1. Во многих лингвистических процессорах требуется установить не только синтаксические связи между словоформами в предложении, но также и тип этих связей: смысловой, трансформационный или переводный [1:171; 8:252–267]. В [4:94–97] предпринято общесемиотическое, абстрактноалфавитное изложение указанной проблемы. Такой подход соответствует генотипической ступени [11:97–100] описания языков, т.е. описанию идеальных лингвистических объектов, отражающих существенные свойства языка и позволяющих изучать язык in vitro [12:47]. В данной работе проблема интерпретации синтаксических связей рассматривается на фенотипическом уровне, при котором идеальные генотипические объекты заменяются реальными объектами естественных языков. Такой двухступенчатый подход является средством описания сверхсложных систем, «средством управления сложностью объекта» (Н.М. Амосов), каким является естественный язык.

2. Как известно, одна и та же система и ее составные части могут быть описаны неединственным образом [3]. Выбор одного из нескольких описаний зависит от его цели и внешних критериев внешней оправданности, таких, как полнота, адекватность, экстраполяция результатов функционирования, от экономности, логической и дескриптивной простоты. Ниже рассматриваются три различных модели интерпретации синтаксических связей: 1) модель **семантических шаблонов**, 2) **семантических валентностей** и 3) **семантической преференции** [13]. Эти модели рассматриваются с точки зрения процедурной простоты их построения и применимости для синтактико-семантического анализа.

3. Наиболее естественным способом интерпретации синтаксических связей является использование **семантических шаблонов** (patterns, *B*-структур словосочетаний [9:62–63]), представляющих собой цепочки классов $B(x_1)$ $B(x_2)$, соответствующие данному словосочетанию *A* при данном разбиении *B*. При этом имеются в виду семантические классы [7:42;588–589; 10:380]. Для адекватности интерпретации синтаксических связей такая классификация должна быть непересекающейся [ср. 7:42]. В противном случае одному словосочетанию будут соответствовать различные семантические шаблоны с разными типами отношений между компонентами шаблона, т.е. омонимия отношений между компонентами словосочетаний останется неразрешенной. Добавим сюда и чисто эвристические трудности интуитивного распределения слов по заранее заданной совокупности семантических классов.

4. Классификацию слов, адекватную задаче интерпретации синтаксических связей, можно получить, задав процедуру «извлечения грамматики из текста» (ср. с дескриптивной и корпусной [1:112–137] лингвистикой). Исходными данными для такой процедуры является множество предварительно проанализированных примеров, а также форма «грамматики» для их анализа. Задается также алгоритм для заполнения этой «грамматики» соответствующей информацией (декларативными данными), извлеченной из исходных примеров (истории обучения). Соответствующая модель «грамматики» рассмотрена в [4:96–97] и может быть названа **моделью семантических валентностей**. Напомним, что в качестве информации к словам для интерпретации синтаксических связей в данном случае используются не семантико-синтаксические классы компонентов, как в п. 3, а набор семантических валентностей компонентов. В этот набор включаются типы семантико-синтаксических отношений таких словосочетаний, в которые может входить данный компонент. Для автоматической интерпретации синтаксических связей образуется логическое пересечение наборов валентностей компонентов данного словосочетания [4:96]. Сходная методика используется итальянской школой корреляционного анализа текста [2] и кембриджской школой семантики для разрешения лексической многозначности с помощью тезауруса [6].

Рассмотрим пример, иллюстрирующий описанную модель интерпретации синтаксических связей. Пусть будет задано некоторое

множество словосочетаний, соответствующих одной и той же синтаксической конфигурации, например, *V+из+Sg*, и каждому словосочетанию сопоставлен номер типа отношений между его компонентами из следующего списка:

1) пространственное: *выйти из аудитории*;

2) переносно-пространственное: *выйти из задумчивости*;

3) указание на источник информации: *узнать из книг*;

4) причинное: сказать из зависти;

5) объектное: сделать из глины.

Эти данные (мини-корпус проанализированных словосочетаний) можно представить в виде матрицы (см. таблицу 1).

Таблица 1

		аудитория	*театр*	*беда*	*рассказ*	*разговор*	*зависть*	*любопытство*	*интернат*	*глина*
	прийти		1					4	1	
	выбраться			2				4		
	узнать				3			4		
	увидеть	1				3				
	сказать						4			
	вылепить						4			5
	получиться	5	5		5	5				
	стоять							4		
	выползать									1

В столбце 1 табл. 1 перечислены главные компоненты словосочетаний, в строке 1 – зависимые компоненты. В клетках на пересечении строк и столбцов от 2 до 10 указаны типы семантических отношений в соответствующих словосочетаниях.

В табл. 1 каждому компоненту сопоставлен кортеж номеров. Заменим кортежи множествами номеров, устранив в каждом столбце и каждой строке лишь по одному вхождению каждого номера и для удобстве расположим и главные, и зависимые компоненты в строках. В результате получим таблицу следующего вида (см. табл. 2).

В этой таблице главным и зависимым компонентам словосочетаний, соответствующих определенной конфигурации, автоматически приписаны валентности на тип отношений между компонентами. Заметим, что валентностная классификация слов может не совпадать с семантической.

Таблица 2

	1	2	3	4
	Главные компоненты	Валентности главных компонентов	Зависимые компоненты	Валентности зависимых компонентов
1	*прийти*	1, 4	*аудитория*	1, 5
2	*выбраться*	2, 4	*театр*	1, 5
3	*узнать*	3, 4	*беда*	2
4	*увидеть*	1, 3	*рассказ*	3, 5
5	*сказать*	4	*разговор*	3, 5
6	*вылепить*	4, 5	*зависть*	4
7	*получиться*	5	*любопытство*	4
8	*стоять*	4	*интернат*	1
9	*выползать*	1	*глина*	1, 5

Например, зависимые компоненты *аудитория, театр, глина* входят в один валентностный класс для конфигурации *V+из+Sg*, но в разные тематические классы: ‘вместилище’ и ‘вещество’ [7:588]. Точно так же обстоит дело и с глаголами *сказать, стоять*. Они относятся к одному сочетаемостному классу в пределах названной конфигурации, но к разным тематическим классам: глаголам речи [7:42] и стативным глаголам [7:386]. Это еще раз подтверждает бóльшую адекватность для интерпретации синтаксических связей дистрибутивной классификации по сравнению с чисто семантической (см. п. 1).

Информация к компонентам словосочетаний из табл. 2 используется следующим образом. Предположим, что нам необходимо определить (автоматически, программно) значение словосочетания *вылепить из глины*. Для этого выбираем из табл. 2 информацию для каждого компонента этого словосочетания и образуем пересечение выбранных информаций: $\{4, 5\} \cap \{1, 5\} = \{5\}$. Верность результата подтверждается проверкой по исходному корпусу проанализированных словосочетаний. Действительно, на пересечении строки 7 и столбца 9 табл. 1 находится 5. Аналогичную проверку можно провести и для оставшейся части списка из табл. 1.

Описанный метод автоматической интерпретации синтаксической связи обладает важным свойством экстраполяции. Он позволяет проанализировать словосочетания, не использовавшиеся при формировании информации к словам. Например, устанавливаются

такие значения для словосочетаний, отсутствовавших в исходном корпусе: *выбраться из театра* → *1, узнать из разговора* → *3, получиться из глины* → *5* и др.

5. Рассматривая столбцы 2 и 4 в табл. 2, можно заметить, что распределение номеров в строках, помимо условия (1) [4:96], обнаруживает еще некоторые закономерности. В одном из этих столбцов всегда есть строка, которая:

а) содержит единственную валентность;

б) данная валентность не сочетается ни с одной другой валентностью в данном столбце.

Например, строки 3, 6 и 7 столбца 4 в табл. 2. В то же время строка 8 в том же столбце удовлетворяет первому требованию, но не удовлетворяет второму, поскольку номер 1 сочетается с номером 5 в 1-й, 2-й и 9-й строках 4-го столбца. То же относится и к строкам 5, 7–9 столбца 2, в которых единичные валентности: {4}, {5}, {1} – связаны с другими валентностями в строках 1–4 и 6 столбца 2.

Появление таких одиночных валентностей в столбце 2 связано с сильным управлением со стороны главного компонента словосочетания, а появление в 4-м столбце связано с тем, что предлог с зависимым словом образует устойчивое сочетание, связь которого с главным компонентом близка к примыканию. В обоих случаях один из компонентов словосочетания однозначно предсказывает тип связи между компонентами [10:380–381] и, таким образом, «единолично» определяет тип семантических отношений между компонентами. Аналогичный подход использован в алгоритме французско-русского машинного перевода: «Перевод многозначного предлога зависит от свойства слова, которое им управляет (обозначим его УПР), и от свойства, которое ему подчинено (обозначим его ППР). Влияние УПР на перевод подчиненного ему предлога определяется правым предложным кодом, а влияние ППР на перевод предлога определяется левым предложным кодом.... Воздействие предложных кодов в порядке убывания следующее: сильный левый код ППР, правый код УПР, слабый левый код ППР [5:181]». Однако информация о вышеуказанных кодах не дана в прямом наблюдении, и поэтому в данном случае требуется процедура «извлечения грамматики из текста». Эта процедура может основываться на таких соображениях.

Указанные выше «единоличные» валентности можно не указывать (т.е. стереть) в противоположном столбце информации, поскольку в подобных случаях для установления типа связи достаточно информации лишь при одном-единственном компоненте. Этот итеративный процесс стирания валентностей следует продолжать до тех пор, пока в табл. 2 не останется ни одного набора валентностей, удовлетворяющего сформулированному свойству «единоличности». При этом в каждом цикле стирания валентностей «единоличным» валентностям приписывается очередной индекс силы управления, или приоритет валентности. После применения описанной процедуры к материалу табл. 2 она примет следующий вид:

Таблица 3

	1	2	3	4
	Главные компоненты	Валентности главных компонентов	Зависимые компоненты	Валентности зависимых компонентов
1	*прийти*	–	*аудитория*	1^3
2	*выбраться*	–	*театр*	1^3
3	*узнать*	–	*беда*	2^1
4	*увидеть*	–	*рассказ*	3^3
5	*сказать*	–	*разговор*	3^3
6	*вылепить*	5^2	*зависть*	4^1
7	*получиться*	5^2	*любопытство*	4^1
8	*стоять*	–	*интернат*	1^3
9	*выползать*	–	*глина*	1^3

В качестве информации в табл. 3 компонентам словосочетаний приписана информация, состоящая из номера валентности и индекса силы управления: чем меньше индекс, тем больше сила управления. Для интерпретации синтаксических связей выбирается информация к компонентам и сравниваются индексы силы управления: словосочетанию приписывается тот тип, которому соответствует индекс с большей силой управления. Пусть, например, дано словосочетание *вылепить из глины.* Выбираем информацию к компонентам: 5^2 и 1^3, сравниваем индексы: $2 < 3$, поэтому для данного словосочетания выбираем тип связи 5, правильность чего проверяется по табл. 1.

Описанную модель интерпретации синтаксических связей будем называть моделью **семантической преференции** (preference semantics [13:114–151]).

6. Сравним три предложенные модели интерпретации.

Для модели семантических шаблонов требуется предварительная семантическая классификация компонентов словосочетаний. Однако при этом часто слова, одинаково ведущие себя с точки зрения интерпретации синтаксических связей, образуют классы, не совпадающие с теми, которые можно выделить на основе семантики. Кроме того, такая модель не допускает автоматической классификации компонентов на основе корпуса текстов.

Две другие модели, как было показано, такую возможность допускают. При этом модель преференции позволяет при одинаковых исходных данных увеличить полноту описания по сравнению с моделью семантических валентностей. Например, на основе табл. 3 для словосочетаний *выбраться из аудитории* и *выбраться из театра* правильно устанавливается пространственный тип отношений, в то время как на основе табл. 2 в этих случаях в связи с неполнотой исходного корпуса словосочетаний (см. табл. 1) вообще не устанавливается никаких отношений.

В то же время 2-я модель обладает тем существенным преимуществом, что она позволяет обнаруживать семантически не отмеченные словосочетания. Если наборы валентностей компонентов являются исчерпывающими и если их логическое произведение пусто, то сочетание является неотмеченным.

Такая проверка может быть использована при синтаксическом анализе методом фильтров. Пусть будет дано предложение ((*Мальчик* (*из интерната*)) (*стоял* (*у причала*))). Если исходить только из грамматической информации к изолированным словоформам, то этому предложению может быть приписана еще одна синтаксическая структура, такая, как, например, в предложении (*Мальчик* ((*из любопытства*) (*стоял* (*у причала*)))). На самом же деле первое предложение является синтаксически однозначным. Правильное решение в данном случае может быть принято так. Выбираем из табл.2 информацию к словам *выбирать* и *интернат* и образуем пересечение информаций: $\{4\} \cap \{1\} = \varnothing$. Поскольку произведение оказывается пустым, связь между этими словами оказывается невозможной. Поэтому правильным считается лишь первый вариант анализа.

Таким образом, для решения большинства лингвистических задач предпочтительной оказывается 2-я модель – модель семантических валентностей.

Литература

1. *Баранов А.Н.* Введение в прикладную лингвистику. – М., 2001. – С. 112–137. **2.** *Глазерсфельд Э.* «Мультистор» – система корреляционного анализа для английского языка // Автоматический перевод / Сб. статей. – М., 1971. – С. 281–318. **3.** *Иванов Вяч.Вс.* О приемлемости фонологических моделей. – Машинный перевод / Труды Института точной механики и вычислительной техники АН СССР. – М., 1961. – С. 396–412. **4.** *Кравчук И.С.* Автоматическое распознавание типов синтаксических связей // Вісник ХНУ імені В.Н. Каразіна. – № 627. – С. 94–97. **5.** *Кулагина О.С.* Исследования по машинному переводу. – М., 1979. **6.** *Мастерман М.* Тезаурус в синтаксисе и семантике // Математическая лингвистика / Сб. переводов. – М., 1964. – С. 160–176. **7.** *Падучева Е.В.* Динамические модели в семантике лексики. – М., 2004. – С. 5–607. **8.** *Попов Э.В.* Общение с ЭВМ на естественном языке. – М., 1982. – С. 252–267. **9.** *Ревзин И.И.* Модели языка. – М., 1962. – С. 62–66. **10.** *Тузов В.А.* Компьютерная лингвистика. – Internet-издание, 1998. **11.** *Шаумян С.К.* Генотипический язык и формальная семантика // Проблемы структурной лингвистики. – М., 1973. – С. 92–164. **12.** *Шрейдер Ю.А.* Топологические модели языка // Проблемы структурной лингвистики. – М., 1973. – С. 47–67. **13.** *Wilks Y.* An artificial intelligence approach to machine translation // Computer models of thought and langnage. – San Francisco: Freeman and Co, 1973. – P. 114–151.

АНОТАЦІЯ

Запропоновано 3 моделі автоматичного розпізнавання семантичних типів синтаксичних відношень: 1) модель семантичних шаблонів, 2) семантичних валентностей, 3) семантичних преференцій. Обговорюються подібності та відмінності між ними.

SUMMARY

Three models of automatic recognition of the semantic type of syntactic relations are proposed: 1) semantic patterns, 2) semantic valences, 3) preference semantics. The similarities and differences between them are discussed.