

О приименных придаточных при автоматическом анализе текстов

Т.Б.Агранат, О.С.Кулагина

ИПМ РАН

agran@sosh.mccme.ru

Автоматический синтаксический анализ текстов на естественных языках дает в разных лингвистических процессорах различные синтаксические представления. Чаще всего с этой целью используются деревья составляющих и деревья зависимостей. В системе Анализа Русских Текстов (АРТ) синтаксический анализ строит дерево зависимостей с узлами-словоформами и согласованное с ним дерево фрагментов (подцепочек входного предложения). Об этом см.[1,2,3].

В последнее время начата разработка правил установления отношений между простыми предложениями в составе сложного. Эта работа ведется в рамках проекта 99-01-01191 РФФИ.

Вообще говоря, установление отношений между простыми предложениями в составе сложного базируется как на общих языковых закономерностях, так и на свойствах отдельных слов. К числу общих закономерностей относятся возможные структурные схемы сложных предложений и допустимые сочетания типов отношений между простыми предложениями при той или иной структурной схеме. Релевантные свойства конкретных слов - это возможность слов иметь сентенциальные актанты или сирконстанты.

Ниже рассматриваются один тип придаточных, а именно присловные придаточные, подчиняющиеся именам существительным, прилагательным и предикативным наречиям.

Существительные.

С точки зрения допустимых существительных конструкции с придаточным можно разделить на общие (в которых может участвовать практически любое существительное), групповые и индивидуальные.

К общим можно отнести конструкцию

«С, чтобы Гинф», где С обозначает существительное, а Гинф обозначает глагол в инфинитиве с подчиненными ему словами.

Например: «станок, чтобы сверлить», «карандаш, чтобы штриховать», «секретарь, чтобы писать письма» и т.п.

Ограничения на использование такой конструкции имеются, но они определяются не свойствами самого С, а осмысленностью сочетания глагола, обозначенного Гинф, и С.

Если С обозначает некоторый объект, то такое придаточное имеет значение «назначение» и союз «чтобы» можно понимать как «предназначенный для того, чтобы». Если С обозначает действие, то мы имеем придаточное цели, а «чтобы» раскрывается как «имеющий целью». Однако граница между назначением и целью является в данной случае весьма трудно определимой.

К общим можно отнести также известную конструкцию

«такой/=ая/=ое С, что П», где П обозначает придаточное предложение. Возможность употребить эту конструкцию определяется наличием слова «такой», поэтому может показаться, что П следует подчинять ему. Однако представляется более целесообразным подчинять П все-таки существительному С. В пользу такого решения можно привести следующие соображения. Формальное - подчинение П слову «такой» ведет к непроективности структуры, что вызывает неудобства. Содержательное - П раскрывает особенности С, на наличие которых указывает «такой».

В некоторых конструкциях могут присутствовать слова определенного класса. Такие конструкции мы называем групповыми, а возможность участвовать в них можно назвать групповым признаком таких слов. Примером является конструкция вида:

«<Это> <Гсв> С, что П».

Здесь «Гсв» в угловых скобках это глагол-связка («быть, становиться, являться» и т.п.), который может отсутствовать. В этой конструкции П является сентенциальным субъектом, а С именной частью сложного сказуемого. Здесь допустимы С, обозначающие некоторое событие или ситуацию и имеющие (за исключением самого слова «событие») положительную или отрицательную окраску. К допустимым С с положительной оценкой события относятся «победа, успех, удача, достижение, триумф, счастье, радость» и т.п., отрицательную окраску имеют слова «поражение, провал, катастрофа, неудача, несчастье, беда» и т.п.

Наиболее индивидуализирована возможность слова иметь сентенциальные актанты, присоединяемые определенными союзами. Соответственно эти возможности естественно описывать как словарные признаки слов.

Для существительных синтаксического словаря системы АРТ описана возможность иметь в качестве актанта придаточное, вводимое союзами «что» или «чтобы». При этом не только отмечался факт наличия такой возможности, но и оценивалась «степень возможности», т.е. различались случаи, когда исследуемый актант вполне возможен, и случаи, когда он только допустим. Для краткости значения указанного признака «степень

возможности» выражались числами. Значение 2 («сильная» оценка) показывает, что для данного существительного актант, вводимый союзом «что», вполне возможен, значение 1 («слабая» оценка) - что такой актант допустим, соответствующие значения для «чтобы» - это 12 и 11. Если при данном существительном может быть как тот, так и другой союз, ему приписывалось два значения (через запятую).

Заметим, что приводимые оценки отражают субъективное мнение авторов, статистических обследований не проводилось, причем, устанавливая эти оценки, авторы ориентировались на тексты научно-технического характера.

Следует также отметить, что оценка «степени возможности» определенного актанта делалась для нейтрального контекста. Замечено, что некоторые специфические контексты повышают эту степень. К таким контекстам относится, например, сочетание существительного с глаголом, являющимся от него лексической функцией Oper или Func. Видимо тут сказывается следующее обстоятельство. Вообще говоря, способность иметь сентенциальные актанты у глаголов выше, чем у существительных, а сочетание существительного с глаголом, являющимся его 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 без такого контекста «торжество, что П» сочетание плохое. Однако несмотря на то, что способность иметь сентенциальный актант в указанных сочетаниях повышается за счет употребления определенных глаголов, этот актант подчиняется все-таки существительному.

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.

Возможность или невозможность иметь сентенциальные актанты во многом определяется тем, к какому семантическому классу относится рассматриваемое существительное. Установлено, что такие актанты возможны для слов, которые в словаре системы АРТ отнесены к классам «состояния», «информационные объекты» и «действия». В этих трех классах содержится 1190 существительных из общего числа 2500 существительных в словаре системы АРТ.

В семантический класс «СОСТОЯНИЯ» вошли 113 слов. В нем введено деление на подклассы. Например, в подкласс «действие и состояние, как его результат» вошли, в частности, «возбуждение, огорчение, осознание» и т.п.; в подкласс «состояние=чувство по некоторому поводу» вошли, в частности, «беспокойство, боязнь, тоска» и т.п. Из 113 рассмотренных слов могут иметь рассматриваемые придаточные 89 слов, а 25 слов таковых не имеют. Оценки распределились следующим образом: для «что» оценку 2 (сильная) получили 53 слова, оценку 1 (слабая) - 22 слова, для «чтобы» оценку 12 (сильная) - 4 слова, оценку 11 (слабая) - 2 слова. Могут управлять как «что», так и «чтобы» - 8 слов.

Например, слова «боязнь, страх» получили две сильные оценки, поскольку можно сказать «боязнь, что он не справится...», а можно: «...боязнь, чтобы не получилось хуже». Близкие к ним по смыслу слова «беспокойство, опасение» получили сильную оценку для «что» и слабую для «чтобы».

Заметим, что для слов, которые могут выражать как действие или процесс, так и состояние, являющееся его результатом, ненулевые оценки относятся к состоянию. 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, названный «ИНФОРМАЦИОННЫЕ ОБЪЕКТЫ», вошли слова обозначающие информацию, сообщения, тексты и т.п. Например, в этот класс включены такие слова как «информация, мысль, гипотеза, высказывание, обращение, объявление» и др., всего 280 слов. Из них могут иметь рассматриваемые придаточные 151 слово, а 129 слов таковых не имеют. Распределение оценок следующее: для «что» оценку 2 (сильная) получили 87

слов, оценку 1(слабая) - 37 слов, для «чтобы» оценку 12(сильная) - 8 слов, оценку 11(слабая) - 6 слов. Могут управлять как «что», так и «чтобы»- 13 слов, из которых получили оценку 2,12 - 6 слов, 2,11 - 3 слова, 1,12 - 1 слово, 1,11 - 3 слова. Например, слово «предупреждение» получило две сильные оценки, поскольку можно сказать «предупреждение, чтобы так не делали, что это может иметь плохие последствия».

Заметим, что также как для омонимии «действие/состояние» при омонимии «действие/сообщение» ненулевые оценки относятся к сообщению, а не к действию, по той же причине: сентенциальный актант раскрывает содержание сообщения. Ср., например: «Объяснение проходило при закрытых дверях» и «Объяснение, что у него не хватило времени, всех успокоило». В первом случае «объяснение» означает действие, во втором - это текст.

В семантическом классе «ДЕЙСТВИЯ» содержится 797 слов, которые не вошли в пересечение с классами «СОСТОЯНИЯ» и «ИНФОРМАЦИОННЫЕ ОБЪЕКТЫ». Из них получили оценки 125 слов, и также, как в двух предыдущих классах, в основном получили ненулевые оценки слова, выражающие и действие, и его результат, причем возможность иметь сентенциальные актанты относится ко второму значению. Распределение оценок следующее: для «что» оценку 2 (сильная) получили 74 слова, оценку 1(слабая) - 40 слов, для «чтобы» оценку 11(слабая) - 4 слова. Могут управлять как «что», так и «чтобы»- 7 слов, из которых получили оценку 2,12 - 1 слово, 2,11 - 3 слова, 1,11 - 3 слова. Например, две ненулевые оценки получило слово «оговаривание»: «оговаривание, что функция непрерывна, не требуется», «оговаривание, чтобы не меняли последовательность действий».

Проверено, что среди семантических классов, различаемых системой АРТ, рассматриваемых актантов не имеют слова следующих шести семантических классов (всего в них 1320 слов).

Класс «ЧЕЛОВЕК» (320 слов) включает различные «наименования» человека с точки зрения профессии (например, «учитель, биолог»), статуса («директор, аспирант»), отношения к другим людям («помощник, друг») и др.

Класс «СВОЙСТВА» включает различные наименования свойств (525 слов, например, «простота, выпуклость, непрерывность, устойчивость» и т.п.).

Класс «ОРГАНИЗАЦИИ» включает различные наименования организаций, учреждений (40 слов, например, «комитет, совет» и т.п.).

Класс «МНОЖЕСТВА» включает различные наименования множеств (53 слова, например, «совокупность, толпа, собрание» и т.п.).

Класс «ПОНЯТИЯ» включает различные наименования понятий (152 слова), которые разделяются на подклассы: подкласс PL='место' содержит 45 слов

(например, «верх, зенит» и т.п.), подкласс ТЕМР='время' содержит 57 слов (например, «день, минута, год» и т.п.), подкласс SCI содержит различные названия наук - 50 слов (например, «математика, химия» и т.п.).

Класс «ПРЕДМЕТ» включает различные наименования материальных объектов (230 слов, например, «компьютер, пирамида, карандаш» и т.п.).

Прилагательные и предикативные наречия.

Как и в случае с существительными, здесь речь пойдет о возможности присоединения сентенциальных актантов, а не любых придаточных с данными союзами. Например, предложения с союзом «что» в значении «как», совершенно не допустимые в текстах деловой прозы, не учитываются. Что касается полных прилагательных, то они (кроме одного класса, о котором будет сказано ниже) могут сочетаться с вышеуказанными придаточными предложениями только при наличии слов «такой», «не такой» и «слишком» (присутствие этих слов всегда делает возможным присоединение данных придаточных). При первом возможно придаточное, вводимое союзом «что», при втором и третьем - вводимое союзом «чтобы». Перечисленные слова могут присоединяться ко всем качественным прилагательным, которые сочетаются с наречиями «более» и «менее». Сочетаемость с этими наречиями помечена в синтаксическом словаре системы АРТ. Из 2500 прилагательных, содержащихся в данном словаре, такую сочетаемость имеют 674.

Краткие прилагательные присоединяют придаточные с союзом «что», только имея определение «так», а с союзом «чтобы» - определения «не так» и «слишком».

(Аналогично подобным конструкциям с существительными сентенциальные актанты подчиняются прилагательным, а не словам, определяющим их.).

Как краткие, так и полные прилагательные, относящиеся к классу НАЗНАЧЕНИЕ (о чем есть соответствующая помета в синтаксическом словаре системы АРТ), могут присоединять к себе придаточное с союзом «чтобы»: «пригодный / пригоден, чтобы писать».

Предикативные наречия могут присоединять к себе рассматриваемые придаточные как с модификаторами «так», «не так» и «слишком», так и без них.

Так как предикативных наречий в русском языке относительно немного, сочетаемость можно задать списком, при этом прослеживается некоторая закономерность. Предикативные наречия физического состояния (например, «тепло», «холодно», «щекотно») и состояния окружающей среды (например, «ветрено», «пустынно») не могут присоединять придаточные с данными союзами; со значением возможности и качественной оценки («возможно», «хорошо») - присоединяют придаточные с обоими союзами, долженствования и необходимости («нужно», «необходимо») - только с союзом «чтобы», предикативные наречия эмоционального состояния («грустно», «стыдно») - только придаточные с союзом «что».

Также можно заметить, что семантические классы существительных и предикативных наречий, присоединяющие к себе сентенциальные актанты при помощи изучаемых союзов, во многом пересекаются. Предикативные наречия физического состояния часто бывают

омонимичны наречиям эмоционального состояния: «больно» и качественной оценки: «дурно». При автоматическом анализе текстов присутствие омонима в рассматриваемой конструкции будет являться критерием отнесения его к определенному семантическому классу. Во всех рассмотренных конструкциях (с существительными, прилагательными и предикативными наречиями) слова «так», «такой», «слишком» помогают правильно диагностировать придаточное предложение при автоматическом анализе. При отсутствии этих слов надежным критерием для дифференцирования различных «что» и «чтобы» будет форма предиката главного предложения. Если в главном предложении есть «полноценный» с семантической точки зрения глагол, то он сам присоединяет придаточное. Если же глагол - связка или он и вовсе отсутствует, то придаточное - сентенциальный актант, присоединяемый к существительному или предикативному наречию.

Литература.

Кулагина О.С. Об автоматическом синтаксическом анализе русских текстов. // Препринт Института прикладной математики им. М.В.Келдыша АН СССР, N 205, М., 1987, 22 с.

Кулагина О.С. О синтаксическом анализе на основе предпочтений //Препринт Института прикладной математики им. М.В.Келдыша АН СССР, N 3, М., 1990, 20 с.

Кулагина О.С. Синтаксический анализ на основе предпочтений. //Festchrift fur Victor Jul'evic Rozenecvejg. Wiener Slawistischer Almanach, sonderband 33, Wien 1992, p.43-61.