{зп}_1 \Pi_2 \} \Pi_1^* \{ \text{зп}_2 \Pi_3 \} \Pi_1^*$

При вставке могут использоваться запятая, тире, скобки и кавычки, причем, как уже говорилось, при скобках и кавычках вставление имеет место даже при отсутствии  $\Pi_1^*$  или  $\Pi_2^*$ . Как и в случае двух простых, естественно считать, что вставное предложение подчинено объемлющему. При следовании, как и в случае двух простых, удобно считать, что отношения направлены слева направо, причем тип отношения определяется знаком препинания.

Поскольку вставка диктует подчинение, схеме ЛС 3.1.5 может соответствовать только  $\text{СТС} = \Pi_1 > \Pi_2 > \Pi_3$ ; а схеме ЛС 3.1.6 — только

СТС =  $P_1 \gg P_2 P_3$ . Для ЛС 3.1.4 возможны две СТС: СТС =  $P_1 > P_2 > P_3$  и СТС =  $P_1 \gg P_2 P_3$ . Для ЛС 3.1.2 возможны те СТС, в которых  $P_1 > P_2$ , а ЛС для 3.1.3 — те СТС, где  $P_2 > P_3$ . Для ЛС 3.1.1 теоретически возможны все СТС, однако для нее, как и для 3.1.2 и 3.1.3, наиболее реально последовательное подчинение: СТС =  $P_1 > P_2 > P_3$ .

Мы не будем больше на этом останавливаться, поскольку два бессоюзных сочленения — это случай достаточно редкий.

**3.2. Одно сочленение без союза и сочинительные союзы.** При одном сочинительном союзе теоретически каждая обобщенная линейная схема может дать две линейные с указанием положения союза, т. е. всего 12 линейных схем:

3.2.1.  $P_1 \text{зп } P_2 \text{сс } P_3$

3.2.2.  $P_1 \text{сс } P_2 \text{зп } P_3$

3.2.3.  $P_1 \{ \text{зп } P_2 \} P_1^* \text{сс } P_3$

3.2.4.  $P_1 \text{сс } P_2 \{ \text{зп } P_3 \} P_2^*$

3.2.5.  $P_1 \{ \text{зп } P_2 \text{сс } P_3 \} P_1^*$

3.2.6.  $P_1 \{ \text{сс } P_2 \} P_1^* \text{зп } P_3$

3.2.7.  $P_1 \text{зп } P_2 \{ \text{сс } P_3 \} P_2^*$

3.2.8.  $P_1 \{ \text{сс } P_2 \text{зп } P_3 \} P_1^*$

3.2.9.  $P_1 \{ \text{зп } P_2 \{ \text{сс } P_3 \} P_2^* \} P_1^*$

3.2.10.  $P_1 \{ \text{сс } P_2 \{ \text{зп } P_3 \} P_2^* \} P_1^*$

3.2.11.  $P_1 \{ \text{зп } P_2 \} P_1^* \{ \text{сс } P_3 \} P_1^*$

3.2.12.  $P_1 \{ \text{сс } P_2 \} P_1^* \{ \text{зп } P_3 \} P_1^*$

Однако, как уже было отмечено, вставка, оформленная сочинительным союзом, для рассматриваемого типа текстов нехарактерна, поэтому схемы 3.2.6–3.2.12 не разбираются, т. е. мы считаем реальными первые пять схем.

В приводимых ниже табл. 1–5 в правом столбце показано, какие СТС могут соответствовать линейной схеме, стоящей в левом столбце, при тех или иных зп. Косая черта разделяет альтернативные варианты СТС.

Таблица 1

3.2.1. $P_1 \text{зп } P_2 \text{сс } P_3$	$P_1 > P_2 + P_3$ (зп=точка с запятой или двоеточие)
	$P_1 > P_2 + P_3 / P_1 > \&!P_2 P_3$ (зп=тире)
	$P_1 \& P_2 \& P_3 / P_1 > P_2 + P_3 / P_1 > \&!P_2 P_3$ (зп=запятая)
3.2.2. $P_1 \text{сс } P_2 \text{зп } P_3$	$P_1 \& P_2 > P_3 / P_1 \& P_2 \& P_3$ (зп=запятая)
	$P_1 + P_2 > P_3$ (зп=точка с запятой)
	$P_1 \& P_2 > P_3 / P_1 + P_2 > P_3$ (зп=тире или двоеточие)
3.2.3. $P_1 \{ \text{зп } P_2 \} P_1^* \text{сс } P_3$	$P_1 > \&!P_2 P_3$
3.2.4. $P_1 \text{сс } P_2 \{ \text{зп } P_3 \} P_2^*$	$P_1 \& P_2 > P_3$
3.2.5. $P_1 \{ \text{зп } P_2 \text{сс } P_3 \} P_1^*$	$P_1 > P_2 + P_3$

Заметим, что при ЛС 3.2.1 и зп=запятая наиболее естественна первая из указанных СТС, т. е. сочинены все три П.

В ЛС 3.2.2 при зп=запятая сочиненность всех трех П маловероятна, тут естественнее первая из указанных СТС. Если зп=точка с запятой, то  $P_3$  независимо, и, как уже оговаривалось, подчиняется из соображений связ-

ности. Если зп=тире или двоеточие, то  $P_3$  может пояснять как одно  $P_2$ , так и пару  $P_1 + P_2$ , т. е. возможны две СТС.

При трех предложениях и одном бессоюзном сочленении еще возможна ситуация двух сочинительных союзов — парных или повторяющихся. Для простоты в обоих случаях будем называть их парой союзов и обозначать  $cc_1$  и  $cc_2$ . С учетом сказанного выше о вставке здесь возможны пять линейных схем, которые вместе с соответствующими им СТС приведены в табл. 2. К сожалению, тут нельзя указать формальные признаки для выбора из альтернативных СТС, он определяется семантикой.

Таблица 2

3.2.13. $P_1 зп cc_1 P_2 cc_2 P_3$	$P_1 > P_2 + P_3 / P_1 \& P_2 + P_3$
3.2.14. $cc_1 P_1 cc_2 P_2 зп P_3$	$P_1 \& P_2 > P_3 / P_1 + P_2 > P_3 / P_1 \& P_2 + P_3$
3.2.15. $cc_1 P_1 cc_2 P_2 \{зп P_3\} P_2^*$	$P_1 \& P_2 > P_3$
3.2.16. $cc_1 P_1 зп P_2 cc_2 P_3$	$P_1 > \& P_2 P_3 / P_1 + P_2 \& P_3$
3.2.17. $cc_1 P_1 \{зп P_2\} P_1^* cc_2 P_3$	$P_1 > \& P_2 P_3$

При ЛС 3.2.13 имеется в виду, что зп тут "значащий" (двоеточие, точка с запятой или тире), т. е. оформляющий данное сочленение, например: «Он знал: либо  $P_2$ , либо  $P_3$ ». Если же зп просто обязательный между самостоятельными предложениями, то имеет место ситуация 3.4.1. (см. ниже).

**3.3. Сочленение без союза и один подчинительный союз.** В этом случае возможны 15 линейных схем, которые вместе с соответствующими СТС приведены в табл. 3.

Как уже оговаривалось ранее, знак препинания перед подчинительным союзом не указан, хотя он обязателен. Напомним, что для связности СТС мы связываем условным подчинением слева направо предложения, которые, в сущности, независимы (что должен отражать тип подчинения).

Если при ЛС 3.3.11 зп = точка с запятой, то это значит, что союз в начале фразы связывает ее с предшествующим текстом, т. е. мы оказываемся в ситуации, разобранный в п. 3.1. Аналогичная ситуация возможна при ЛС 3.3.13.

Линейные схемы 3.3.14 и 3.3.15 возможны только если союз в начале предложения связывает его с предшествующим текстом, т. е. это ситуация, разобранный в п. 3.1.

**3.4. Сочинительные союзы.** В отличие от п. 3.2. здесь рассматривается ситуация, когда все сочленения оформлены сс. При двух сочинительных союзах возможна только одна линейная схема:

**3.4.1.**  $P_1 cc_1 P_2 cc_2 P_3$

Наиболее естественная структурная схема: СТС =  $P_1 \& P_2 \& P_3$ , т. е. три сочиненных предложения. Однако тут возможны  $P_1 + P_2 \& P_3$ ,  $P_1 \& P_2 + P_3$ , см. также 3.2.13.

При парных и повторяющихся союзах возможны еще схемы с тремя союзами:

**3.4.2.**  $cc_1 P_1 cc_2 P_2 cc_3 P_3$

**3.4.3.**  $P_1 cc_1 cc_2 P_2 cc_3 P_3$

При ЛС 3.4.2 СТС зависит от того, какие союзы образуют пару или повторяются. Если все три союза одинаковые, то имеем три сочиненных предложения, СТС =  $P_1 \& P_2 \& P_3$  (например, «Либо  $P_1$ , либо  $P_2$ , либо  $P_3$ »).

Таблица 3

3.3.1. $\Pi_1 \text{зп} \Pi_2 \text{сп} \Pi_3$	$\Pi_1 > \Pi_2 > \Pi_3 / \Pi_1 \ \& \ \Pi_2 > \Pi_3 / \Pi_1 + \Pi_2 > \Pi_3$ (зп=запятая)
	$\Pi_1 > \Pi_2 > \Pi_3$ (остальные зп)
3.3.2. $\Pi_1 \text{сп} \Pi_2 \text{зп} \Pi_3$	$\Pi_1 >> \Pi_2 \Pi_3$ (зп=точка с запятой)
	$\Pi_1 > \Pi_2 > \Pi_3 / \Pi_1 >> \Pi_2 \Pi_3$ (зп=двоеточие, тире, запятая)
	$\Pi_1 >> \Pi_2 \Pi_3 / \Pi_1 > \Pi_2 + \Pi_3 / \Pi_1 > \Pi_2 < \Pi_3$ (зп=запятая)
3.3.3. $\Pi_1 \text{зп} \Pi_2 \{\text{сп} \Pi_3\} \Pi_2^*$	$\Pi_1 > \Pi_2 > \Pi_3 / \Pi_1 \ \& \ \Pi_2 > \Pi_3$ (зп=запятая)
	$\Pi_1 > \Pi_2 > \Pi_3$ (остальные зп)
3.3.4. $\Pi_1 \text{сп} \Pi_2 \{\text{зп} \Pi_3\} \Pi_2^*$	$\Pi_1 > \Pi_2 > \Pi_3$
3.3.5. $\Pi_1 \{\text{зп} \Pi_2\} \Pi_1^* \text{сп} \Pi_3$	$\Pi_1 >> \Pi_2 \Pi_3$
3.3.6. $\Pi_1 \{\text{сп} \Pi_2\} \Pi_1^* \text{зп} \Pi_3$	$\Pi_1 >> \Pi_2 \Pi_3 / \Pi_1 > \& \Pi_2 \Pi_3$ (зп=запятая)
	$\Pi_1 >> \Pi_2 \Pi_3$ (остальные зп)
3.3.7. $\Pi_1 \{\text{зп} \Pi_2 \text{сп} \Pi_3\} \Pi_1^*$	$\Pi_1 > \Pi_2 > \Pi_3$
3.3.8. $\Pi_1 \{\text{сп} \Pi_2 \text{зп} \Pi_3\} \Pi_1^*$	$\Pi_1 > \Pi_2 > \Pi_3 / \Pi_1 >> \Pi_2 \Pi_3 / \Pi_1 > \Pi_2 + \Pi_3$
3.3.9. $\Pi_1 \{\text{зп} \Pi_2 \{\text{сп} \Pi_3\} \Pi_2^*\} \Pi_1^*$	$\Pi_1 > \Pi_2 > \Pi_3$
3.3.10. $\Pi_1 \{\text{сп} \Pi_2 \{\text{зп} \Pi_3\} \Pi_2^*\} \Pi_1^*$	$\Pi_1 > \Pi_2 > \Pi_3$
3.3.11. $\text{сп} \Pi_1 \text{зп}_1 \Pi_2 \text{зп}_2 \Pi_3$	$\Pi_1 > \Pi_2 < \Pi_3$ (зп <sub>1</sub> =двоеточие, тире)
	$\Pi_1 + \Pi_2 < \Pi_3 / \Pi_1 > \Pi_2 < \Pi_3$ (зп <sub>1</sub> =запятая)
	$\Pi_1 < \Pi_2 > \Pi_3$ (зп <sub>1</sub> =запятая, зп <sub>2</sub> =запятая, двоеточие или точка с запятой)
3.3.12. $\text{сп} \Pi_1 \{\text{зп}_1 \Pi_2\} \Pi_1^* \text{зп}_2 \Pi_3$	$\Pi_1 > \Pi_2 < \Pi_3$
3.3.13. $\text{сп} \Pi_1 \text{зп}_1 \Pi_2 \{\text{зп}_2 \Pi_3\} \Pi_2^*$	$\Pi_1 < \Pi_2 > \Pi_3$
3.3.14. $\text{сп} \Pi_1 \{\text{зп}_1 \Pi_2 \text{зп}_2 \Pi_3\} \Pi_1^*$	
3.3.15. $\text{сп} \Pi_1 \{\text{зп}_1 \Pi_2 \{\text{зп}_2 \Pi_3\} \Pi_2^*\} \Pi_1^*$	

Если парой являются  $сс_1$  и  $сс_2$  (например: «Либо  $\Pi_1$ , либо  $\Pi_2$ , и  $\Pi_3$ »), то  $СТС = \Pi_1 + \Pi_2 \ \& \ \Pi_3$  или  $СТС = \Pi_1 \ \& \ \Pi_2 + \Pi_3$ , а если пара — это  $сс_1$  и  $сс_3$  (например: «Либо  $\Pi_1$ , а  $\Pi_2$ , либо  $\Pi_3$ »), то  $СТС = \Pi_1 + \Pi_2 \ \& \ \Pi_3$ . Если же пара — это  $сс_2$  и  $сс_3$ , то  $сс_1$  связывает всю фразу с предшествующим текстом, т. е. это случай 3.4.1.

В ЛС 3.4.3 парой могут быть только  $сс_2$  и  $сс_3$  (например, « $\Pi_1$ , но то  $\Pi_2$ , то  $\Pi_3$ »), поэтому  $СТС = \Pi_1 \ \& \ \Pi_2 + \Pi_3$ .

**3.5. Сочинительные и подчинительные союзы.** Это сочетание представлено в табл. 4. Если имеется один подчинительный союз и один сочинительный (неповторяющийся) союз, то возможны 7 линейных схем: 3.5.1–3.5.7. Как уже говорилось, вставка, начинающаяся с сочинительного союза, не рассматривается. При парных или повторяющихся сочинительных союзах возможны линейные схемы 3.5.8–3.5.15. При двух подчинительных союзах и одном сочинительном возможны линейные схемы 3.5.16–3.5.18.

Сочетание двух сочинительных союзов (непарных и неповторяющихся) и одного подчинительного (ЛС 3.5.19) возможно не при любых сп, сс<sub>2</sub> здесь является как бы усилителем, но отношение определяет подчинительный союз. Например, «П<sub>1</sub> и П<sub>2</sub>, а поэтому П<sub>3</sub>».

Таблица 4

3.5.1. П <sub>1</sub> сс П <sub>2</sub> сп П <sub>3</sub>	$P_1 \& P_2 > P_3 / P_1 + P_2 > P_3$
3.5.2. П <sub>1</sub> сп П <sub>2</sub> сс П <sub>3</sub>	$P_1 > P_2 + P_3 / P_1 > \& ! P_2 P_3 / P_1 > P_2 \$ P_3$
3.5.3. сп П <sub>1</sub> зп П <sub>2</sub> сс П <sub>3</sub>	$P_1 < P_2 \& P_3 / P_1 < P_2 + P_3 / P_1 < P_2 \$ P_3$
3.5.4. сп П <sub>1</sub> сс П <sub>2</sub> зп П <sub>3</sub>	$P_1 + P_2 < P_3$
3.5.5. П <sub>1</sub> {сп П <sub>2</sub> }П <sub>1</sub> *сс П <sub>3</sub>	$P_1 > \& ! P_2 P_3$
3.5.6. П <sub>1</sub> сс П <sub>2</sub> {сп П <sub>3</sub> }П <sub>2</sub> *	$P_1 \& P_2 > P_3$
3.5.7. П <sub>1</sub> {сп П <sub>2</sub> сс П <sub>3</sub> }П <sub>1</sub> *	$P_1 > P_2 + P_3$
3.5.8. сс <sub>1</sub> П <sub>1</sub> сс <sub>2</sub> П <sub>2</sub> сп П <sub>3</sub>	$P_1 \& P_2 > P_3 / P_1 + P_2 > P_3$
3.5.9. сс <sub>1</sub> П <sub>1</sub> сс <sub>2</sub> П <sub>2</sub> {сп П <sub>3</sub> }П <sub>2</sub> *	$P_1 \& P_2 > P_3$
3.5.10. сс <sub>1</sub> П <sub>1</sub> сп П <sub>2</sub> сс <sub>2</sub> П <sub>3</sub>	$P_1 > P_2 \& ! P_3$
3.5.11. сс <sub>1</sub> П <sub>1</sub> {сп П <sub>2</sub> }П <sub>1</sub> *сс <sub>2</sub> П <sub>3</sub>	$P_1 > P_2 \& ! P_3$
3.5.12. П <sub>1</sub> сп сс <sub>1</sub> П <sub>2</sub> сс <sub>2</sub> П <sub>3</sub>	$P_1 > P_2 + P_3$
3.5.13. П <sub>1</sub> {сп сс <sub>1</sub> П <sub>2</sub> сс <sub>2</sub> П <sub>3</sub> }П <sub>1</sub> *	$P_1 > P_2 + P_3$
3.5.14. сп П <sub>1</sub> сс <sub>1</sub> П <sub>2</sub> сс <sub>2</sub> П <sub>3</sub>	$P_1 < P_2 + P_3$
3.5.15. сп сс <sub>1</sub> П <sub>1</sub> сс <sub>2</sub> П <sub>2</sub> П <sub>3</sub>	$P_1 + P_2 < P_3$
3.5.16. П <sub>1</sub> сп <sub>1</sub> П <sub>2</sub> сс сп <sub>2</sub> П <sub>3</sub>	$P_1 \gg P_2 P_3$
3.5.17. П <sub>1</sub> {сп <sub>1</sub> П <sub>2</sub> сс сп <sub>2</sub> П <sub>3</sub> }П <sub>1</sub> *	$P_1 \gg P_2 P_3$
3.5.18. сп <sub>1</sub> П <sub>1</sub> сс сп <sub>2</sub> П <sub>2</sub> П <sub>3</sub>	$P_1 P_2 \ll P_3$
3.5.19. П <sub>1</sub> сс <sub>1</sub> П <sub>2</sub> сс <sub>2</sub> сп П <sub>3</sub>	$P_1 + P_2 > P_3$

**3.6. Два подчинительных союза.** Возможные линейные и соответствующие им структурные схемы приведены в табл. 5.

Заметим, что при ЛС 3.6.1 и ЛС 3.6.6 та СТС, в которой первому предложению подчинены два других, менее вероятна, чем другая СТС.

3.6.1. $P_1 c_{P_1} P_2 c_{P_2} P_3$	$P_1 > P_2 > P_3 / P_1 \gg P_2 P_3$
3.6.2. $c_{P_1} P_1 z_{P_1} P_2 c_{P_2} P_3$	$P_1 < P_2 > P_3$
3.6.3. $c_{P_1} P_1 c_{P_2} P_2 z_{P_2} P_3$	$P_1 < P_2 < P_3 / P_1 P_2 \ll P_3 / P_1 > P_2 < ! P_3$
3.6.4. $P_1 \{c_{P_1} P_2\} P_1^* c_{P_2} P_3$	$P_1 \gg P_2 P_3$
3.6.5. $P_1 c_{P_1} P_2 \{c_{P_2} P_3\} P_2^*$	$P_1 > P_2 > P_3$
3.6.6. $P_1 \{c_{P_1} P_2 c_{P_2} P_3\} P_1^*$	$P_1 > P_2 > P_3 / P_1 \gg P_2 P_3$
3.6.7. $P_1 \{c_{P_1} P_2 \{c_{P_2} P_3\} P_2^*\} P_1^*$	$P_1 > P_2 > P_3$
3.6.8. $c_{P_1} P_1 \{c_{P_2} P_2\} P_1^* z_{P_1} P_3$	$P_1 > P_2 < ! P_3$
3.6.9. $c_{P_1} P_1 z_{P_1} P_2 \{c_{P_2} P_3\} P_2^*$	$P_1 < P_2 > P_3$

В рассмотренных схемах, как и в случае двух простых, говоря о числе союзов, мы считаем, что союз в самом начале связывает первое простое предложение с другими внутри сложного. Другая возможность — считать, что этот союз связывает сложное предложение в целом с предшествующим текстом. Тогда, кроме указанных, применимы схемы, в которых на один союз меньше (т. е. из ситуации п. 3.6 мы попадаем в ситуацию, разобранный в п. 3.3, из 3.5 в 3.2, из 3.3 в 3.1).

#### § 4. Фрагментирование

Ниже мы называем фразой анализируемую часть текста "от точки до точки", которая может быть простым или сложным предложением, а слово "предложение" употребляется в основном в смысле "простое предложение".

Цель фрагментирования — разделение исходной фразы, если она является сложным предложением, на простые предложения или на такие, по возможности крупные, независимые части, что они связаны друг с другом в синтаксической структуре фразы только через их вершины. *Вершиной* цепочки слов мы, как обычно, называем элемент этой цепочки, от которого непосредственно или опосредствованно зависят остальные элементы цепочки. Иными словами, правила фрагментирования устанавливают линейную структуру фразы.

Наличие простого предложения определяется, вообще говоря, наличием предикативной пары: "подлежащее+сказуемое". Вершиной в этой паре (и во всем предложении) мы считаем сказуемое. К сожалению, любая из компонент предикативной пары может отсутствовать по разным причинам: эллипсис, отсутствие сказуемого в назывных предложениях, а подлежащего — в безличных и др. С другой стороны, возможен и избыток этих компонент при наличии однородных членов. Для текстов рассматриваемого вида отсутствие глагола характерно для заголовков разделов, однако они, как правило, не слишком длинны и не нуждаются в фрагментировании. Что же касается остального текста (не заголовков), то в подавляющем большинстве случаев сказуемое присутствует, поэтому приводимые ниже правила опираются на наличие и число вершин.

Опорными точками для фрагментирования являются возможные границы простых предложений (знаки препинания и союзы) и возможные вершины (глагол, предикативное наречие, краткое прилагательное или причастие

и др.). Следует заметить, что русский язык выгодно отличается от многих других, например, от английского, во-первых, тем, что в русском четкая пунктуация, и, во-вторых, тем, что в нем глагольные формы редко омонимичны. При этом омонимия глагольных форм, редкая в языке в целом, еще сокращается за счет выбранного типа текстов. Так, например, в научно-технических текстах всегда можно считать, что словоформа «были» — это глагол, а не множественное число существительного «быль», напротив, словоформа «полю». может встретиться в математических текстах как дательный падеж от «поле», а не как форма глагола «полоть».

Оговорим сразу, что разработанные правила фрагментирования не могут гарантировать правильный результат при любых исходных данных. Как уже отмечалось выше, естественный язык настолько сложен, что он позволяет для любой системы правил построить противоречащий пример. С учетом этого в ряде случаев отмечены те ситуации, которые сочтены настолько маловероятными, что ими можно пренебречь. В основном имеющиеся правила рассчитаны на сложные предложения из двух или трех простых, причем такие, которые построены "естественным" образом. В некоторых ситуациях выдается более одного варианта фрагментирования фразы.

При фрагментировании для каждой выявленной вершины предложения надо установить левую и правую границы этого предложения. Заметим, что левые границы обычно указаны явно наличием не только знаков препинания, но и союзов. Гораздо труднее бывает установить правую границу предложения. Понятно, что чем меньше возможных границ, тем меньше вариантов фрагментирования фразы.

**4.1. Правила корректирования.** Работа правил фрагментирования начинается с этапа предварительной корректировки анализируемой фразы, цель которого сократить, по возможности, число опорных точек.

Мы называем *сильными* границами те границы, которые заведомо делят фразу на крупные независимые части. Независимые в том смысле, что только вершина может быть связана со словами вне данной части. Такие части представляют собой либо предложения, либо развернутые именные группы.

К сильным границам изначально относятся скобки, кавычки и точка с запятой. Даже если точки с запятой или правые скобки без соответствующих левых употреблены для разделения пунктов перечисления, разделяемые ими части фразы независимы в указанном выше смысле. В процессе фрагментирования к сильным могут быть причислены и другие границы.

Будем называть *разделами* те части, на которые исходную фразу делят сильные границы, либо саму исходную фразу, если сильных границ в ней нет.

Сокращение числа учитываемых границ регулируется следующими правилами. Считается, что не может быть границей простых предложений запятая, если это первая граница в разделе и либо ей предшествует вводное слово, вводный оборот (задаются списком) или деепричастный оборот, либо за ней следует причастный оборот. Последнее означает, что, снимая такую границу, мы присоединяем причастный оборот к предшествующему тексту, считая маловероятным, что второе простое предложение фразы начинается с препозитивного причастного оборота. Из числа границ исключается также последняя граница раздела, если за ней следует причастный оборот, а ближайшему влево подчинительному союзу не предшествует именная группа. Это последнее правило может дать неверный результат при двойной вставке, после которой продолжается объемлющее предложение, что тоже считается маловероятным.

В корректирующей части обрабатываются также сочетания союзов. Сочетание сочинительного и подчинительного союзов (например, «и что») считается одним подчинительным союзом, причем учитывается, что такому со-

четанию может не предшествовать запятая. Сочетание «как то:» считается единым оборотом, примерно означающими «а именно» и т. п.

Рассматриваются также повторяющиеся союзы: «или», «либо», «ни». Если между двумя такими союзами нет возможной вершины предложения, то знаки препинания, предшествующие не первому из них, убираются из числа возможных границ простых предложений. Аналогично — для парных союзов: «не только..., но и», «как..., так и» и т. п.

В этой же корректирующей части происходит предварительный анализ тире, которое может быть знаком препинания, а может быть заменителем глагола, т. е. вершиной простого предложения. Так, если раздел имеет вид "ИГим — ИГим", где ИГим — это именная группа в именительном падеже, то такое тире — глагол. Напротив, если тире стоит сразу после глагольной группы или в разделе присутствуют два тире, между которыми стоит именная группа, то такое тире — это знак препинания. Тире, не разобранные на этом этапе, остаются как возможная вершина, что учитывается в правилах фрагментирования.

Не считаются возможными вершинами простых предложений предикативные наречия, краткие прилагательные и причастия, если они не отделены никакими границами от глагола в личной форме. Например, в предложении «Задача решена.» вершина «решена», а в предложении «Задача была решена.» вершина «была».

Выявляются также явно однородные глаголы: два глагола, между которыми есть сочинительный союз и нет других границ и союзов. Такая пара глаголов образует одну возможную вершину простого предложения. Аналогично для предикативных наречий, кратких прилагательных и причастий.

Причисляется к числу вершин глагол в неопределенной форме, которому предшествует союз «чтобы» или сочетания с ним («для того чтобы», «с тем чтобы». и т. п.).

**4.2. Правила фрагментирования.** Мы не будем приводить все имеющиеся правила, но постараемся дать достаточно полное представление о том, на каких основаниях проводится фрагментирование анализируемой фразы.

Работа правил фрагментирования начинается с рассмотрения знаков препинания (называемых ниже для краткости границами) и определения для них некоторых числовых оценок, называемых ниже весом границ. Иными словами, в этом случае мы действуем в рамках того подхода с ориентацией на правила предпочтения, а не правила запрещения, который был развит ранее и изложен в ряде публикаций (см. [5–9]).

Выше было сказано, что точка с запятой считается сильной границей: она получает максимальный вес, равный девяти, скобки и кавычки получают вес восемь.

Вводится понятие *левая граница*. Это граница, открывающая придаточное предложение и выраженная знаком препинания перед подчинительным союзом или сочетанием сочинительного и подчинительного союзов. Левые границы получают вес пять, а если левая граница является последней границей в разделе, то она считается сильной, т. е. получает вес девять.

Назовем *сегментом* часть фразы между двумя границами, не содержащую границ, а *В-сегментом* такой сегмент, в котором есть явная вершина. Двоеточие перед В-сегментом считается левой границей. Сочетание "запятая + тире" перед В-сегментом получает вес четыре. Запятые, ограничивающие вводный, деепричастный или причастный оборот, если они не получили более высокой оценки, получают минимальный вес, равный единице. Остальные границы получают вес два.

Если во фразе есть скобки или кавычки, то отдельно анализируется то, что внутри них, и то, что вне их, по одним и тем же правилам, рассматриваемым ниже. Внутри скобок или кавычек могут оказаться: словосочетание,

не являющееся предложением, одно или несколько простых предложений. Теоретически, в скобках может быть любое число простых предложений, но в пределах рассматриваемой ситуации можно считать, что их не более двух.

Как было сказано выше, сильные границы делят фразу на части (названные разделами), которые анализируются независимо друг от друга.

При отсутствии вершин в разделе мы имеем дело со словосочетанием, не являющимся предложением. Например, это именная группа в скобках.

Если в разделе имеется одна вершина, то он считается одним простым предложением.

Очевидно, что если исходная фраза содержит две вершины, и либо между ними есть сильная граница, либо во фразе всего одна граница, то исходная фраза делится этой границей на два последовательных простых предложения, т. е. задача фрагментирования решена. Аналогично, если в исходной фразе было три вершины и две сильные границы между ними или всего две границы, то разбиение фразы на три последовательных простых предложения однозначно.

В остальных случаях сильные границы делят исходную фразу на разделы, которые требуют дальнейшей обработки.

Заметим, что если в исходной фразе было четыре вершины и хотя бы одна сильная граница, то такая фраза может быть обработана до конца имеющимися правилами, поскольку разделится на части, содержащие не более трех простых предложений. Аналогично при большем числе вершин и сильных границ.

После деления фразы по сильным границам рассматриваются левые границы. Ниже левые границы обозначены  $L$ , остальные границы  $G$ , а числовые индексы при них обозначают порядковые номера.

Пусть в разделе две вершины (обозначим содержащие их  $B$ -сегменты через  $V_1$  и  $V_2$ ) и пусть между ними есть левая граница ( $L$ ). Если после  $V_2$  в разделе нет границ, то указанная левая граница делит раздел на два последовательных простых предложения:  $\Pi_1 L \Pi_2$ . Если после  $V_2$  есть границы, то  $\Pi_1$  и  $\Pi_2$  могут быть последовательными, а возможна вставка  $\Pi_2$  в  $\Pi_1$ . Встает вопрос об определении правой границы  $\Pi_2$ . Если после  $V_2$  все границы имеют вес, равный единице, то считаем, что вставки нет. Если после  $V_2$  есть одна граница с весом больше единицы (обозначим ее  $Gp$ ), то выдается два варианта фрагментирования:  $\Pi_1 L \Pi_2$  и  $\Pi_1 \{L \Pi_2 Gp\} \Pi_2^*$ . Если после  $V_2$  есть несколько границ с весом больше единицы, то выдается фрагментирование  $\Pi_1 L \Pi_2$  как наиболее вероятное и указание, что в этом случае много вариантов.

Если в разделе два  $B$ -сегмента и нет левой границы, то мы имеем дело со сложносочиненным предложением. В этом случае труднее установить левую границу второго простого предложения, но зато маловероятна вставка, т. е. можно считать, что правая граница второго предложения совпадает с концом раздела. Если между двумя  $B$ -сегментами есть одна граница с весом больше единицы, то она считается границей этих двух простых предложений. Если таких границ две, то выдается два варианта, если же их больше двух, то в качестве наиболее вероятной берется та граница, которая ближе к второму  $B$ -сегменту, и отмечается, что возможны другие варианты.

Разберем еще ситуацию трех  $B$ -сегментов и двух левых границ. Это сочетание соответствует рассмотренному выше в п. 3.6 с тем отличием, что там выписаны схемы в предположении, что деление на простые предложения известно, а при фрагментировании нам известно расположение вершин и левых границ.

Отобразим строение анализируемого раздела с помощью схемы, которую для отличия от схем, введенных в параграфе 3, назовем  $B$ -схемой (сокращенно  $BC$ ).  $B$ -схема — это образ фразы в виде цепочки, в которой  $B$ -сегменты обозначены  $V$  с порядковыми индексами, левые границы обо-

значены  $L$  с порядковыми индексами, а  $T$  и  $N$  с порядковыми индексами обозначают последовательности элементов раздела, не содержащие  $V$  и  $L$ . Различие  $N$  и  $T$  в следующем: цепочки, обозначенные  $N$ , нейтральны в том смысле, что их состав не влияет на результат фрагментирования, в то время как строение цепочек, обозначенных  $T$ , влияет на результат, и, соответственно, ниже указаны условия на цепочки  $T$ , помогающие решению вопроса.

Для фразы с тремя  $B$ -сегментами и двумя левыми границами возможны  $B$ -схемы, приводимые ниже. Отметим сразу, что для трех последних  $B$ -схем ( $BC$  6,  $BC$  7,  $BC$  8) общим условием является непустота  $T_0$ . Этим они отличаются от  $BC$  3,  $BC$  4,  $BC$  5. Разбирая  $BC$  3,  $BC$  4,  $BC$  5, мы исходим из предположения, что союз  $L_1$  служит для связи внутри данной фразы, а не для связи всей этой фразы с предшествующим текстом. Последняя ситуация соответствует другому случаю: три  $B$ -сегмента и одна левая граница.

**BC 1.**  $N_0 V_1 T_1 L_1 N_2 V_2 T_2 L_2 N_3 V_3 T_3$

**BC 2.**  $N_0 V_1 T_1 L_1 N_1 L_2 N_2 V_2 T_2 V_3 T_3$

**BC 3.**  $L_1 N_1 V_1 T_1 V_2 N_2 L_2 N_3 V_3 T_3$

**BC 4.**  $L_1 N_1 V_1 T_1 L_2 N_2 V_2 T_2 V_3 N_3$

**BC 5.**  $L_1 N_1 L_2 N_2 V_1 T_1 V_2 T_2 V_3 T_3$

**BC 6.**  $T_0 L_1 N_1 V_1 T_1 V_2 T_2 L_2 N_3 V_3 T_3$

**BC 7.**  $T_0 L_1 N_1 V_1 T_1 L_2 N_2 V_2 T_2 V_3 N_3$

**BC 8.**  $T_0 L_1 N_1 L_2 N_2 V_1 T_1 V_2 T_2 V_3 T_3$

Как было сказано выше, задача фрагментирования состоит в переходе от  $BC$  к  $LC$ . Рассмотрим соотношения между ними.

Больше всего вариантов возможно для  $BC$  1, которая может давать  $LC$  3.6.1,  $LC$  3.6.4,  $LC$  3.6.5,  $LC$  3.6.6,  $LC$  3.6.7,  $LC$  3.6.10. Однако, если  $T_3$  не содержит никаких границ, то возможны только  $LC$  3.6.1,  $LC$  3.6.4, а если  $T_2$  не содержит границ, то возможны  $LC$  3.6.1,  $LC$  3.6.5,  $LC$  3.6.6. В простейшем случае, когда ни  $T_2$ , ни  $T_3$  не содержат границ, мы имеем три последовательных простых предложения с делением по указанным левым границам.

$BC$  2 может давать  $LC$  3.6.5,  $LC$  3.6.7. Здесь  $T_2$  должно обязательно содержать хотя бы одну границу, а если  $T_3$  не содержит границ, то возможна только  $LC$  3.6.5.

$BC$  3 может давать  $LC$  3.6.2,  $LC$  3.6.9. Здесь  $T_1$  должно обязательно содержать хотя бы одну границу, а если  $T_3$  не содержит границ, то возможна только  $LC$  3.6.2.

$BC$  4 может давать  $LC$  3.6.3,  $LC$  3.6.8,  $LC$  3.6.9. Здесь  $T_2$  должно обязательно содержать хотя бы одну границу, причем если в  $T_2$  только одна граница, то  $LC$  3.6.8. исключается. Если  $T_1$  не содержит границ, то исключается  $LC$  3.6.9.

$BC$  5 может давать  $LC$  3.6.6 и  $LC$  3.6.8. Здесь  $T_1$  и  $T_2$  должны обязательно содержать границы, причем если в  $T_1$  только одна граница непосредственно перед сегментом  $V_2$ , то  $LC$  3.6.6 исключается.

$BC$  6 может давать  $LC$  3.6.4,  $LC$  3.6.10. Здесь  $T_1$  должно обязательно содержать хотя бы одну границу и обязательно имеет место вставка  $P_1\{P_2\}P_1^*$ . Если  $T_3$  не содержит границ, то исключается  $LC$  3.6.10.

$BC$  7 может давать  $LC$  3.6.6,  $LC$  3.6.7,  $LC$  3.6.10. Здесь если  $T_1$  или  $T_2$  не содержат границ, то исключается  $LC$  3.6.10. Если в  $T_2$  есть только одна граница, то  $LC$  3.6.7 исключается.

$BC$  8 дает  $LC$  3.6.7. Здесь  $T_1$  и  $T_2$  должны обязательно содержать границы.

Как уже отмечалось, правые границы предложений установить труднее, чем левые. При решении вопроса о том, имеет ли место вставка, полезно учитывать следующие сочетания.

Пусть фраза или раздел начинаются с последовательности

$$R_1 L R_2 V_1 G V_2 R_3,$$

где  $V_1$  и  $V_2$  — это В-сегменты, а  $R_1$ ,  $R_2$  и  $R_3$  — цепочки сегментов, не содержащие ни вершин, ни левых границ. Тогда имеет место вставка  $\Pi_i \{ \Pi_{i+1} \} \Pi_i^*$ . Важно, что между  $V_1$  и  $V_2$  есть только одна граница. Здесь  $\Pi_i = R_1 V_2 R_3$ , а  $\Pi_{i+1} = R_2 V_1$ .

С другой стороны, вставке препятствуют такие левые границы как двоеточие и сочинительный союз (что уже отмечалось), а также повторение союзов. Например, в последовательностях вида «, что  $\Pi_i$ , что  $\Pi_j$ » или «, что  $\Pi_i$  и что  $\Pi_j$ » предложение  $\Pi_j$  не будет вставлено в  $\Pi_i$ .

Из сформулированных условий понятно, как работают правила фрагментирования в рассмотренной ситуации. Строится В-схема фразы, если в ней три вершины и две левых границы, то проверяются условия на наличие и число границ в указанных выше местах.

Сходным образом, выявляя если не единственное, то наиболее вероятное членение, работают правила фрагментирования при одной левой границе или отсутствии левых границ.

## § 5. Синтаксический словарь (ССлов)

Синтаксический словарь содержит сведения о синтаксических свойствах слов. Центральную часть этих сведений составляют модели управления, описанные в [5].

В данной работе мы охарактеризуем те свойства, которые релевантны для рассматриваемой задачи, т. е. возможность слов управлять придаточными предложениями. Придаточное предложение может относиться к отдельному слову главного или к главному предложению в целом. В последнем случае для единообразия можно считать, что оно подчиняется вершине главного предложения.

Ниже иногда для краткости говорится "такое-то слово управляет таким-то союзом" вместо "такое-то слово управляет придаточным предложением, вводимым таким-то союзом".

Придаточные предложения могут выполнять роль актантов, сирконстантов, определений.

Наиболее индивидуализирована возможность слова иметь сентенциальные актанты, присоединяемые определенными союзами. Поэтому эти возможности естественно описывать как словарные признаки слов. В синтаксическом словаре, разработанном в рамках описываемого подхода, для существительных, глаголов, прилагательных и наречий охарактеризована возможность иметь сентенциальные актанты, вводимые теми или иными союзами. При этом не только отмечался сам факт наличия такой возможности, но и оценивалась "степень возможности" по трех- или четырехбалльной шкале. Эти оценки нужны для определения наиболее вероятной структуры при наличии для одного придаточного нескольких потенциальных управляющих на этапе перехода от линейной схемы к структурной, а затем к структуре. Иными словами, этот переход также находится в рамках подхода "анализ на основе предпочтений", который описан в работах [5–9].

Часть словаря, содержащая существительные, разработана автором, а часть, содержащая прилагательные и наречия, Т. Б. Агранат. Некоторые данные о приименных придаточных опубликованы в [2]. Свойства глаголов описаны Е. Л. Козьминой.

**5.1. Существительные.** Для существительных в ССлов описана возможность иметь в качестве актанта придаточное, вводимое союзами «что»

или «чтобы». При этом различались случаи, когда исследуемый актант вполне возможен — сильная оценка, и случаи, когда он только допустим — слабая оценка.

Следует отметить, что оценка степени возможности определенного актанта делалась для нейтрального контекста. Замечено, что некоторые специфические контексты повышают эту степень. К таким контекстам относится сочетание существительного с глаголом, являющимся от него лексической функцией *Орег* или *Func*. Видимо, тут сказывается следующее обстоятельство. Вообще говоря, способность иметь сентенциальные актанты у глаголов выше, чем у существительных, а сочетание существительного с глаголом=лексической функцией воспринимается приблизительно как единый глагол. Например, «ощущать беспокойство»=«беспокоиться», «отдавать приказ»=«приказывать», «испытывать торжество»= «торжествовать» и т. д. Соответственно сочетание «Он испытал торжество, что так все получилось.» возможно, хотя сочетание «торжество, что» плохое. Такое повышение возможности иметь сентенциальный актант должно учитываться не словарными, а общими правилами.

Возможность или невозможность иметь сентенциальные актанты во многом определяется тем, к какому семантическому классу относится рассматриваемое существительное.

Семантический класс **СОСТОЯНИЯ** в *ССлов* включает 111 слов. В нем введено деление на подклассы. Например, в подкласс слов, имеющих смысл 'действие и состояние как его результат', вошли, в частности, «возбуждение», «осознание» и т. п.; в подкласс слов со смыслом 'состояние=чувство по некоторому поводу' вошли, в частности, «беспокойство», «боязнь» и т. п.

Из 111 слов этого класса 25 не управляют ни «что», ни «чтобы». Для 86 слов, имеющих эти возможности, оценки распределились следующим образом: для «что» сильную оценку получили 54 слова, а слабую оценку — 19 слов, для «чтобы» сильная оценка у четырех слов, а слабая у двух слов. Могут управлять как «что», так и «чтобы» — семь слов. Например, слова «беспокойство», «опасение» получили сильную оценку для «что» и слабую для «чтобы». Слова «боязнь», «страх» получили две сильные оценки, иными словами сочтены одинаково возможными выражения «боязнь, что он не справится» и «боязнь, чтобы не получилось хуже».

Заметим, что для слов, которые могут выражать как процесс или действие, так и состояние, являющееся его результатом, ненулевые оценки относятся к состоянию=результату. Это объясняется тем, что рассматриваемые актанты как правило поясняют управляющее слово, раскрывая его содержание. Например, в предложении «Понимание, что нужно делать, пришло позднее» придаточное поясняет, что именно было понято, а не процесс понимания.

В семантический класс **ИНФОРМАЦИОННЫЕ ОБЪЕКТЫ** вошли слова, обозначающие информацию, сообщения, тексты и т. п. Например, в этот класс включены такие слова как «информация», «дезинформация», «мысль», «гипотеза», «высказывание», «обращение», «объявление» и др. Многие из этих слов имеют омонимы, означающие действие или процесс. Ср., например, «Объяснение длилось два часа» и «Он дал ясное объяснение ситуации». В этом классе всего 274 слова, из которых 132 не управляют ни «что», ни «чтобы». Остальным 142 словам даны следующие оценки: для «что» сильную оценку получили 85 слов, слабую — 33 слова; для «чтобы» сильная оценка у семи слов, слабая у шести слов. Могут управлять и «что», и «чтобы» 11 слов, причем обе сильные оценки у шести слов.

Заметим, что, так же как для омонимии 'процесс/результат', при омонимии 'действие/сообщение' ненулевые оценки относятся к сообщению, а не к действию, по той же причине: сентенциальный актант раскрывает содержание сообщения. Ср., например, «Пояснение, что найденное реше-

ние положительно, относится к первому случаю» и «Пояснение примеров продолжалось целый час». В первом предложении придаточное раскрывает, в чем именно состояло пояснение, а во втором, где подразумевается процесс с указанием субъекта, — актант, вводимый «что», недопустим.

В семантическом классе ДЕЙСТВИЯ содержится 795 слов, которые не вошли в пересечение с классами СОСТОЯНИЯ и ИНФОРМАЦИОННЫЕ ОБЪЕКТЫ. Из них получили ненулевые оценки 124 слова, причем, как и в двух предыдущих классах, в основном получили ненулевые оценки слова, выражающие одновременно и действие, и его результат, причем возможность иметь сентенциальные актанты относится ко второму значению. Для «что» сильную оценку получили 75 слов, слабую — 40 слов, для «чтобы» по одному слову с сильной и слабой оценкой. Могут управлять как «что», так и «чтобы» — 7 слов.

Проверено, что среди семантических классов, различаемых в Синтаксическом словаре, слова следующих шести семантических классов рассматриваемых актантов не имеют. Класс ЧЕЛОВЕК (320 слов), включающий наименования профессий («учитель», «биолог»), статусов («директор», «аспирант») и др. Класс УЧРЕЖДЕНИЕ (40 слов), например, «институт», «совет». Класс МНОЖЕСТВО (53 слова), например, «совокупность», «группа». Класс СВОЙСТВА (525 слов), например, «простота», «выпуклость». Класс ПОНЯТИЯ (152 слова), разделяющийся на подклассы: подкласс МЕСТО содержит 45 слов (например, «верх», «зенит»), подкласс ВРЕМЯ содержит 57 слов (например, «день», «минута», «год»), подкласс НАУКА содержит 50 слов (например, «математика», «геология»). Класс МАТЕРИАЛЬНЫЙ ОБЪЕКТ (230 слов), например, «компьютер», «пирамида», «станок».

Подчеркнем еще раз, что рассматривались именно сентенциальные актанты, а не любые придаточные, вводимые указанными союзами. Так не учитывалась возможность иметь придаточное, вводимое союзом «что», употребленном в значении «который». Сочетания вида: «Книга, что лежит на столе, принадлежит Петрову» — допустимы в разговорной речи, но не употребляются в письменных текстах научного и делового характера.

В ССлов не индивидуализирована возможность существительных иметь сирконстанты, вводимые данными союзами. Практически у любого существительного, обозначающего материальный объект, возможно придаточное, имеющее смысл 'назначение', вводимое союзом «чтобы» в значении «для того, чтобы». Например: «специальный секретарь, чтобы писать письма», «карандаш, чтобы он мог штриховать» и т. п. Аналогично широко распространена возможность иметь придаточное цели для существительных, обозначающих действие («переписывание, чтобы упорядочить» и т. п.).

Подобные свойства, присущие словам определенного класса, естественно назвать групповыми.

К числу групповых относится также возможность существительных, смысл которых включает оценку ситуации (как например, «успех», «радость», «провал», «поражение» и т. п.), участвовать в конструкции вида

Это (быть) С, что П.

Здесь С обозначает существительное, П — придаточное предложение, глагол «быть» указанный в угловых скобках, если он присутствует, согласуется с С. В этой конструкции П является сентенциальным субъектом, а С — именной частью сложного сказуемого.

Групповые возможности характеризуются в словаре указанием принадлежности к соответствующему классу. Они являются промежуточными между индивидуальными свойствами и общими, присущими всем существительным.

Так любое существительное может подчинять придаточное — определение, вводимое союзным словом «который». Поскольку это свойство явля-

ется общим, оно не помечается в ССлов. Для установления того, к какому существительному относится такое определение, проверяется согласование в роде и числе существительного с «который» или с прилагательным, если предложение, содержащее «который», начинается с конструкции

**ПРИЛАГАТЕЛЬНОЕ+«из которых»,**

например, «первый из которых» и т. п.

Если таких существительных несколько, то предпочтение отдается тому, которому предшествует указательное «тот». Так, например, в сочетании «... то решение уравнения, которое ...» придаточное относится к слову «решение», а не к «уравнения».

**5.2. Прилагательные и наречия.** Большой класс прилагательных может участвовать в конструкциях вида

**МОДИФИКАТОР+ПРИЛАГАТЕЛЬНОЕ+«чтобы» П.**

Здесь модификаторы — это: «вполне», «довольно», «слишком», «очень», «достаточно», «чересчур». Возможность сочетания с ними указана в ССлов. В сущности, этой возможностью обладают прилагательные, выражающие такое качество определяемого объекта, которое может быть представлено в разной степени. Например, к ним относятся «простой», «понятный», «яркий» и т. д., но не относятся «английский», «отрицательный» и др. В ССлов прилагательных первого вида 674. Например, можно сказать «Случай достаточно простой, чтобы служить примером».

При наличии модификаторов «такой», «не такой» прилагательные приобретают способность управлять придаточным с «что». Для кратких прилагательных аналогичную роль играет модификатор «так». Например, «Случай такой простой (так прост), что может служить примером». Отметим, что модификатор «не так» дает краткому прилагательному еще возможность управлять придаточным с «чтобы»: «Случай не так прост, чтобы служить примером».

Из общего числа русских прилагательных (более 20 тысяч) было выявлено 3791 таких, что образуемые от них наречия могут управлять придаточными предложениями.

Для 3356 наречий на «-о» была оценена возможность введения придаточных союзами «что» и «чтобы», причем отдельно для следующих трех конструкций (где Н=наречие). Первая: «это Н, что...», вторая: «Н, что...», третья: «Н, чтобы...». Разделение двух первых основано на наблюдении, что возможность участвовать в первой встречается чаще, чем для второй.

Из указанного числа прилагательных 81 могут участвовать во всех трех конструкциях. Например, «допустимый» позволяет сказать: «это допустимо, что...», «допустимо, что...», «допустимо, чтобы...». Три прилагательных могут участвовать во второй и третьей конструкциях. Например, наречие от «достаточный» допускает выражения: «достаточно, что...» и «достаточно, чтобы...».

Наречия от 12 прилагательных могут участвовать в конструкции с «чтобы», но не с «что». Например, «желательно, чтобы...».

Больше всего прилагательных, наречия от которых участвуют в двух первых конструкциях, — их 2437. Например, «это абсурдно, что...», «абсурдно, что...».

Наречия от 818 прилагательных участвуют только в первой конструкции, например, «это иллюзорно, что...».

Наречия от пяти прилагательных участвуют только во второй конструкции, например, «беспокойно, что...».

Кроме того, от 435 прилагательных образуются наречия на «-и», которые могут подчинять придаточные в конструкции типа: «это по-Н, что...». Например, «это по-дилетантски, что он так поступил».

У предикативных наречий прослежены следующие закономерности. Наречия со значением возможности и качественной оценки («возможно», «хорошо») присоединяют придаточные с обоими союзами; наречия со значением необходимости и долженствования («надо», «необходимо», «нужно», «нельзя», «невозможно») — только с союзом «чтобы»; наречия эмоционального состояния («грустно») — только с союзом «что»; наречия физического состояния («тепло», «холодно») не могут присоединять придаточные с данными союзами.

**5.3. Глаголы.** Как известно, глаголы по сравнению с другими синтаксическими классами обладают наибольшими возможностями управления придаточными. Соответственно эта часть словаря разработана наиболее детально. Для глаголов в ССлов указана возможность управлять придаточными, вводимыми следующими союзами, союзными словами или словосочетаниями: «где», «зачем», «как», «как бы не», «каков», «какой», «когда», «который», «кто», «куда», «ли», «насколько», «откуда», «отчего», «почему», «сколь», «сколько», «чей», «что1», «что2», «чтобы», а также возможность управлять прямой речью.

Были оценены 2000 глаголов. Понятно, что при этом глагол брался в одном определенном значении. Например, отдельно оценивались «выводить» в значении 'делать вывод' и «выводить» в значении 'удалять'. Выяснилось, что около 70% из рассмотренных глаголов не обладают релевантными возможностями. Так, сентенциальных актантов не имеют безличные глаголы, глаголы движения (например, «двигаться», «идти»), глаголы изменения фазы или состояния («начинать», «заканчивать», «увеличивать», «ухудшать»), глаголы физического действия («строить», «резать»), некоторые другие. Сентенциальные актанты характерны для глаголов, выражающих передачу, получение, переработку или хранение информации («доказывать», «знать», «помнить», «сообщать», «узнавать», «выводить1» и т. д.)

Для тех глаголов, которые могут иметь сентенциальные актанты, отдельно оценивались возможности глагола в активе и в пассиве или в возвратной форме, если она не является просто пассивом.

Для глаголов возможность управления сентенциальными актантами оценивалась по четырехбалльной системе: "вполне возможно", "возможно", "возможно с трудом", "невозможно".

Например, для глагола «думать» указано, "возможно" для всех перечисленных союзов, кроме «как бы не» и «чтобы». Для «как бы не» дана оценка "возможно с трудом", для «чтобы» оценка "невозможно". Однако отмечено, что при наличии отрицания управление придаточным с «чтобы» становится возможным. Например, можно сказать «Я не думаю, чтобы он этого хотел». Кроме отрицания, дополнительные возможности для управления придаточным, чаще всего придаточным с «чтобы», глаголу придают выражения модальности.

Отдельно оценивались возможности глагола управлять сентенциальными актантами при наличии соединительного «то», иногда для нескольких вариантов оформления «то» (падеж, предлог), если несколько мест в модели управления глагола допускают такое заполнение. Например, для глагола «аргументировать» в таблице оценок имеется четыре строки: 1) «аргументировать», 2) «аргументироваться», 3) «аргументировать то», 4) «аргументировать тем».

Наличие соединительного «то» может и повышать, и понижать оценки, и не влиять на них. Например, глагол «беспокоить» присоединяет сентенциальные актанты одинаково как с «то», так и без него, сочетаясь с любым из перечисленных союзов, кроме «чтобы». Глагол «видеть» более свободно присоединяет такие актанты без помощи «то»: например, для союза «что» без соединительного «то» дана оценка "вполне возможно", а с «то» оценка "возможно с трудом". Напротив, глагол «заинтересовывать» (в активе), мо-

жет управлять придаточными только посредством соединительного «тем», например, можно сказать «Он всех заинтересовал тем, как он решил задачу», а без «тем» — невозможно.

Заметим, что среди перечисленных союзов и союзных слов особняком стоит союз «чтобы», возможность управления которым отличается от того, что имеет место для остальных союзов. С одной стороны, очень большое число глаголов допускают все союзы за исключением «чтобы». С другой стороны, есть такие, как, например, «желать», которые допускают только «чтобы». Различные роли придаточных, вводимых «чтобы», рассмотрены Е. Л. Козьминой в работе [4].

#### СПИСОК ЛИТЕРАТУРЫ

1. Агранат Т. Б. Способы различения простого и сложного предложения при автоматическом анализе текстов // Русский язык: исторические судьбы и современность. Международный конгресс исследователей русского языка. Москва, филологический факультет МГУ им. М. В. Ломоносова 13–16 марта 2001 г. Труды и материалы. — М.: Изд-во МГУ, 2001. — С. 393.
2. Агранат Т. Б., Кулагина О. С. О приименных придаточных при автоматическом анализе текстов // Диалог'2000. Труды международного семинара Диалог'2000 по компьютерной лингвистике и ее приложениям. Т. 1. Теоретические проблемы., Протвино. — 2000. — С. 11–15.
3. Агранат Т. Б., Кулагина О. С. Об электронном словаре сочетаемости сложносочиненных и сложноподчиненных предложений // Диалог'2001. Труды международного семинара Диалог'2001 по компьютерной лингвистике и ее приложениям. Т. 2. Прикладные проблемы., Аксаково. — 2001. — С. 13–15.
4. Козьмина Е. Л. Семантические роли сентенциальных актантов, вводимых союзом «чтобы» // Русский язык: исторические судьбы и современность. Международный конгресс исследователей русского языка. Москва, филологический факультет МГУ им. М. В. Ломоносова 13–16 марта 2001 г. Труды и материалы. — М.: Изд-во МГУ, 2001. — С. 209.
5. Кулагина О. С. Об автоматическом синтаксическом анализе русских текстов. — Препринт / ИПМ им. М. В. Келдыша АН СССР. — М., 1987. — № 205.
6. Кулагина О. С. О синтаксическом анализе на основе предпочтений. — Препринт / ИПМ им. М. В. Келдыша АН СССР. — М., 1990. — № 3.
7. Кулагина О. С. Об аспекте меры в лингвистическом знании // Вопросы языкознания. — 1991. — № 1. — С. 48–60.
8. Кулагина О. С. Синтаксический анализ на основе предпочтений // Festschrift für Victor Jul'evic Rozenčevjg zum 80 Geburtstag. — Wiener Slawistischer Almanach. — Sonderband 33. — Wien, 1992. — P. 43–61.
9. Кулагина О. С. О проблемах автоматической обработки текстов на естественных языках // Интеллектуальные системы: Т. 1, Вып. 1–4. — М., 1996. — С. 109–116.

Поступило в редакцию 15 VIII 2001